UNIVERSITY OF CRETE, SCHOOL OF SCIENCE AND TECHNOLOGY

Department of Physics

Bachelor Thesis

Methods for melanin quantification on fish scales and color regression on Porgus skin using Hybrid Microscopy and Machine Learning.

Author: Oikonomou Athanasios

Advisor: Dr. G. Tsironis
Supervisors: Dr. G. Zacharakis, Prof. G. Tsironis, Dr. G. Tserevelakis, Dr. G. Barmparis
Committee: Prof. Giorgos Tsironis, Dr. Giannis Zacharakis, Prof. Michael Pavlidis

Approved by the examining committee on May, 2021

Acknowledgments

First and foremost, I want to thank Dr. Zacharakis, Prof. Tsironis and Prof. Pavlidis, for giving me the opportunity and their trust on this collaborative work. Throughout this adventure, I have gained valuable skills and unparalleled knowledge. I want to especially thank my supervisor George Tserevelakis for always being by my side guiding, helping and always teaching me new things throughout every step of this work. I also want to specially thank my "coding" teacher, mentor and supervisor Dr. George Barbaris for all the time, advice and guidance he invested in me. Without his love for Machine Learning and his students, I would not have been in this position now. Last but not least, I want to thank Giannis Draganidis and Kostas Mavrakis for our excellent cooperation and teamwork as well as the exceptional companionship during the time we spent together in the lab. Thank you for the genuine interest in all my crazy thoughts and ideas. The lab felt like home because of you both and Mikis Mylonakis.

To my family and their support.

Abstract

Fishes display sophisticated skin chromatic properties that are of considerable ecological, physiological and behavioural importance. From behavioral regulations, to crypsis and adaptiveness to light, color patterns play an important role to the wellbeing of fishes. Contemporary methods of studying this subject either at a microscopic or macroscopic level, have many disadvantages and drawbacks. In this work, combining three disciplines, Physics, Computer Science and Biology, we propose novel procedures that measure and analyse the chromatic properties of fish skin, either at a microscopic or macroscopic level. Using a hybrid photoacoustic and fluorescence configuration we were able to create a method to quantify melanin concentration on a fish scales. Hybrid microscopy's advantages are clearly demonstrated in our results. The complementarity of such a configuration creates datasets with information properly separated and easy to work with. Furthermore, photoacoustic microscopy in particular, gives us quantified information due to the fact that each signal on an image is proportional to the melanin concentration. Using Machine Learning and image processing algorithms we were able to analyse the quantified results and prove statistical important, intra class, differences. A Convolutional Neural Network was created in order to automatically distinguish an input to its respective class, opening a road to further investigations on the subject. Finally, on a macroscopic level with the use of image processing and Convolutional Neural Networks, a model was created and trained that was able to robustly calculate the chromatic parameters on specific positions on a fish's body, possibly replacing the use of a colorimeter, the go to instrument when it comes to color measurements. Every model and method that is demonstrated here, is fully automated hoping to create useful and valuable tools for biologists studying the respective fields.

Contents:

1) Introduction

Color importance

Studying the coloration patterns in fishes is an important task. Fish display sophisticated skin chromatic properties that are of considerable ecological, physiological and behavioural importance. Fish species show high variability in skin color patterns and also high intra and inter species differentiations in coloration.

Intra species differentiations, although expected, exist due to the differences in feeding patterns, natural habitat, fish life style, UV exposure, predatory or survival needs etc. that differentiate from species to species[1].

As for the inter species differentiations, it has been observed that environmental conditions (background color, lighting spectrum, light intensity and water temperature) play an important role[2]. Other parameters that affect inter differences are fish ontogenetic phase and physiological condition, camouflage, thermoregulation and protection against solar radiation. Stress can also play an important role in the coloration patterns of a fish [4] as well as behavioral parameters like differences in and out of mating seasons [3] and a connection to aggression with melanin pigmentation[9].

Color production

At a molecular level, color patterns on fish are caused by large star-shaped pigment containing cells, chromatophores, which are located in the skin. The chromatophores are grouped into subclasses based on the color of their pigment containing organelles: xanthophores (yellow), erythrophores (red/orange), iridophores (reflective/iridescent), leucophores (white), melanophores (black/brown) and the more rare cyanophores (blue)[10].

Fish chromatophores are found in dermis, epidermis and on the scales that cover the skin.

Colors are produced by the combination of the chromatophores and their interaction with each other and light.

Melanin

Melanophores, with their distinctive dark melanin-containing organelles, melanosomes, are relatively well studied compared to the other chromatophore types. Their main characteristic. In lower vertebrates, is their ability to aggregate and disperse. Aggregated melanin in melanophores produces a lighter color while a dispersion of melanin in the melanophore produces a darker color. In general pelagic fishes have a darker dorsal area and a lighter ventral area due to the number of melanophores, melanin content and/or dispersion. [2]

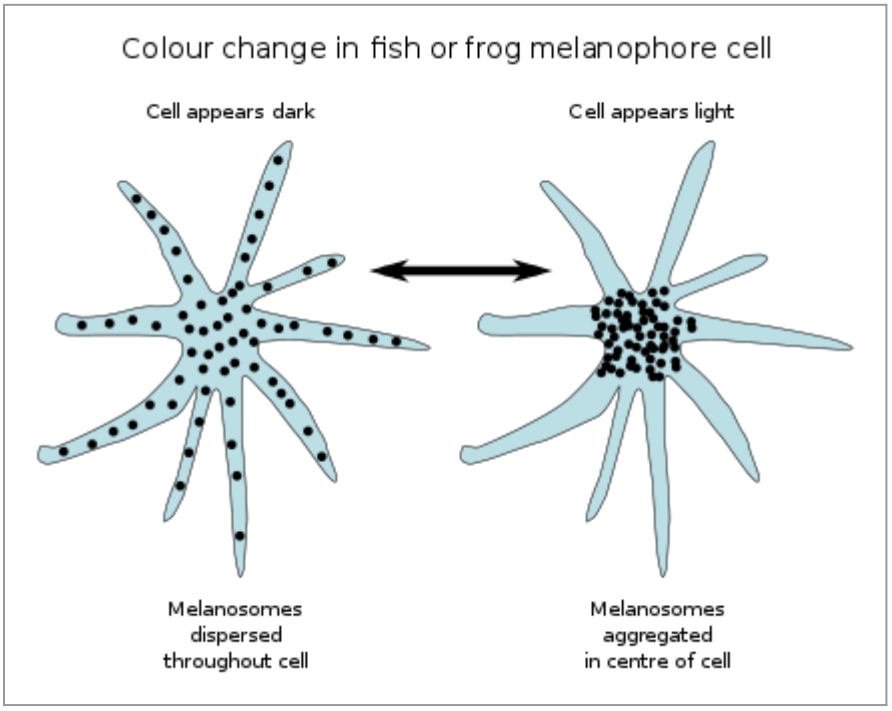


Figure 1: Left we can see the dispersion of melanosomes and on the right we can see an aggregation of melanosomes in a cell. [25]

One of the most important roles of melanin in fish, similar to humans, is the protection of cells' DNA from UV radiation. Adaptive mechanisms in fish adjust the melanin on their skins depending on the depth they live and disperse/aggregate it in a way that best protects the organism. As we already mentioned, melanin is a tool for crypsis for predatory reasons or safety from predation.

Melanin has an immediate correlation with the survivability of fish when it comes to global warming. It has been observed that, due to global warming, UV radiation exposure, temperature and humidity rise, and thus fish that carry non-negligible quantities of melanin adapt and survive better against global warming's effects. [7]

Finally, melanin has the ability to provide resistance to a fish from parasites and infections[11] Higher levels of melanin in the skin can deter the multiplication of parasites on the skin, due to its

toxicity as a substance, or signal immune responses to defer parasites from multiplying[8]

At a macroscopic level, studying the color of the fish can provide us with crucial information on its well-being, whether we're talking about stress or adaptation to UV exposure etc. When, due to conditions melanin cells are dispersed, the fish tend to look darker. Especially in aquaculture, monitoring the well being of the fish is important for both economical as well as biological reasons.

Economically, especially for reddish fishes like *Pargus pargus* and *Pagrus major* changes in color can affect the sales of the fish. It is reasonable as well to determine that when the color of a fish has been changed from the wild standard, it is an indication that something is wrong with it's living conditions[5]

Macroscopic color and melanin measurement procedures

As of now the standard ways of melanin quantification methods applied by biologists has many drawbacks. First and foremost, the fish on which the measurement will be conducted on, has to be sacrificed in order to dissect a part of its skin. Furthermore, the dissected part of the skin has to be processed with chemicals and be quantified by sophisticated specialized analytical techniques. Lastly the measurements are conducted by a stereoscope in order to obtain quantified measurements of melanin distribution[2]. It's clear that such procedures are human laboring, unfriendly to the fishes and nature and better approaches are needed. Hybrid Photoacoustic and Autofluorescence Microscopy on fish scales is a novelty approach for the quantification of melanin.

On the other hand, on a macroscopic level, a chromatometer is used in order to measure the chromatic parameters of a fish in specific spots (La*b* color space)[5]. Many samples are required for statistical analysis of the results. The product of those procedures are not tangible results, but indicators to see if different approaches or changes on specific conditions on the living conditions, food etc. of the fish in an aquaculture facility have to be made.

Our approach

This work is a combination of three disciplines, Biology, Physics and Machine Learning. Physics could provide really useful tools to biologists, like advanced microscopy and imaging methods, while Machine Learning and sophisticated programming could be an improvement on the data analysis required in every experiment. Accurate data analysis in the most automatic way possible are some of the many things Machine Learning can offer, especially in color and melanin measurements at both microscopic and macroscopic level.

We created a novel approach in melanin measurement with the use of hybrid photoacoustic and fluorescence microscopy. This method is less laboring and more accurate than the already used methods. It also provides more information as to the spatial distribution of the melanin cells and the actual melanin concentration levels. It is the first time that hybrid microscopy has been used to solve this specific problem and usually most of the works on hybrid microscopy are demonstrations, in contrast to our work. Machine Learning is a core part of our experiment as we

have used it in making the procedures and results we acquired fully automatic but also that the produced material from such a microscope is an excellent candidate for industrial applications using Artificial Neural Networks. In this part of the work physics is the cornerstone and Artificial Intelligence are excellent supplementary tools for analysis.

Furthermore, on a macroscopic level the color measurement is usually done manually through a chromatometer. As a supplementary part in our work we developed a fully automatic way to measure the chromatic parameters of a fish through digital high resolution images. The main interest of this part of the work is highly practical and could be combined with the microscopic findings in the feature. In this second part Artificial Neural Networks and Machine Learning are the cornerstones instead.

2) Theoretical Background:

2.1) Physics:

2.1.1) Photoacoustics

Photoacoustic microscopy is based on the photoacoustic effect. The information in the photoacoustic microscopy comes from an ultrasound wave produced by the oscillation of a sample (a cell, a tissue etc) as a result of absorbing the intensity of a pulsed laser beam. In contrast to other microscopy techniques that completely work on radiation and light detection to produce an image, photoacoustics detect soundwaves reflect significantly less than light when it comes to small structures. This way we can achieve greater penetration lengths in a given sample and even have the ability of in-vivo or 3-D imaging.

2.1.1.1) Photoacoustic Effect

A pulsed laser beam, laser with intensity changes in terms of time, hits a sample where it is absorbed by specific cells or tissues, depending on the application. The energy of the absorbed pulsed laser beam converts, mainly or entirely, to thermal energy resulting in the thermoelastic expansion of the absorbers. A pressure wave that travels through the sample is created. The pressure wave a photoacoustic microscope creates has typical frequencies from 1 to 150 MHz, commonly known as ultrasound waves, which in the end are detected by an ultrasound transducer.

The resulted ultrasound waves can be described by the following mathematical formula:

$$\left(\nabla^2 - \frac{1}{u_s^{\,2}}\frac{\partial^2}{\partial t^2}\right) p(r,t) = -\frac{\beta}{C_p}\frac{\partial H(r,t)}{\partial t} \quad (1)$$

where $p(r;t)$ stands for the spatiotemporal pressure function, $H(r;t)$ is the respective heating function defined as the thermal energy converted per unit volume and unit time following light exposure, u_s denotes the speed of sound in the propagation medium, β is the thermal coefficient of volume expansion and C_p represents the specific heat capacity at p, constant pressure. The source term of Eq. (1) is proportional to the first time derivative of the heating function, implying thus that the thermoelastic expansion of the medium will give rise to photoacoustic wave propagation only when the heating through the absorption of the incident irradiation changes over time - a constant light intensity would not cause any photoacoustic effect. This explains the reason why pulsed or intensity-modulated lasers resulting in time-variant heating are exclusively employed for this purpose. [12][13]

On laser excitation, the fractional volume expansion dV/V can be expressed as:

$$\frac{dV}{V} = -\kappa p + \beta T$$

Here, κ denotes the isothermal compressibility and β denotes the thermal coefficient of volume expansion p and T denotes the changes in pressure (Pascal) and temperature (Kelvin), respectively.

If the laser excitation is in both the thermal and stress confinements, the fractional volume expansion is negligible and the local pressure rise immediately after the laser pulse, p can be derived from:

$$p_0 = \frac{\beta T}{\kappa}$$

Furthermore, from Beer-Lambert law $I = I_o e^{-\mu x}$ and $Q = cmT$ we can derive the temperature-lasers energy flux equation:

$$T = \frac{\eta F \mu}{\rho c}$$

where η is the optical to thermal energy coefficient, F is the flux of optical energy and μ is the absorption coefficient. In standard photoacoustic experiments the temperature rises a few tenths of mK while the pressure changes are hundreds of $mbar$.

Finally the propagated sound waves are detected by an ultrasound transducer connected to

an oscilloscope. As seen in Figure 2, the detected sound waves result in a positive and a negative peak in Volts, representing respectively denses and thinings of the medium between the sample and the transducer, generated by the sample. So the peak to peak distance entails information of the radiation absorption rate of the sample. Different geometrical configurations of a transducer result in different waveforms and different sizes of transducer result in different active acoustical frequencies. The smaller peaks seen in Figure 2 are negligible and are formed due to reflections. Finally, an important factor is the duration of a pulse. A smaller pulse (in respect to time) results in more confined signals in time, thus better spatial resolution.

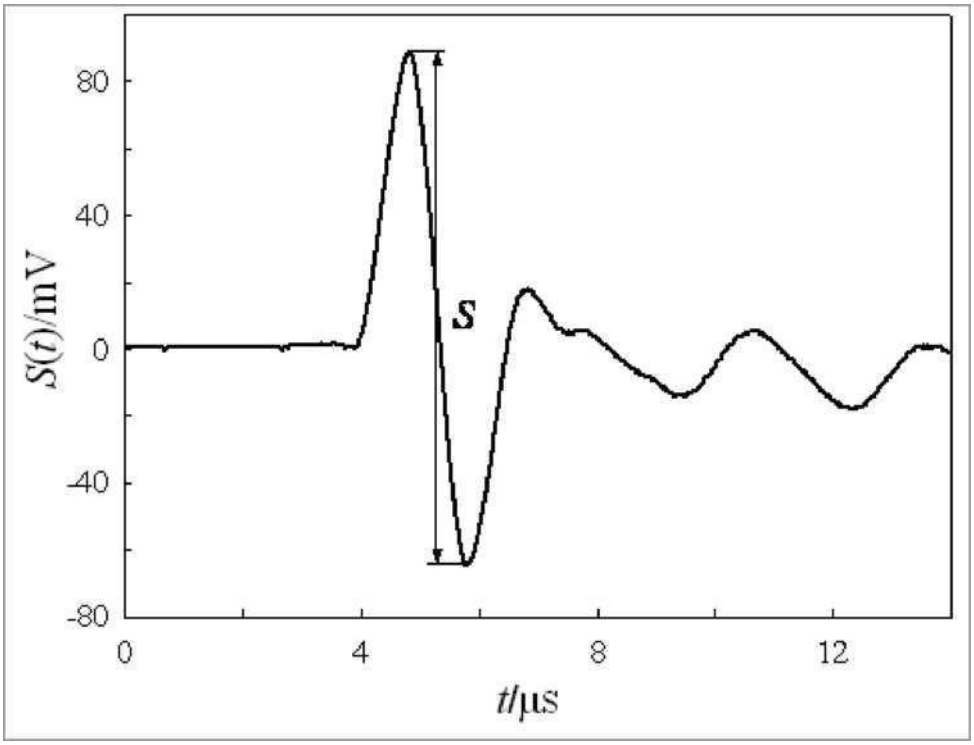


Figure 2: A typical photoacoustic waveform of amplitude S. The smaller peaks are a by-product of the reflections. [13][26]

2.1.1.2) Axial and Lateral Resolution of Photoacoustic Microscopy

Axial resolution of an photoacoustic microscope is defined by:

$$R_{axial} = 0.88 \cdot \frac{v_a}{Df_a}$$

where v_a is the speed of sound through a given medium and Df_a is the photoacoustic signal bandwidth. The main practical parameters that determine the axial resolution is the bandwidth of the ultrasound transducer used for a given experiment and the medium through which the sound

wave traverses. The bandwidth of the transducer must always match the bandwidth of the detected photoacoustic signal.

Lateral resolution is dependent on the optical and acoustical foci of the system. Two main categories of photoacoustic microscopy can be referenced on that differentiation, optical-resolution photoacoustic microscopy (OR-PAM) and acoustic-resolution photoacoustic microscopy (AR-PAM).

OR-PAM uses a tighter optical focus and it is more useful for imaging up to 1mm. It's lateral resolution is determined by the formula:

$$R_{lateral} = 0.51 \cdot \frac{\lambda_o}{NA_o}$$

where λ_o is the optical wavelength and NA_o numerical aperture of the optical objective lens.

AR-PAM has a greater optical scattering and it's more useful for imaging from 1mm up to 3mm. The lateral resolution of AR-PAM is determined by the formula:

$$R_{lateral} = 0.71 \cdot \frac{\lambda_\alpha}{NA_a}$$

where λ_α is the central wavelength of the photoacoustic wave and NA_α is the numerical aperture of the ultrasound transducer. [14]

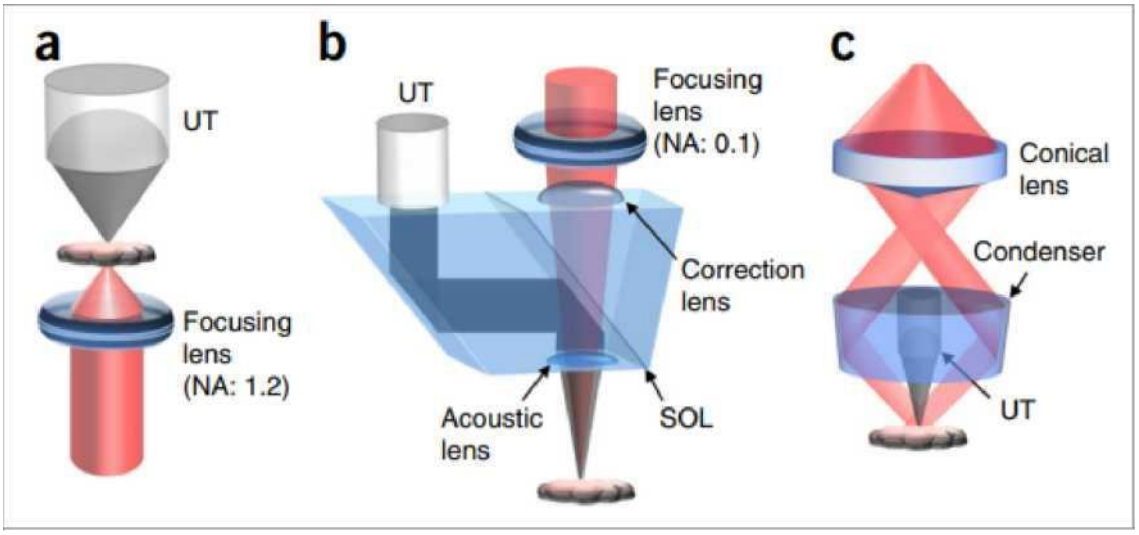


Image3:(a) Transmission-mode OR-PAM system, where the ultrasonic transducer (UT) and the water-immersion focusing lens are on opposite sides of the object98. (b) Reflection-mode OR-PAM system with an optical-acoustic combiner that transmits light but reflects sound. (c) AR-PAM system with a dark-field illumination12. The laser light is only weakly focused, with the UT in the dark cone [14]

2.2) Fluorescence

2.2.1) Fluorescence

Fluorescence is the emission of light from a substance or a molecule that has absorbed light or has absorbed electromagnetic radiation in general. The characteristic that makes fluorescence useful for microscopy is that the emitted light has a longer wavelength than the absorbed light due to Stokes shift. When a molecule absorbs electromagnetic radiation there are two ways to release the energy it receives. One is to immediately re-emit the photon and another is to lose energy, in the form of thermal energy through vibrational relaxation or collisions, and then emit a photon with less energy, thus the Stokes shift occurs. A Jablonski diagram depicts clearly the different types of energy emissions of an atom or molecule.

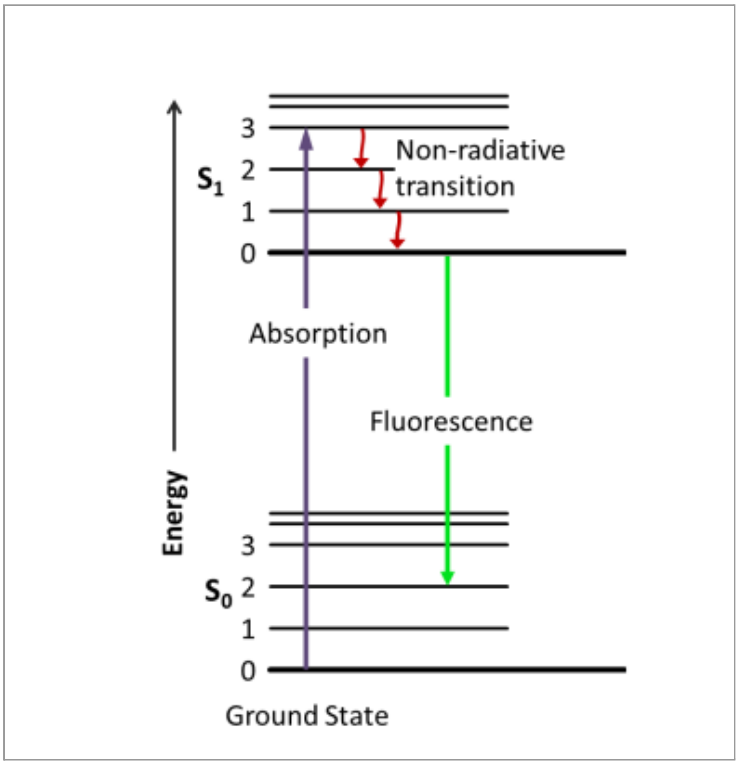


Figure 4: Jablonski diagram of the absorption and fluorescence re-emission at different wavelengths/energy. [27]

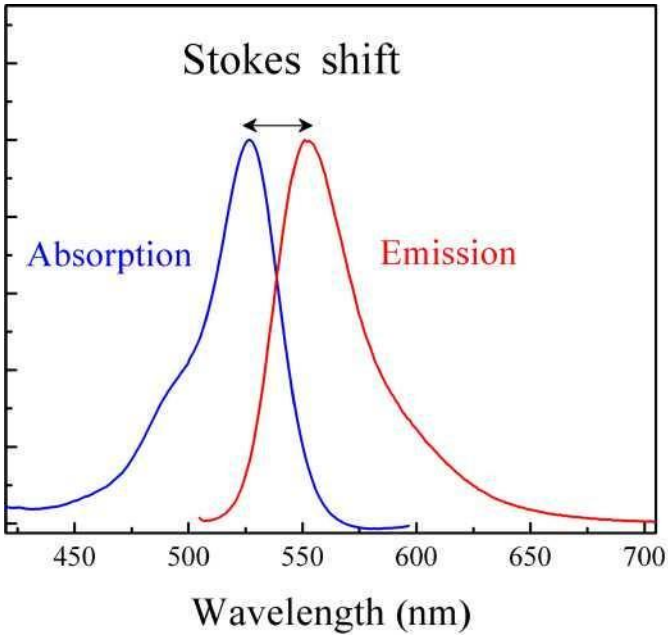


Figure 5: Stokes shift is the distance between the peaks of emission and absorption.[28]

Autofluorescence is the natural emission of light from substances that have been radiated. Such substances are collagen, lipofuscin, chlorophyll or NADH that can be found in biological samples during a microscopy experiment. If a sample does not have inherently such substances, appropriate fluorescent sub-substances are inserted to the sample. In a case where our sample carries inherently an autofluorescence substance then problems might occur to the imaging process if we add fluorescent agents. The autofluorescence substance might emit light and overshadow the light emissions of artificial fluorescence added with the purpose of an index on the structure we want to depict. [15]

Another difference between autofluorescence and fluorescence substances is that artificial fluorescent agents re emit photons after the radiation in approximately 10 s and thus the emission stops when the radiation from the laser stops, in contrast to autofluorescent substances where the emission continuous after the radiation stops [16]

Finally, in our experiment we did not use any artificial agents because fish scales contain collagen which, as mentioned, is an autofluorescent substance[17].

2.1.2.2) Fluorescence Microscopy

Fluorescence microscopy provides highly selective imaging properties and highly contrasted images when it comes to microscopy, thus it is widely used. In many cases, artificial fluorophores are inserted to the sample in a proper way in order to attach themselves to the areas of interest to re-emit light when radiated by a laser. Here confocal microscopy will be explained in a simple manner as stepping stone to a hybrid fluorescence and photoacoustic microscope used for the datasets creation in this work.

2.1.2.3) Confocal Microscopy

First, a light source passes through a pinhole e.g. a xenon or a mercury lamp radiates multichromatic light[18], and then passes through an excitation filter turning it to monochromatic. Such a filter allows only specific wavelengths of light to pass through, thus the monochromatic attribute. The now monochromatic light beam is reflected by a dichroic mirror towards the sample and is focused by an objective lens on a focal plane in the sample. Dichroic mirrors have the attribute of reflecting light at a specific wavelength and lets everything else pass undisturbed. Specific dichroic mirrors can be made for the relevant wavelength regions of excitation and fluorescence. The specimen is radiated and due to fluorescent/autofluorescent substances, light with different wavelength is emitted from the sample. The excitation area is at a focal plane but also above and below it. The emitted photons enter the objective and are focused on a detector. Before reaching the detector, they pass undisturbed from the dichroic mirror, an emission filter letting only the emitted theoretical wavelength pass, and a pinhole which is placed at the objective's focal distance.

The purpose of the second pinhole is to cut off emissions from below and above the focal plane, due to the geometry of the configuration, producing clearer images without noise. [19] In confocal microscopy the whole specimen has to be scanned, "pixel by pixel". This is achieved either by using galvanomirrors or using mechanical stages. In the first case the light beam scans the sample where in the second case the light beam is in a fixed position and the sample moves. The final image is reconstructed with the help of computer programs such as ImageJ, Matlab, Labview etc.

The main difference between wide field and confocal microscopy is that in wide field microscopy the sample is radiated whole. This produces noisy images because of different z axis planes radiating at the same time on the same (x,y) lateral space, aka "pixel", detected by the detector. The final product in either case is a 2D image. More specifically the lateral resolution of the wide field microscope is given by Rayleigh's criterion

$$r_{x,y} = 0.61 \frac{\lambda}{\mathrm{NA}}$$

where ***r*** is the minimum distance between two point objects (resolution), ***λ*** is the wavelength and ***NA*** is the numerical aperture of the objective lens. In optimal conditions, large numerical aperture and small wavelength, a confocal microscope can, theoretically, achieve resolution at ~200nm[20]. The lateral resolution of a confocal microscope is improved by a factor of ^2. The axial resolution of the confocal microscope is given by:

$$r_z = 1.4 \frac{\lambda n}{\mathrm{NA}^2}$$

In order to use with maximum efficiency the resolution of a given configuration, the pixel

size of an image has to be at least half of the resolution of the configuration (Nyquist criterion).

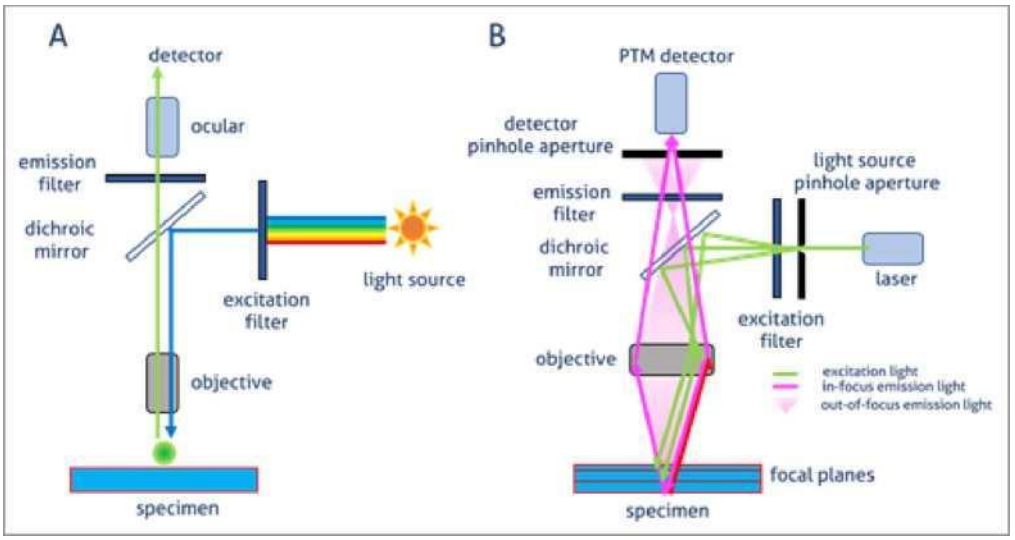


Figure 6: Comparison of wide-field (A) and Confocal (B) [21][29]

2.1.3) Hybrid Microscopy

In recent years photoacoustic microscopy has been used as a part of hybrid imaging configuration with incremental rate. The main benefit of a hybrid imaging configuration is the fact that it can give us more information than using the two separate imaging techniques separately.

Most commonly photoacoustic microscopy is used in combination with fluorescence microscopy. Two experiments can be used as an example here. The first, imaging on vegetal tissues [12] and the second one imaging on ciliary muscle [23]. In both experiments it has been observed that the hybrid imaging configuration returns really detailed results and in fact the signal of the two techniques are highly complementary. In a third research the use of the same configuration on ocular tumors led to a new way of classification on the tumors using the complementary nature of the returned signal [24].

Another type of hybrid configurations are those that combine photoacoustic microscopy and ultrasound microscopy. In the biomicroscopy field, such hybrid configuration could prove to be a useful tool for research on micro blood vessels. In a research that used such a configuration, for the imaging of a human finger, once more produced excellent results by depicting with the photoacoustic modulation parts of the finger (small blood vessels) that ultrasound microscopy could not.[25] In another research using the above configuration on zebrafish, the complementary nature of hybrid microscopy was again observed[26].

Concluding, the combination of two imaging techniques creating a hybrid configuration can surpass the problems an imaging technique could have on its own resulting in better results and uses that could not have been achieved any other way.

2.3) Digital Images

Digital images are nothing more than nxm matrices, with a value in every element of the matrix. The values on the matrix of a digital image are constrained between 0 and 255 where 0 is representing black and 255 is representing white. Images can have multiple channels, stacked matrices spatially coherent. Hyperspectral images can have many channels, each representing a different thing but in general digital images have 3 channels abiding to the laws of a specific colorspace. The most common one is RGB. Non-linear color spaces have been introduced throughout the years, such as L*a*b* (LAB) which will be described briefly as it has been used further below in this thesis.

RGB is an additive color model. Each RGB image has 3 channels. Each channel's pixel has a value for each one of the basic colors Reg, Green, Blue respectively. That value represents how much of that color will have the final color product. Again, in this colorspace the values of the pixels are represented in hexadecimal and can have values from 0 to 255. If for example an RGB image has 255 in the Red channel and 0 in the other two then the final product would be red. A spectrum of approximetly 16 million colors can be represented in the RGB colorspace per pixel but even with that there are colors that lie outside of the RGB spectrum. Another problem is that different devices can display the same triplet of RGB in a different manner. So other color models have been created.

LAB is a non linear color model and was built in order to make a standardized color space. In the LAB space the L* coordinate represents lightness, the a* coordinate is on a green-magenta axis, and the b* coordinate is on a blue-yellow axis. The lightness coordinate L* ranges from 0 to 100. The a and b coordinates are typically in the range -128 to 128 (though not explicitly constrained) and a* = b* = 0 denotes lack of color, i.e., a shade of gray from black (L* = 0) to white (L* = 100). In colorimetry, color differences are expressed as Euclidean distances between LAB coordinates and are denoted ΔE. [22]

2.4) Machine Learning

Machine learning is the science of programming computers so they learn from data. A

simplified translation of the process of a machine learning program is that it's algorithms try to find a function f: X -> Y that maps a set of input data X onto an output domain Y The type of function f is dependent on the nature and complexity of each given problem. A general definition of machine learning was given by Tom Mitchell in 1997 as:

A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

There are so many different types of Machine Learning (ML) systems and algorithms that is best to classify them in broad categories:

Supervised/Unsupervised Learning with Machine Learning

Supervised learning systems make use of labeled datasets (x, y) e (X ,Y), where x represents a data point and y the corresponding true prediction for x. This training set of input-output pairs is used to optimize the parameters of a function that maps a given input to a given output, predicting future input-output observations while minimizing errors as much as possible. In that case, the labeling process has to be done manually. We could say that when it comes to supervised learning, if an individual has done a process manually many times and kept a log throughout the process, a ML model can be trained so that he never has to do that same process again.

Unsupervised learning systems use unlabeled datasets to train the system. The objective of unsupervised learning is to derive structure from unlabeled data by investigating the similarity between pairs of objects, and is usually associated with density estimation or data clustering. Unsupervised learning is most often used in order to gain insights on a dataset. In many cases, especially when it comes to datasets with many dimensions (many features characterizing each sample instance) which are hard to visualize, could give information that could save time or direct towards a proper question about the data in question that was hard to ask just relying on theory. [23]

2.4.1) Supervised Machine Learning Algorithms

There are two main categories of supervised ML algorithms, Regression and Classification. Regression is used when through data we try to predict specific real values, float numbers (e.g. given some characteristics on an apartment what is the proper rent value for it) whereas Classification is used when we try to teach a computer to classify labeled objects (e.g. to tell us if a given picture is a dog or a cat). The whole point of this theory part of the thesis is to outline the basic concepts of machine learning that are used on the actual applications.

The main differences between a classification and a regression problem are the format of

our target values, the loss function and the Metrics that determine the performance when tested on test data.

With regression models we try to predict a specific continuous value thus the target values used in a regression problem are continuous. The most commonly used loss functions are the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE). These loss functions measure the distance of a predicted value from the actual value, tied to the input values. Using that information, regression models try to best change their variables in order to create the best function that connects input to output, minimizing the loss function. The metrics used in regression problems are the same as the loss functions.

With classification models we try to predict discrete values that denote in which class an instance of our data belongs to, thus the target values in a classification problem are discrete. The most common loss function used in classification problems is the Cross-entropy when dealing with binary classification problems or the Categorical Cross Entropy when dealing with multi-output classification problems. The metrics used to examine the usefulness of a model are always dependent on the problem. Some of the most common are Accuracy, Recall, Precision and F1 score. [23]

2.4.1.1) Loss Functions

In order to train a model we need a loss function. Two of the most known loss functions for regression are the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE). Depending on the problem, other loss functions could be more suitable but elaborating more on the subject is counterproductive for the purposes of this thesis. The two above mentioned metrics go by the names l1 and l2 norm respectively. For classification purposes Cross Entropy and Categorical Cross Entropy are the most common loss functions.

- Root Mean Square Error (MSE):

The formula for RMSE is given by:

$$RMSE(\hat{y}, y) = \sqrt{\frac{1}{M}\sum_{i=1}^{m}(\hat{y}^{(i)} - y^{(i)})^2}$$

where X is a matrix containing all the feature values (excluding labels) of all instances in the dataset. $x^{(i)}$ is a vector of all the feature values (excluding the label) of the ith instance in the dataset, and $y^{(i)}$ is its label (the desired output value for that instance). $\hat{y}^{(i)}$ is your system's prediction. When your system is given an instance's feature vector $x^{(i)}$, it outputs a predicted value $\hat{y}^{(i)} = h(x^{(i)})$ for that instance. m is the number of instances in the dataset you are measuring the RMSE on. M is the total number of instances in a given dataset.

RMSE calculates the euclidean distance of the predicted value for instance i by the model from the true value $y^{(i)}$.

- Mean Absolute Error (MAE)

The formula for MAE is given by:

$$MAE(\hat{y}, y) = \frac{1}{M}\sum_{i=1}^{m}\left|\hat{y}^{(i)} - y^{(i)}\right|$$

where the notations are the same for RMSE. It is sometimes called the Manhattan norm because it measures the distance between two points in a city if you can only travel along orthogonal city blocks.

In general loss functions measure the distance (in different mathematical spaces each) of a model's predicted value from the true value. [24]

- Cross Entropy

Cross Entropy is defined as:

$$CE = -\sum_{i}^{C} y_i log(\hat{y}^{(i)})$$

where y_i is the ground truth and $\hat{y}^{(i)}$ is the output of the model for each class i in C where C are the two classes. Usually an output function is used before the cross entropy is applied. These functions could be a sigmoid or a softmax for binary or multi labeled classifications respectively. [35]

2.3.1.2) Optimization

Gradient descent is an optimization algorithm and it is used in order to optimize the parameters (weights) of a given model. In order to implement Gradient Descent, you need to compute the gradient of the cost function with regards to each model parameter θj . In other words, you need to calculate how much the cost function will change if you change θj just a little bit. The equation for partial derivatives of the cost function in regard of each θj parameter is:

$$\frac{\partial}{\partial \theta_j} MSE(\theta) = \frac{2}{M} \sum_{i=1}^{m} (\theta^{\mathrm{T}} x^{(i)} - y^{(i)}) x_j^{(i)}$$

After calculating all the derivatives for all the training instances and by representing it as a vector:

$$\nabla_\theta MSE(\theta) = \frac{2}{m} X^T (X \cdot \theta - y)$$

we subtract that vector from the θ vector in order to minimize the loss function:

$$\theta^{(next\ step)} = \theta - \eta \nabla_\theta MSE(\theta)$$

Where *η* is called the learning rate of the procedure. If the loss function is a convex function then you are guaranteed to reach a global minimum. [23]

2.4.1.2) Evaluation

After obtaining the best model we need to evaluate its performance. The safest way to do so is to examine the results of the loss function in data that the model has never "seen" during the training. In order to do so we split the dataset into two subsets, the train set and the test set by randomly separating some data instances with their respective labels (test set) for evaluation purposes after the training phase is done on the rest of the data (train set).

The learning rate and the loss function for the training phase are called hyper parameters of the model. A hyper parameter of a model is a value given by the programmer and can be changed manually in order to improve the performance of the model. Grid Search is a common method used to fine tune the hyperparameters of a model. It is an algorithm that tries different combinations of different values for different hyperparameters in order to find the best solution. Different hyper parameters could result in different evaluation scores.

In regression problems the evaluation is quite straight forward. Most often we use the same

loss function used during the train process but on data our model has never seen before.

In classification processes the things are a little bit different. We want to know how many wrong predictions (or right predictions) our model has achieved again on data it has never seen before.

Classification Metrics

Some terminology must be given first before explaining the various classification metrics. That terminology is best applicable in confusion matrices [27]. Given a binary classification where the only results could be 0 (Negative) and 1 (Positive) we will denote a positive condition as P and a Negative condition as N. Then, if a prediction is correctly classified as P when it is indeed positive then we denote that prediction as TP (True Positive). Similarly if a prediction is classified as N and it is indeed Negative then we denote that prediction as TN (True Negative). If a prediction is classified as N but it should have been P then we denote that as FN (False Negative) and similarly we denote FP (False Positive) if a prediction is classified as P when in reality it is N. FP and FN are also called type I (underestimation) and type II (overestimation) errors respectively.

A confusion matrix is a proper arrangement of TP, TN, FP and FN results a classification process produces. An example can be seen in Figure 7.

		Predicted	
		Negative (**N**) -	Positive (**P**) +
Actual	Negative -	True Negatives (T**N**)	False Positives (F**P**) **Type I error**
	Positive +	False Negatives (F**N**) **Type II error**	True Positives (T**P**)

Figure 7: An example of confusion matrix and the arrangement of TP,TN,FP,FN results of a binary classification procedure. [30]

Finally we can explain the metrics used in a classification procedure.

- Accuracy

Accuracy is the proportion of the correct predictions over the total number of predictions we did on our test set.

$$\mathbf{Acc = (TP + TN)/(FP + FN + TP + TN)}$$

- Recall

Recall is the proportion of the images belonging to a label given for prediction, for each label, correctly classified as that label.

$$\mathbf{Recall = TP/(TP + FN)}$$

- Precision

Precision is the proportion of all the predicted images as a label that are actually that label.

$$\mathbf{Precision = TP/(TP+FP)}$$

- F1 Score

F1 Score is the harmonic mean of Recall and Precision.

F1 = 2* (Precision*Recall) / (Precision + Recall)

Concluding, the above notations can be generalized in multiclass classification procedures as well. There the only thing that stays the same is the general Accuracy. Precision, Recall and F1 can be derived for every specific class of the classification separately using only the predictions revolving around that specific class.

2.4.2) Unsupervised Learning

There are many unsupervised models used in various applications. [40] Usually, unsupervised models are used to gain insights on a problem or dataset. Maybe an unsupervised model will separate your data into 4 categories even though theory denotes that the specific data contains only 3 categories. It could also reveal information on how to approach the statement of a problem. In image processing unsupervised learning is used on images . to create clusters of colors or objects on images. This is what we used them for in this work so we will continue only explaining the simplest unsupervised model that exists, K-Means. This is also the one we used for image segmentation.

- K-Means Clustering

The k-means algorithm divides a set of N samples X into K disjoint clusters, each described by the mean μj of the samples in the cluster. The means are commonly called the cluster "centroids"; note that they are not, in general, points from X, although they live in the same space.
The K-means algorithm aims to choose centroids that minimise the inertia, or within-cluster sum-of-squares criterion:

$$\sum_{i=0}^{N} min(||x_i - \mu_j||^2)$$

where n is the number of data instances, x_i is the value of the i_{th} instance and μ_j is the mean value of the samples belonging to the j_{th} cluster. Inertia can be recognized as a measure of how internally coherent clusters are.

K-means is often referred to as Lloyd's algorithm. In basic terms, the algorithm has three steps. The first step chooses the initial centroids, with the most basic method being to choose k samples

from the dataset X. After initialization, K-means consists of looping between the two other steps. The first step assigns each sample to its nearest centroid. The second step creates new centroids by taking the mean value of all of the samples assigned to each previous centroid. The difference between the old and the new centroids are computed and the algorithm repeats these last two steps until this value is less than a threshold. [35]

2.4.3) Neural Networks

An Artificial Neural Network (ANN) is an architecture of neurons, represented as mathematical functions in a computer program, that are arranged in layers (in most cases) and are connected with each other in a proper way, in order to allow communication with each other. A neural network usually has an input layer and an output layer. Their size (number of neurons they contain) depends on the problem. An input layer has an equal number of neurons as the number of features a single instance in a dataset has. Similarly, the output layer has an equal number of neurons as the number of values we want to predict (or the categories it has to make a decision from). We will discuss all of these thematics on Neural Networks in the following paragraphs in an introductory manner in order to understand how they work and how we used them.

2.4.3.1) Perceptron

A perceptron receives a set of x values, represented here as a vector, as an input and computes the output $\hat{y}$. Each x_i parameter of the vector x is accompanied by a weight w_i multiplying the x_i value and a bias parameter, b, added to the product of the previous multiplication. Both w and b are represented here as vectors. Each neuron has an activation function. A mathematical function that takes as input the input vector z:

$$z = w_1 x_1 + w_2 x_2 + \ldots + w_n x_n + b = w^T \cdot x + b$$

and returns as an output the result of that function:

$$\hat{y} = g(z)$$

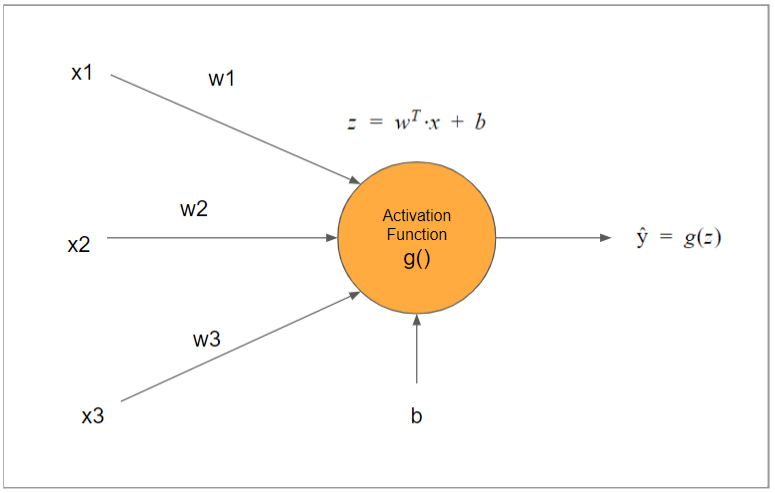


Figure 8: A perceptron with 3 inputs and 1 output. [custom]

2.3.3.2) Layers

A layer of is a configuration of perceptrons. If every neurons' output in a layer is connected with weight with all the neurons of the following layer it is called a Fully Connected Neural Network (FNN). We can generalize the previous notation we used on perceptrons to layers as follows. We denote a hidden layer (every layer between the input and output layer) as l and it's neurons as $n^{[l]}$. We denote the activation function of the i_{th} neuron in layer l as $a_i^{[l]}$. In that regard, the input of a neuron in layer l in regards to the output of the previous layer we have:

$$z^{[l]} = w^{[l]T} a^{[l-1]} + b^{[l]}$$

where:

$$a^{[l]} = g^{[l]}(z^{[l]})$$

and $g^{[l]}()$ is the activation function of the layer l. It is a common practice to use the same activation function throughout all the neurons of a layer.
Each neuron's output from the input layer to the next is the value of the data instance intact.

After a data instance passes through all the layers, reaching the output layer, following the above notations to make the proper calculations and save the results in proper matrices, we say that it finishes it's forward propagation.

2.3.3.3) Backpropagation

Backpropagation is the process of learning for a neural network. It is the process that updates it's weights on every layer in order to minimize the loss function. A loss function here is similar to the loss functions discussed on the general machine learning models discussed in previous paragraphs.

Backpropagation takes place after the forward propagation of either a single instance or multiple batches of instances from our data. Finally when every dataset instance has been fed to our network, then we say that an epoch has finished. A neural network usually trains for several epochs.

First step for gradient descent is to calculate the cost function (normalized loss function) and then, based on that result, change the parameters of the model in order to improve the loss function in further iterations. Let's say that loss function for our example is:

$$J = \frac{1}{M} \sum_{m=1}^{M} \sum_{k=1}^{k} (\hat{y}_k^{(n)} - y_k^{(n)})^2$$

where $m$is the m_{th} data point in M and k is the k_{th} node of the output vector in k.

In order to minimize the loss function by changing the weights and biases we must calculate the partial derivatives of J in regards to w and b beginning from the output layer until we reach the input layer. First we will introduce the $\delta^{[l]}_j$ where:

$$\delta^{[l-1]}_j = \frac{dg^{(l-1)}}{dz} \Big|_{z_j^{[l-1]}} \sum_{j'} \delta_{j'}^{[l]} w^{[l]}_{j,j'}$$

where g is the activation function of the lth layer, j' is where the weight ends and j where it comes from during the forward propagation. We end up with:

$$\frac{\partial J}{\partial w_{jj'}^{[l]}} = \frac{\partial J}{\partial z_{j'}^{[l]}} \frac{\partial z_{j'}^{[l]}}{\partial w_{jj'}^{[l]}} = \delta_{jj'}^{[l]} \, a_j^{[l-1]}$$

and:

$$\frac{\partial J}{\partial b_{j'}^{[l]}} = \delta_{j'}^{[l]}$$

Finally we update the weights by:

$$w'^{[l]}_{jj'} = w^{[l]}_{jj'} - \beta \frac{\partial J}{\partial w_{jj'}^{[l]}} = w^{[l]}_{jj'} - \beta \, \delta_{j'}^{[l]} a_j^{[l-1]}$$

and

$$b'^{[l]}_{j'} = b^{[l]}_{j'} - \beta \frac{\partial J}{\partial b_{j'}^{[l]}} = b^{[l]}_{j'} - \beta \, \delta_{j'}^{[j]}$$

2.3.3.4) Convolutional Layers

A convolutional layer consists of a set of filters, nxn matrices which are considered the weights of the layer, a non-linear threshold as the activation function, and a pooling layer which subsamples the information the final filter will extract. Convolution is a process where you “blend” two quantities in order to create a new one. Let’s consider a single filter F of dimensions nxn and a part of an image I again nxn in dimensions. The convolution of those two matrices is determined as:

$$F * I = \sum_{j=1}^{n} \sum_{k=1}^{n} w_{jk} i_{jk}$$

where w_{jk} and i_{jk} are the elements of each matrice, respectively.

The idea behind the CNNs is that a set of filters scan the original image and extract features through convolution and store the resulting in a new matrix, called the feature map of the filter. A nonlinear activation function is applied on the feature map and then it is pooled. The most common pooling function in CNNs is MaxPooling which finds the maximum value in a patch of the feature map, which works as a downsampling mechanism in the whole process. Finally, having many convolutional layers in a network, every next layer does the same procedures in the same order and takes as an input the feature maps of the previous layer.

In the end of a Convolutional part of a Neural Network all the final feature maps are flattened and are inserted as input on the fully connected part of the network (the classifier).

2.3.4) Dimensionality Reduction

In many cases, in order to train a Neural Network (or an ML algorithm in general) we must use data too big in size. The size of an input is measured by the number of the dependent variables it has. As a result, during the training, the calculations become time consuming or the resulting model has not reached its optimal capabilities. To solve that we use dimensionality reduction methods. One of the most common is Principal Component Analysis.

The central idea of Principal Component Analysis (PCA) is to reduce the dimensionality of a dataset where its dimension is the number of interrelated features it has. This is achieved by transforming our original feature values into new variables (principal components) that preserve best the total variance of our data. Principal components are uncorrelated with each other as they are orthogonal to each other.

The steps of a PCA algorithm are the following:

- Compute the mean value of every dimension in the dataset.
- Compute the covariance matrix between every dimension.
- Compute the eigenvalues and eigenvectors of the covariance matrix.
- Sort the eigenvectors by decreasing eigenvalues.

- Choose k eigenvectors and create a new orthogonal space as a representation of the original dataset in a new one with dimensions dxk

Calculating the mean values of each dimension is translated as calculating the mean value of each column, which is easy. The covariance of two given columns (X,Y) is calculated by:

$$COV(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{x})(Y_i - \overline{y})$$

where n is the number of samples, i is the ith instance and $\overline{x}, \overline{y}$ are the mean values of X and Y columns respectively. After finding the covariances between every column we can arrange a covariance matrix with dimensions dxd where d is again the dimensions of the original dataset.
Finding the eigenvalues and eigenvectors of this matrix is a simple linear algebra problem. In sort, we have to solve:

$$det(A - \lambda I) = 0$$

where A is the covariance matrix, I is the identity matrix of dimensions dxd and λ are the eigenvalues we want to find, find the eigenvalues and then find the vectors x that satisfy:

$$Ax = \lambda x$$

where x is an 1-D eigenvector.

Because matrix A is a covariance matrix, the eigenvalues represent the proportion of the total variance a given eigenvector preserves by projecting the original data on that vector. Sorting and picking the eigenvectors with a cumulative sum equal to the total variance we want is the final step of representing our data to smaller dimensions.

3) Main Part

3.1) Hybrid Photoacoustic and Autofluorescence Microscopy on fish scales for melanin quantification

3.1.2) Materials and Methods

3.1.2.1) Configuration

As already mentioned, the configuration used in this work was a hybrid microscopy configuration combining two different imaging methods, photoacoustic microscopy and fluorescence microscopy.

For the autofluorescence imaging we used as an excitation source a continuous wave (CW) diode laser (CP450, Thorlabs, Newton, NJ, USA; power output 4.5 mW) emitting at 450 nm. Because of the beam's elliptical shape, we used a two lens system (L1 and L2) and one pinhole to filter the beam and improve its quality. The beam was then aligned by len L4 before it passed thourgh a neutral density filter which decreased the beam's original power output, in order to prevent the destruction of the sample. Then the beam was refracted by a dichroic longpass mirror and then led to a properly modified inverted optical microscope after being magnified by a two lens (L5 and L6) telescope. An objective lens focused the beam on the sample which laid on the bottom part of a water tank with an optically transparent bottom. After the beam hits the sample a part of the returned fluorescence radiation is collected by the objective lens. Having longer wavelength than the lasers radiated beam it is transmitted through two dichroic mirrors (DM1 and DM2) and is reflected by a mirror. A filter for larger wavelengths (LP03-532RU-25, Semrock, Imc, Rochester, NY, USA; edge wavelength: 532nm) is placed in order to cut off the reflected original beam, with smaller wavelength than the autofluorescence and thus isolate the wavelengths of interest. The autofluorescence radiated wave passes through a pinhole of 50μm in order to cut off the light originating out of the focal point of the original beam. Finally it is detected through a PMT (H6780-20, Hamamatsu, Hamamatsu City, Japan). The resulting signals were recorded through a Data Acquisition card (DAQ) (PCIe-9852, ADLINK, Taipei, Taiwan, sample rate 200M/s; 90MHz) and was saved on a computer.

For the photoacoustic imaging a Nd:YAG laser with mutable repetition rate which emits at 1064nm (QIR-1064-200-S, CrystalLaser LC, Reno, NV, USA; pulse energy: 29.4 μJ, pulse

duration: 10 ns, selected repetition rate: 5kHz). The beam was focused on a trivoric lithium crystal of second harmonic (Castech Inc, Fuzhou, China) which transformed a part of the beams wave into a laser beam with half wavelength of the original (532nm). Then, the beam was aligned with L9 (lens) and passed through a transit zone filter (FF01-531/40-25, Semrock, Rochester, NY,USA) where the rest of the infrared radiation was absorbed. A polarizer was used in order to decrease the intensity of the laser to the wanted levels. The beam was reflected by a dichroic mirror (DM2), passed through a second dichroic mirror (DM1) and the systems telescope. Finally the beam was focused on the sample through an objective lens on the sample. The sound waves that propagated from the sample as a result of the radiation were detected by an ultrasound transducer, spherically focused, with centre frequency at 20 MHz (V373-SU, Olympus, Tokyo, Japan; band width: 13-33MHz, focal length: 32mm). The transducer was partially submerged into the water tank, coaxially and confocal with the focal point of the beam.

The water tank was filled with deionized water and was screwed on a base that is capable of moving by a motorized stage XY (8MTF75LS05, Standa, Vilinious, Lithuania) and a stage Z manually adaptable. The ionized water was used because the soundwave propagation is better through it, as medium, compared to air. Every measurement was the mean value of an average of 20 recorded signals.

Finally the detected signals were augmented by a low noise RF amplifier (TB-414-8A+, Mini-Circuits, Camberley, England; gain 31dB) and the acquisition card was coordinated by the signal enabling the laser.

MATLAB scripts and ImageJ were used for the image creation from the detected signals but also for the proper coordination of the different parts in the configuration.

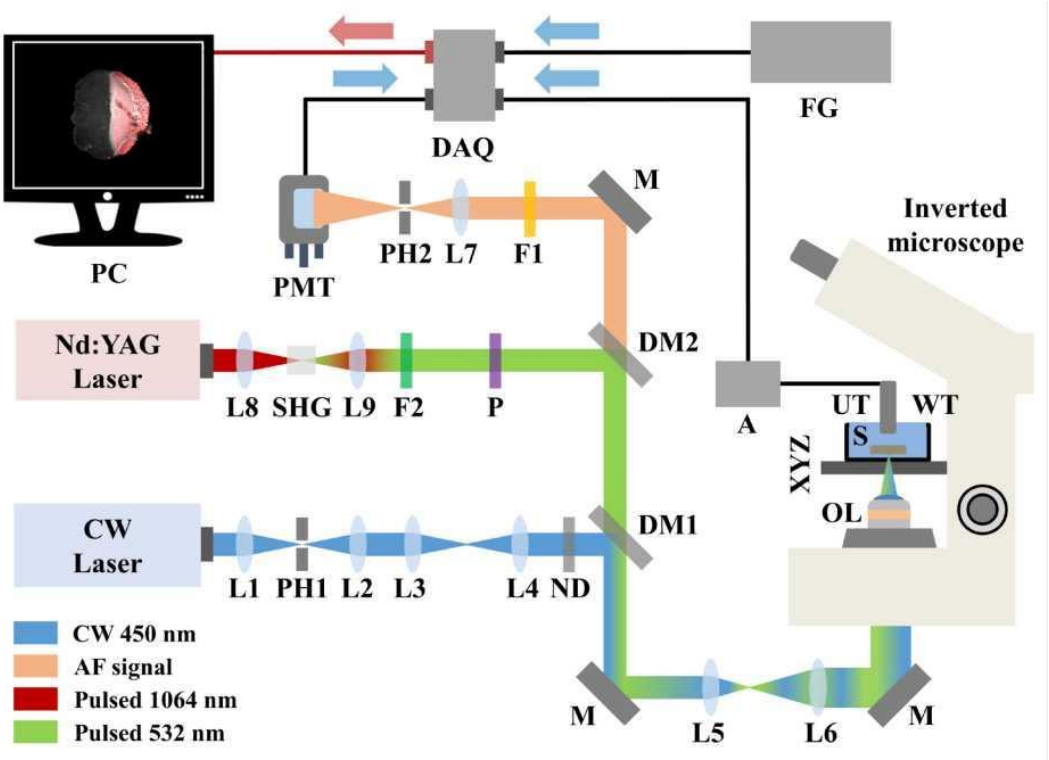


Figure 9: Visualization of the configuration. L, lens; PH, pinhole; ND, neutral density filters; DM, dichroic mirror; M, mirror; OL, objective lens; WT, water tank; XYZ, 3D mechanical stages; UT, ultrasound transducer; A, amplifier; DAQ, data acquisition card; PMT, photomultiplier tube; F1, low frequency passing filter; P, polarizer; F2, band passing filter; SHG, second harmonic generator crystal; PC, computer.

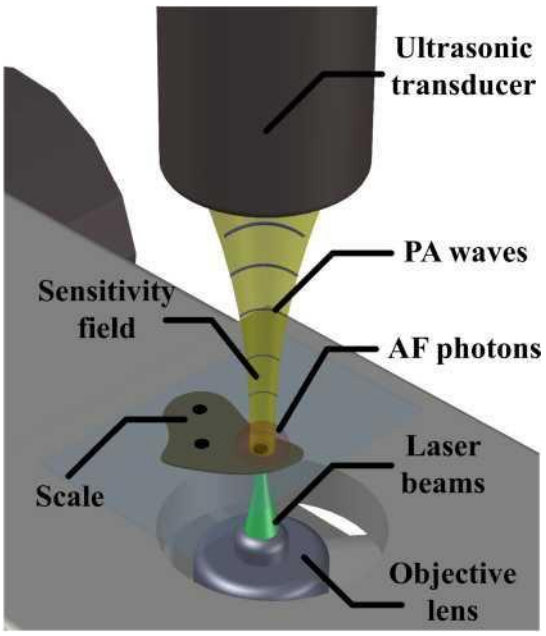


Figure 10: 3D representation of the acquisition process.

3.1.2.2) Sample preparation

The sample preparation begins by extracting the fish scales from the fish and imbuing them in 4% diluted formalin, washing them and then being stored in separate vials per class, containing 70% diluted ethanol. The vials were stored in a refrigerator for the span of the whole experiment which lasted for about 3 months. Before each measurement we placed some fish scales in the bottom of a custom made water tank, in which the ultrasound transducer is submerged for the measurement, and were covered with transparent ultrasound gel so the samples do not move during the scanning phase of the stages and to prevent the contact of the samples with the deionized water that filled the water tank. A side note on the preparation is that the ultrasound gel often stores bubbles of air in its body. Proper 'grooming' of the gel is necessary in order to get clear images during the scanning.

3.1.2.3) Measurement procedure

First we attach the water tank that contains the scales onto the stages and fill the tank with deionized water. We focus the light beam's focal point, axially, on the middle of the sample. The scanning is conducted on the x,y plane. The quality of the resulting fluorescence image is highly dependent on the axial focus of the beam. This is done by taking a sample measurement before obtaining the actual image of the dataset, while adjusting the height in micrometers of the stages on the z axis. Then we adjust the position of the ultrasound transducer in regards to the sample in order to maximize the signal we're receiving from the melanin. This is done by monitoring the peak to peak signal from an oscilloscope while we radiate a strip of black tape on the bottom of the tank, approximately on the same height as the sample. The difference on the thickness of the scale versus the tape is negligible compared to the depth of view the photoacoustic phenomenon has.

3.1.2.4) Programming

All the code written for this work is in Python. For the image segmentation algorithms K Means [31] from sklearn library was used. Filters, contour finding and flood fill algorithms were taken by the OpenCV libraries [32] in order to properly segment the images. The datasets were splitted in appropriate subsets using sklearn[33]. Artificial Neural Networks were created and trained through Tensorflow [34] and Keras[35] libraries. Adam optimizer[36], LeakyRelu activation function [37] as well as EarlyStopping and Checkpoint modules were taken from the respective Keras libraries as well. Finally, all the code was written in Jupyter Notebooks.

3.1.3) Results

We first had to test whether the obtained images from the photoacoustic and fluorescence hybrid microscope were in spatial coherence with the images a biologist would have obtained through a stereoscope. In order to do so we compared a measurement from an image obtained by the hybrid microscope and one obtained by a stereoscope. The two images are on the same scale. The results are shown in Figure 11.

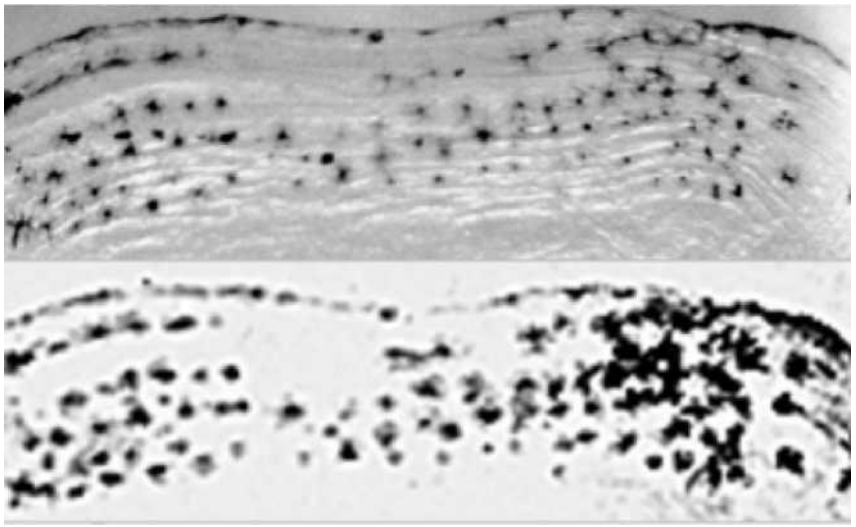


Figure 11: On the top image we see the area of interest on a sea bream scale and on the bottom image we see an image of the same area obtained by the hybrid microscope. The black areas are the melanin cells.

Then we scanned some fish scales from different categories to test if an observer could see the differences in melanin described in theory per category, from the results. In Figure 12 present measurements acquired from three fish scales, each belonging to a different category. On the first column present see an average sized scale from the fluorescence channel per class, on the second column the melanin each scale contains per class, and on the third column a combination of the two channels. From top to bottom we have a class on each row, Red porgy, sea bream and sea bass respectively.

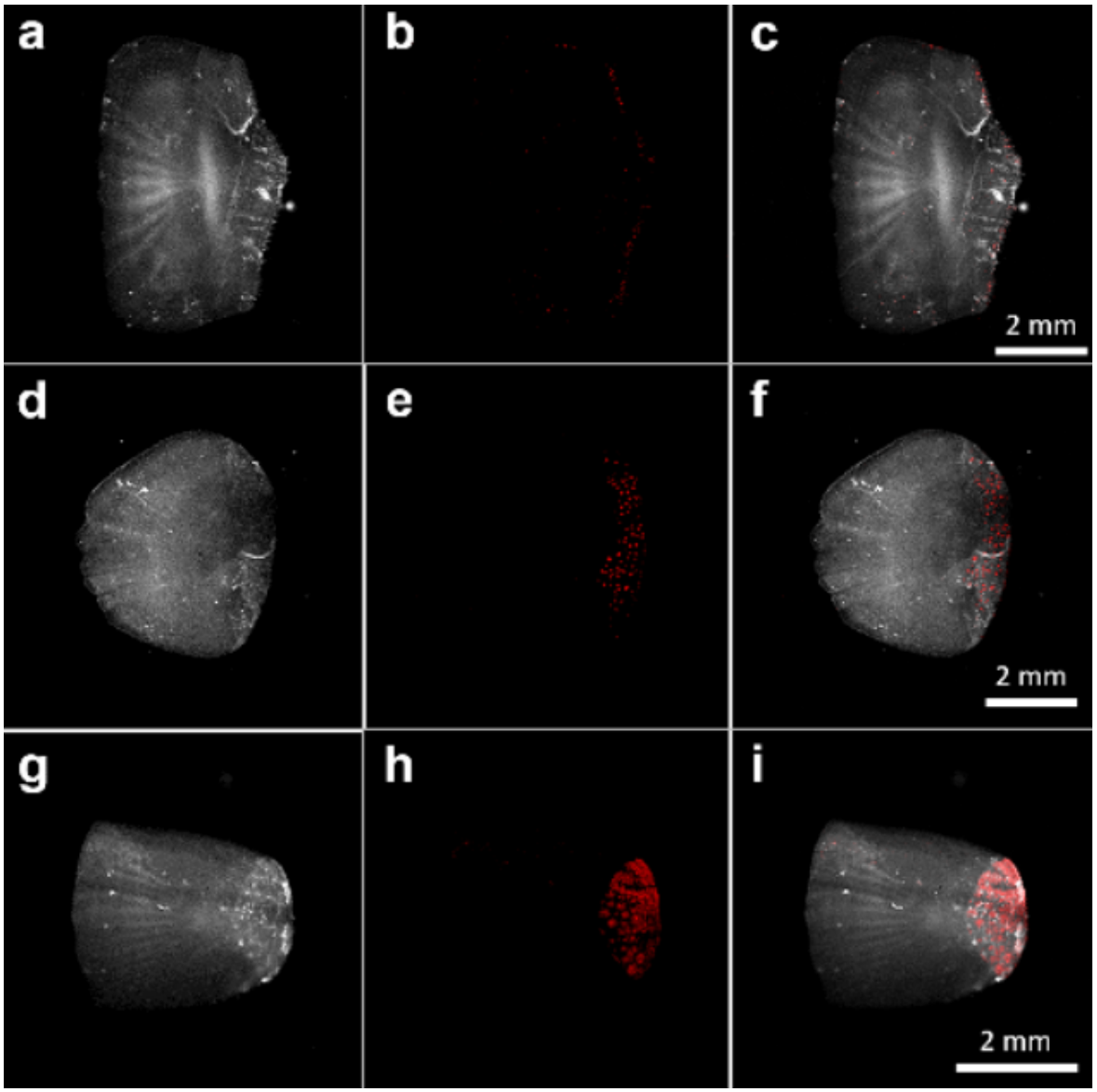


Figure 12: Hybrid imaging of 3 scales belonging to different categories. (a) Autofluorescence image of red porgus scale. (b) Photoacoustic image of the same scale. (c) Hybrid image, combining the two previous images. (d) Autofluorescence image of a sea bream scale. (e) Photoacoustic image of the same scale. (f) Hybrid image, combining the two previous images. (g) Autofluorescence image of a sea bass scale (h) Photoacoustic image of the same scale (i) Hybrid image, combining the two previous images.

We propose five parameters that best describe a fish scale based on the information hybrid microscope provides, the scale area and the melanin signal on it. The values that we used to best describe a scale are the surface of the fish scale, the total melanin each scale has, the area of coverage of melanin on the scale, the density of melanin, the percentage of coverage of surface from melanin and the total melanin per melanin cell. In order to calculate those values we need the area of the scale and the total area of the melanin covering the scales. We preprocessed our data in order to cut off the background noise in a way that doesn't alter the actual measurements when it comes to melanin (Photoaccoustic channel) and applied an object detection algorithm on the scale (Fluorescence channel) in order to find the total area.

Finally, before building an actual model that classifies the images we conducted a statistical analysis on the values that best describe the melanin and the fish scales, to actually

determine if the input images have statistically significant differences per class.

3.1.3.1) Original Dataset

We created a dataset of 300 images. We acquired 100 images of red porgy, 100 images of sea bream and 100 of sea bass. We call them images but they are actually 400x400 2-D matrices containing the raw signal returned for each pixel during the acquisition process in Volts. The biggest difference between our raw signal matrices with an image is that an actual binary grayscale image has values between 0 and 255 as explained in section Theory, while the raw images does not. We chose to work on the raw data and keep those values untouched. Each image has resolution 400x400 with 2 channels each, one containing the scale without the melanin (fluorescence channel) and the other one containing the melanin without the scale(photoacoustic channel).

This is really practical when it comes to machine learning procedures, because we already have seperated the information that will be fed into any given model depending on what we want to learn. In other microscopy techniques, in order to work separately on melanin cells and fish scales an extra image segmentation step would be required.

Each pixel refers to a real area of 20μmx20μm for the categories of red porgy and sea bream but for sea bass it was 15μmx15μm. The average size of sea bass scales is smaller and the difference in pixel size was chosen so that the whole scale could fit in an image for every category.

3.1.3.2) Noise clearing

We constructed an object detection algorithm in order to get the total area of the scale using unsupervised learning and computer vision algorithms. Thus far, in melanin observation experiments it is the first time that the surface of a scale is measured via an unsupervised method and not by hand. Similar approaches on image segmentation using k means, and many variations of it, are very common in image processing. [38,39]

The steps of the image segmentation algorithms are the following:

First we blur the image with a median filter of size 11x11 pixels. This is done in order to blend the values of neighbouring pixels. The median blur also helps in minimizing the impact of unwanted noise pixels with really high values. Those signals come from potential objects that were not cleared properly during the sample preparation process. We then used a gamma correction factor on the whole image. Gamma correction factor has the effect of making values below 1 larger and values above 1 smaller. This will separate more clearly background noise from signals of interest. Using an unsupervised learning method, the k-means method, we clustered the pixels of each image into two clusters and appointed values to each cluster 0 and 255 respectively. Finally we used a contour finding computer vision algorithm that detects the edges of every object on an image, connects them and then fills the empty spaces inside the

object it created. The filling value is again 255. This way we created a binary mask representing the fish scale. By extracting the coordinates of each pixel belonging to the mask we can now go to the actual image, because the mask and the original image are in full spatial coherence, and extract the values of the fluorescence signal on the scale, or even calculate directly the surface of the mask. We noticed that all the masks had actually 1-2% smaller surfaces than the surfaces someone, manually, circled by hand, trying to make as few errors as possible. This is completely logical if we consider that someone who wants not to lose any valuable information is going to circle excess area outside the scale in order not to miss any valuable information.

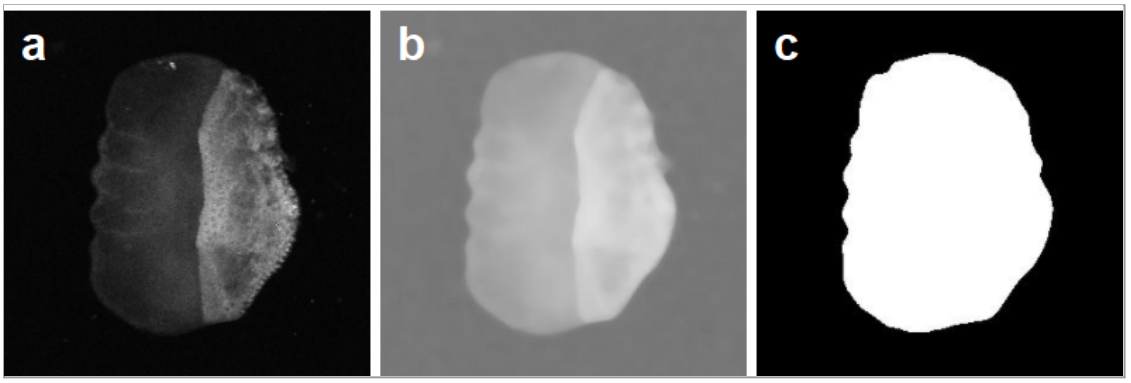


Figure 13: Autofluorescence of sea bream scale. a) The original image, b) The image after median blur and gamma correction, c) The resulting mask after k-means and contour filling.

To separate melanin signals from noise on the melanin channel, we used k-means on all the melanin signals from all the images flattened. This resulted in an 1-D array, representing the melanin space, which the model was instructed to best separate the data on that space into

two clusters, around two centroids.

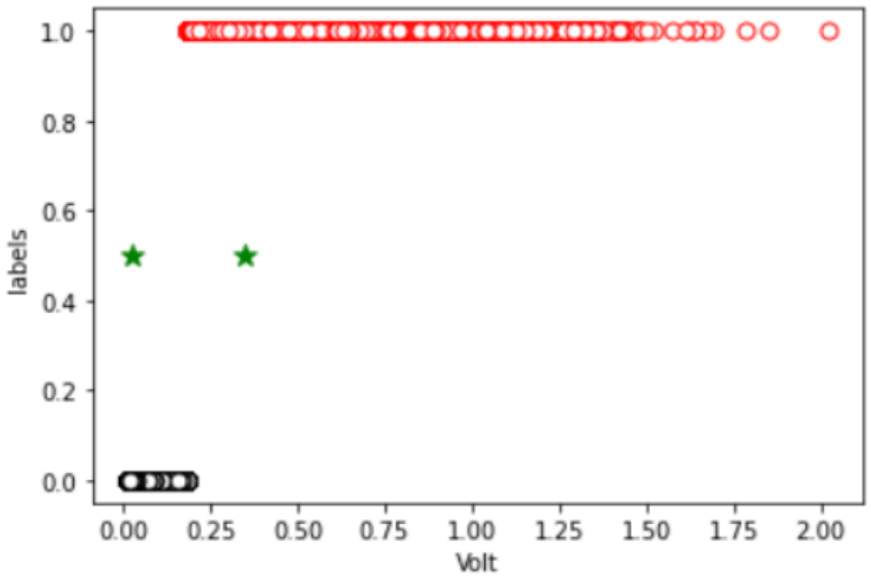


Figure 14: K Means on the melanin. With black we can see the pixel values of the 0th cluster, with red the pixel values that belong to the 1st cluster, with green starts the cluster centers of black and red points.

The threshold value that separates the noise from information was denoted as:

$$V_t = \frac{(c_1 + c_2)}{2}$$

where V_t is the threshold in Volts, $\boldsymbol{c_1}$ and $\boldsymbol{c_2}$ the optimal cluster centers k means returned. The predicted threshold was calculated as 0.19 which distorted the images when applied and we discarded that approach. A more empirical value was used as a threshold of 0.15 decided by experience obtained during the dataset creation and the general knowledge of values returned from our configuration in previous experiments.

3.1.3.3) Statistical Analysis

After creating the original dataset and applying the tools described in the previous section to calculate the best describing parameters of a fish scale, we created **Table 1** containing the mean values of each characteristic value for each class. Then we used multiple t-tests to see which parameter has statistically significant differences between classes. All the classes had statistically significant differences with p-value = 0.0001.

Table I

Class	Scale Surface (mm2)	Total melanin signal (V)	Melanin Surface Coverage (mm2)	Melanin Density (V/mm2)	Percentage of coverage (%)
red porgy	17.67 (± 0.36)	86.83 (± 9.63)	0.129 (± 0.014)	4.83 (± 0.5)	0.718 (± 0.071)
sea bream	14.28 (± 0.40)	411.48 (± 42.5)	0.521 (± 0.047)	27.32 (± 2.38)	3.486 (± 0.262)
sea bass	8.91 (± 0.15)	624.36 (± 39.86)	0.737(± 0.042)	71.60 (± 3.57)	8.419 (± 0.486)

We used some box plots to better visualize the contents of Table1 as shown in Figure 15.

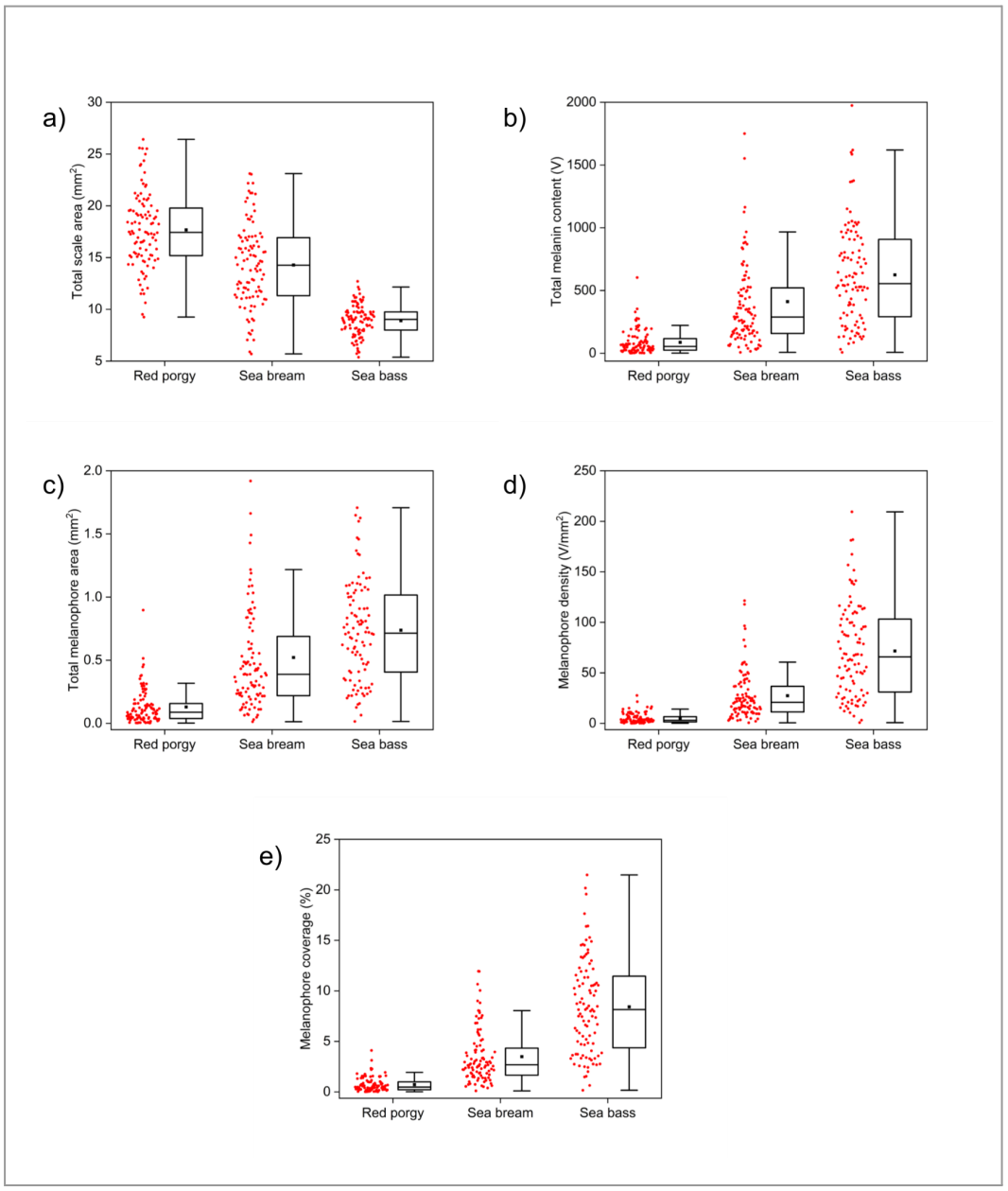


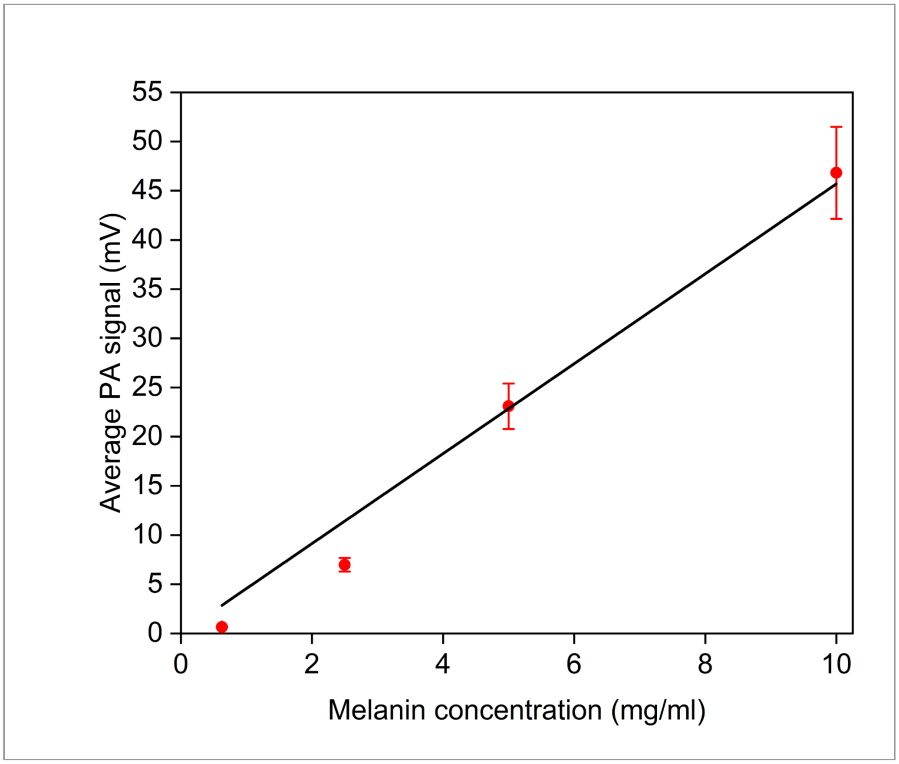
Figure 15: Box plots for each parameter with all the data. a) Total scale area, b) Total melanin content, c) Total melanophore area, d) Melanophore density, e) Melanophore coverage (%)

In Figure 15 present 5 different sets of box plots each for a different feature of Table1 containing a box plot of each fish category. In every plot the red dots show the actual measurements for each class on the specific feature. Figure 15a presents box plots depicting the distribution of scale surfaces for each fish category. The Figure 15b box plots show an index of the total quantity of melanin each scale contains. This index is calculated by adding all the raw values of the photoacoustic signal on an image that exceed the threshold of 0.15V. Box plots Figure 15c shows the distribution of the area melanin cells cover on a fish scale. This is calculated by adding the number of pixels that are spatially inside the borders of the mask found on the fluorescence channel. In Figure 15d we can see the distributions of the density of melanophores on a scale. Finally in the Figure 15e plot we can see the distributions of the percentage of coverage of melanophores on a scale. This is calculated by adding the number of pixels that exceed the threshold in the photoacoustic channel divided by the number of pixels a

mask contains for the same scale. Concluding, the differences of Table 1 are qualitatively confirmed by the boxplots.

3.1.3.4) Volts to Melanin densities

The last analysis tool we wanted to create was a calibration line that transforms the values the configuration measured (Volts) to real values of melanin densities. To achieve that we diluted 4 different concentrations of melanin from a cuttlefish, as this substance is considered the benchmark density of melanin in all biological experiments, into a gel with 2% w/v agarose in it. We then used those mixtures as samples and measured the photoacoustic signal they returned. The following linear graph represents the calibration line:



Graph 1: Calibration line from Volts to melanin concentration

3.1.3.5) Dataset preprocessing for Machine Learning

The 3 categories, red porgy, sea bass and gilt-head bream were labeled on a separate column with 0,1 and 2 respectively. We splitted the original dataset two times to create a train,test and validation subsets. First, we separated 20% of the images, randomly selected, as test images (60 images total). From the remaining 240 images we separated 10% of the images, again randomly selected, as validation images (24 images). The 216 remaining images were augmented by flipping the images up-down and left-right resulting in 648 training images each accompanied by its respective label stored in a different array which was one hot encoded. The augmentation was applied based on the fact that neural networks need a great amount of data in order to be trained properly. Augmenting the dataset helped in this particular case by approximately increasing the final accuracy of our model by 5%. Furthermore, one hot encoding

transforms a single multi-labeled column to multi-single columns that contain only 0 and 1. The number of columns created are equal to the number of labels in the original multi-labeled column. One hot encoding helps significantly when we have to deal with a multi class, categorical labels (in our case are red porgy, sea bream and sea brass). The computer does not understand words and can only work with numbers. We could represent each class with a number e.g. 0,1 and 2 respectively but this creates a hierarchy (by definition) that is not respective to reality (2>1>0). That said, the importance of One Hot Encoding is great. Further augmentation was tested but did not improve our results. The final training set was shuffled once more.

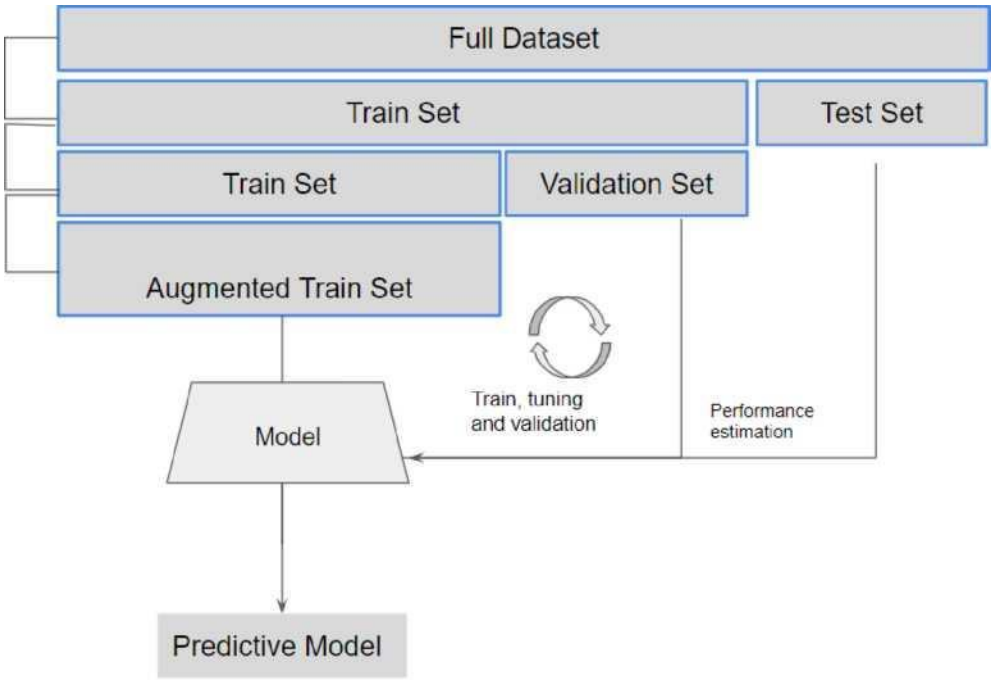


Figure 16: The separation of a dataset to train, validation and test subsets followed by the augmentation step. The model is trained on the augmented train set and validated on the validation set. Finally the performance of the model is tested on the test set. The final train, validation and test sets are not overlapping.

3.1.3.6) Artificial Neural Networks

The differences on how much melanin exists on the different categories of fish scales is clearly distinguishable even by an untrained observer. This is a sign that a neural network could distinguish those classes from the images. From that perspective we created and trained a Convolutional Neural Network that can return the class of a fish scale using an image from the photoacoustic channel as an input.

The model takes as input a single channel 400x400 photoacoustic image and has an output softmax layer with 3 nodes, each node represents one class. The network has 6 convolutional layers. All kernels have sizes of 3x3 pixels and padding applied to them. Every convolutional layer is followed by a MaxPooling layer with kernel size 2x2 to reduce the original size to 12x12

sized images right before the classification part of the model. The only exception is the first convolutional layer which has a striding of (2,2) and a MaxPooling Layer with kernel size of 3x3. The dense part of the model has 2 hidden layers of 100 and 50 nodes respectively leading to a three node softmax output layer. Every node in this network has an activation function of leaky ReLu. We concluded on this specific architecture after many trials and errors.

The loss function minimized at the training process was categorical cross entropy as we are solving a multi class classification problem. The optimizer we used was Adam without restrainments or pre-fixed hyperparameters in order to let it find for itself the best learning rate during the training process.

Finally an early stopping and model checkpoint were used so that the best model during the train was returned before the model overfits. The monitored parameter that would determine the best model was validation loss. These functions also save us a lot of time during the training phase of a model. EarlyStopping, stops the training process of a model if the monitored value does not improve through a number of subsequent epochs. In our case the patience, as it's called, was 20 epochs. ModelCheckpoint keeps the weights of the models' epoch that had the best (smaller in our case) validation loss.

Based on the clear differences between each class we proceeded to train the neural network on the augmented dataset of raw signal obtained from the photoacoustic channel. The model that was described in the previous paragraph was trained on 100 initializations of the datasets to be certain that the way we split the images through the subsets does not affect the efficiency of the model, making it more credible.

The metrics used were the general Accuracy of the model and sub metics for each class, Precision, Recall and F1 for each class independently.

Finally, in Table 2 we present the average values and standard deviations of the metrics over 100 different splits on the train, validation and test sets for each class respectively.

Table2

Labels\Metrics	Precision	Recall	F1
red porgy	**0.89 (± 0.08)**	**0.92(± 0.07)**	**0.90(± 0.05)**
sea bream	**0.8 (± 0.1)**	**0.75(± 0.1)**	**0.76 (± 0.0.09)**
sea bass	**0.87 (± 0.08)**	**0.85 (± 0.1)**	**0.85(± 0.0.07)**
Accuracy	**0.85 (± 0.05)**		

In Table2, the statistical results per class and general accuracy of the model are produced by cross validating our model on 100 different random splits of the original dataset.

2.2) Macro Color Detection with Neural Networks

Coloration patterns can be really versatile throughout fish species but can also differ in the same fish species. The color deferenciations are dependent on many variables. In this part of our work we built a model that can measure the La*b* chromatic parameters at specific positions on the body of a given red porgy. We created a model that could possibly replace a colorimeter, the go to measurement tool when it comes to coloration measurements.

2.2.1) Materials and Methods

2.2.1.1) Dataset

The original dataset for this part of our project consists of 2731 red porgy RGB images. Each image is a photograph taken under controlled environment by a high resolution camera, with 6000x4000 pixels resolution. Each image is represented by triplets of, 12 in total, values. Each triplet contains the color parameters (LAB colorspace) in 4 specific areas of each fish. These values were measured with a colorimeter during the photographing of the fish and are used as target values in our regression problem. The coloration measurements were taken from the same areas of each fish, proportionally to each fish's size as seen in Image17. In the following paragraphs we will refer to these areas as D1,D2,V1,V2 where D is the dorsal area and V the ventral area of a fish.

2.2.1.2) Preprocessing Tools

2.2.1.2.1) Image Segmentation

As a starting point, before building the models in this part of our work, the original dataset had to be preprocessed in order to extract all the valuable (for our task) information of each image and discard all the rest. An image segmentation algorithm was constructed using K means, Computer Vision and colorspace transitions to do so.
In more detail, a preprocessing pipeline was created following the steps:

1) Converted the original RGB channels to three separate, L,a* and b* (LAB color space)
2) Combine the new channels in a non-linear manner that best separates information from background.
3) Used a median blur filter to blend the pixels with each other.
4) Use k means in order to classify the pixels on a given image as fish versus no fish.
5) Find the contours and select the biggest object in the image (find the contour with the biggest area)

6) Calculate the rectangle that contains the biggest contour using the minimum area possible.
7) Find and extract, proportionally to a fish's size, 4 new images containing the D1,D2,V1 ,V2 areas.

Since the areas cropped from an image are proportional to a fish's size, the cropped parts from each image do not have identical dimensions with each other. Smaller fishes have smaller cropped areas while larger fish have larger cropped areas. An additional preprocessing step had to be added, to make all the input images identical in size. It is always easier to downscale a large image to a smaller one than upscale a smaller one to a larger image. In that regard we found the size distribution of all the cropped images through an histogram and discarded the smaller outliers. Then, re-cropped all the larger images into size equal to the threshold that seperated the outliers from the rest of the images.

2.2.1.2.2) Dimensionality Reduction

One of the main difficulties we faced in this part of our work was the large dimensions of both the original images and the cropped sub-images. Dimensionality reduction was applied to the cropped images using principal component analysis (PCA). Every cropped image was flattened and it was inserted to an array as a row. A new matrix was constructed with dimensions $m \times n$ where $m = width * height * N_{channels}$ and $n = number\ of\ images\ in\ dataset$. A part of this matrix was given as an input to PCA. By looking at the cumulative explained variance of the principal components of our new matrix we found the number of components that maintained 0.95 of the total variance. The PCA algorithm was then retrained on the train set in order to transform images into a 1-D array with length equal to the number of principal components that preserve 0.95 of the total variance. Then used to transform the validation and test set as well.

3.2.1.3) Models

We are trying to build a regression model that can predict the three chromatic parameters on areas of a fish. In that regard three different models were constructed, one for each chromatic parameter that takes as an input an image and returns one value (single input/ single output models) All the procedures were conducted three-fold, one for each model, but are identical in all regards.

3.2.1.3.1) Benchmark Model

A benchmark model is the simplest model that can calculate the value in question using the available input. This is constructed in order to have a basic approach and thus a baseline value

in our resulting metrics.

The benchmark model we created here is the mean value of the pixel values an image has after it's transformed to the respective channel in the LAB colorspace. For example, the benchmark model for the calculation of Lightness on the D1 area of fish is the model that takes as an input the transformed, from RGB to L, D1 image and calculates the mean value of it's pixel. It is equivalent with the mean Lightness of that area.

3.2.1.3.2) Fully Connected Neural Network

A Fully Connected Neural Network (FNN) that takes as an input an array with length k and returns a single value as an output was trained and tested on our data. The input array contains the ith values of the k principal components that maintain 0.95 of the original datasets variance. After the input layer of size k follows three hidden layers with 500,250 and 100 nodes respectively and an output layer with a single node. The hidden layers and the output layers have a Relu activation function. The optimization algorithm used was Adam and the loss function was MSE. The models were trained for 100 epochs and a batch size equal to 64.

3.2.1.3.3) Convolutional Neural Network

A Convolutional Neural Network that takes as an input an RGB image and returns a single value was trained and tested on our data. The model has 5 convolutional and max pooling layers. The convolutional layers starting from 8 kernels at the first layer and resulting to 128 kernels in the final convolutional layer incrementing by a multiplication of 2 per layer have kernel dimensions of 3x3 and create a final feature map with elements of size 8x8 using a MaxPooling layer, of dimensions of 2x2 after each convolutional layer. The activation function is LeakyRelu. After the convolutional part of the neural network the final 128 feature maps are flattened and are used as an input to a dense neural network with an input layer, without activation function, of size 8192 and followed by 3 hidden layers with 250,100 and 50 nodes respectively. Their activation function is again LeakyRelu and the output layer has 1 single neuron. The loss function is MSE and the optimization algorithm used is Adam() with free hyper parameters. The neural networks are trained for a maximum of 40 epochs using EarlyStopping with patience 10 and Model Checkpoint.

The models described in this section are a product of trial and error on different architectures and hyperparameters. The results that the best architectures produce are presented in the next chapters of our work.

3.2.2) Results

3.2.2.1) Preprocessing

In Figure 17 we present one original image from our dataset as well as the 4 specific areas of interest, D1,D2,V1,V2.

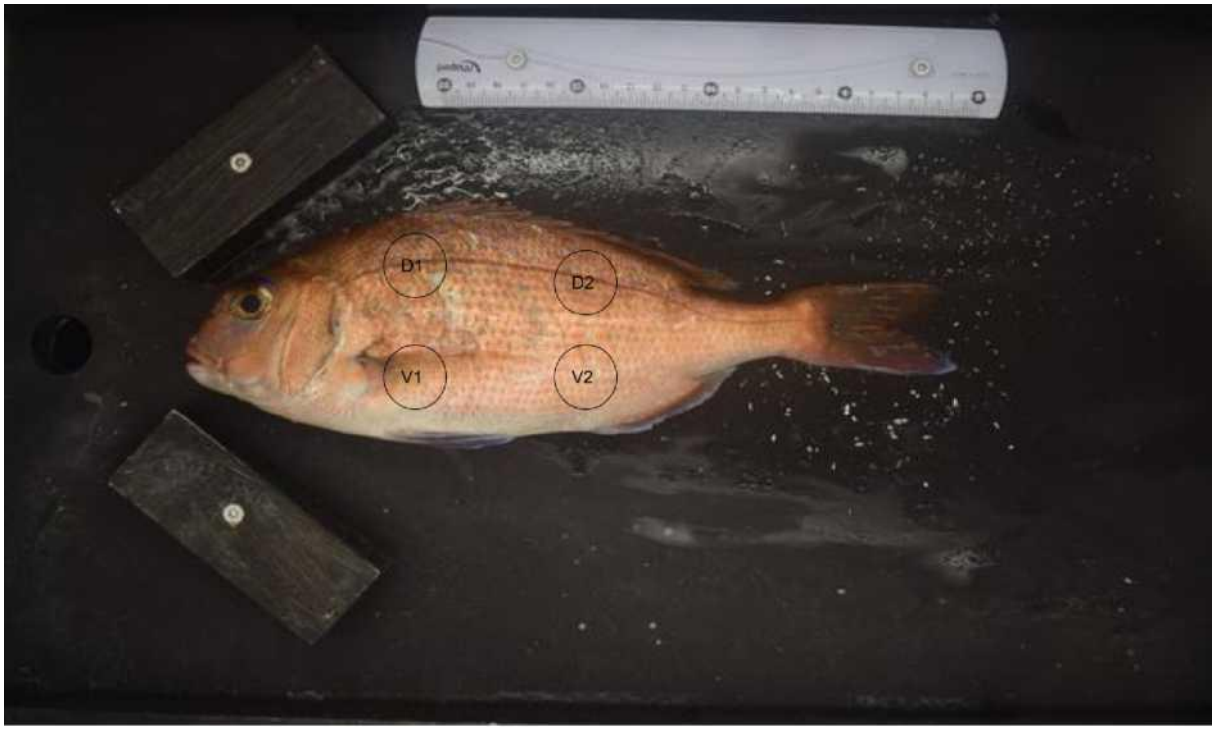


Figure 17: An original image from the dataset. D1 is the dorsal-left area of a color measurement, D2 is the dorsal-right and V1,V2 are the respective areas in the ventral area.

Through the skimage library we can easily extract from the original image the three channels in LAB colorspace. Although L,a* and b* channels carry different information in them as we can see in Figure 2 we want to find a combination of them that only carries information about the fish and not the background. Information gain here is achieved when there is a clear differentiation in pixel intensities on the fish area versus the background. This can be achieved by a non linear combination of L and b* channels shown in Figure 18.d. The non-linear combination used was:

$$z = b^{*2} - L$$

where z is the result of the combination of the channels, b* and L are the respective channels after the transformation from RGB to LAB.

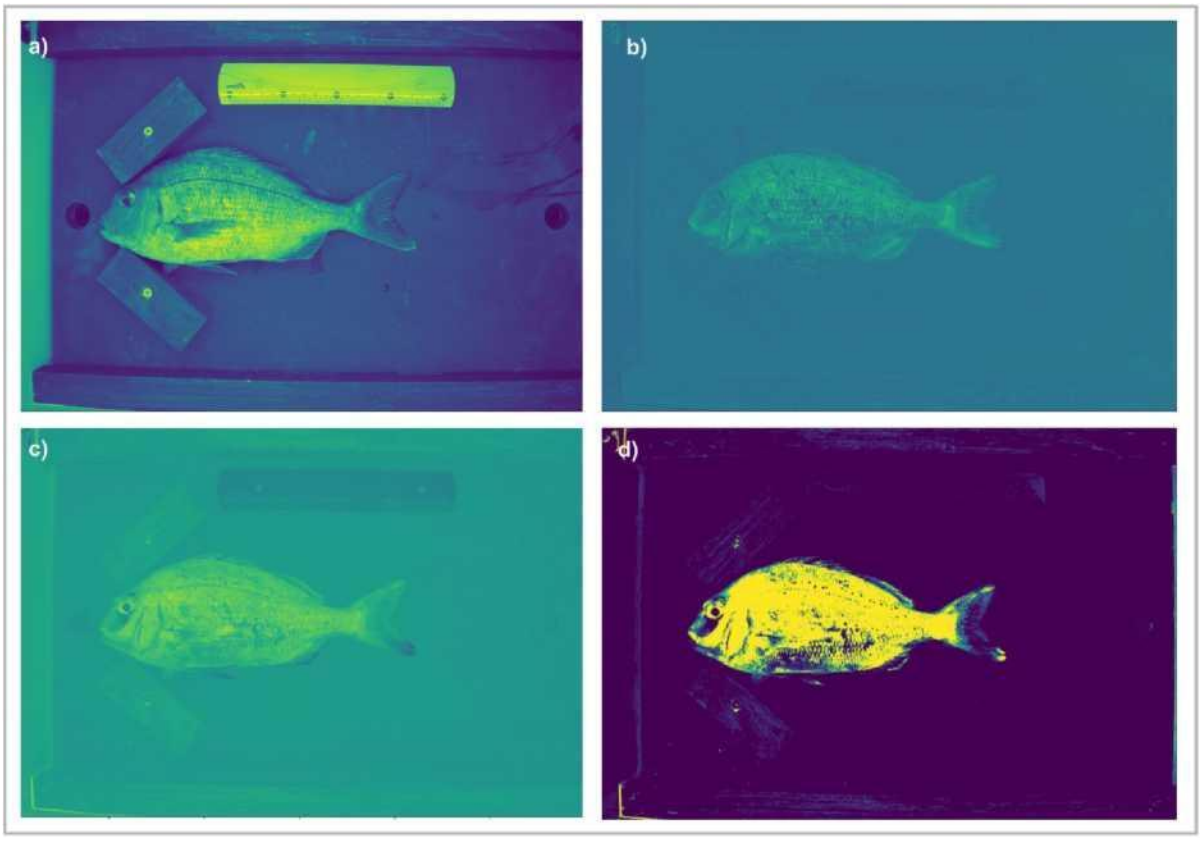


Figure 18: a) L channel on a random fish image, b) a* channel on a random fish image, c) b* channel of a random fish image d) The z custom channel. A non-linear combination of La*b* channels ($z = b^{*2} - L$)

In order to make sure that z can be displayed as an image we used the clip function, making all values above 255 and below 0 equal to the threshold they're exceeding. A median blur filter was passed on z and then a k-means algorithm classified the pixel values of each image into three clusters. The cluster having the minimum value as a cluster center were regarded as the background and the other two as the fish. All the pixels belonging to the background were appointed 0 values and all the pixels belonging to the fish were appointed 255 values, making a binary image. In the now binary image we used a computer vision algorithm from the OpenCV library to find the contours of the fish and then crop the rectangle that contains the contours of the fish while occupying the least area possible. This rectangle is called ROI (Region of interest) of the image. The ROI function that OpenCV provides, finds the ROI and returns two tuples. One with the top left coordinates (x,y) of the rectangle and the other one with its width and height (w,h). The width and height of ROI is proportional to the size of the

fish and the x,y coordinates tells us the position of the fish in the image. With those two tuples we are now in position to create a grid on a cropped ROI and then find the right coordinates based on the grid to robustly extract the desired parts of a fish's skin. The optimal x_step and y_step chosen for creation of such a grid was width/12 and length/5 for x and y axis respectively. The specific spots were then chosen as follows:

The corners of the Ai square area are denoted as $(x_{1,Ai}, y_{1,Ai}), (x_{2,Ai}, y_{2,Ai})$ and $(x_{3,Ai}, y_{3,Ai}), (x_{4,Ai}, y_{4,Ai})$ with the first two coordinated being of the top corners and the last two the bottom corners, going from left to right. As for the Ai notation, A could be either D (dorsal) or V (ventral) while the i index takes either 1 or 2 denoting the left or right area of the D,V areas.
Furthermore, $x_{step} = w/12$ and $y_{step} = height/5$ where $w = ROIs\ width,\ h = ROIs\ height.$
Finally, $(x, y) = (x\ coordinate, y\ coordinate)\ of\ the\ top\ left\ ROI\ corner$ and also, in an image imported by matplotlib, the y axis is incrementing from top to bottom while the x axis is incrementing from left to right. Bigger y values are towards the bottom of an image while bigger x values are towards the right. With that we can define the corners of:

<u>D1 as:</u>

$(x_{1,D1}, y_{1,D1}) = (x + 3* x_{step}, y + y_{step})$ $(x_{2,D1}, y_{2,D1}) = (x_{1,D1} + x_{step}, y_{1,D1})$

$(x_{3,D1}, y_{3,D1}) = (x_{1,D1}, y_{1,D1} + x_{step})$ $(x_{4,D1}, y_{4,D1}) = (x_{2,D1}, y_{3,D1})$

<u>D2 as:</u>

$(x_{1,D2}, y_{1,D2}) = (x_{1,D1} + 3 * x_{step}, y_{1,D1} + y_{step}/2)$ $(x_{2,D2}, y_{2,D2}) = (x_{1,D2} + x_{step}, y_{1,D2})$

$(x_{3,D2}, y_{3,D2}) = (x_{1,D2}, y_{1,D2} + x_{step})$ $(x_{4,D2}, y_{4,D2}) = (x_{2,D2}, y_{3,D2})$

<u>V1 as:</u>

$(x_{1,V1}, y_{1,V1}) = (x_{1,D1}, y_{1,D1} + 2 * x_{step} - y_{step}/2)$ $(x_{2,V1}, y_{2,V1}) = (x_{1,V1} + x_{step}, y_{1,V1})$

$(x_{3,V1}, y_{3,V1}) = (x_{1,V1}, y_{1,V1} + x_{step})$ $(x_{4,V1}, y_{4,V1}) = (x_{2,V1}, y_{3,V1})$

<u>V2 as:</u>

$(x_{1,V2}, y_{1,V2}) = (x_{1,D2}, y_{1,D2} + 2 * x_{step} - y_{step}/2)$ $(x_{2,V2}, y_{2,V2}) = (x_{1,V2} + x_{step}, y_{1,V2})$

$(x_{3,V2}, y_{3,V2}) = (x_{1,V2}, y_{1,V2} + x_{step})$ $(x_{4,V2}, y_{4,V2}) = (x_{2,V2}, y_{3,V2})$

Finally, we crop the areas and save them in a seperate folder in order to use them for the training of our models. In Figure 19 we present the results of each extraction step.

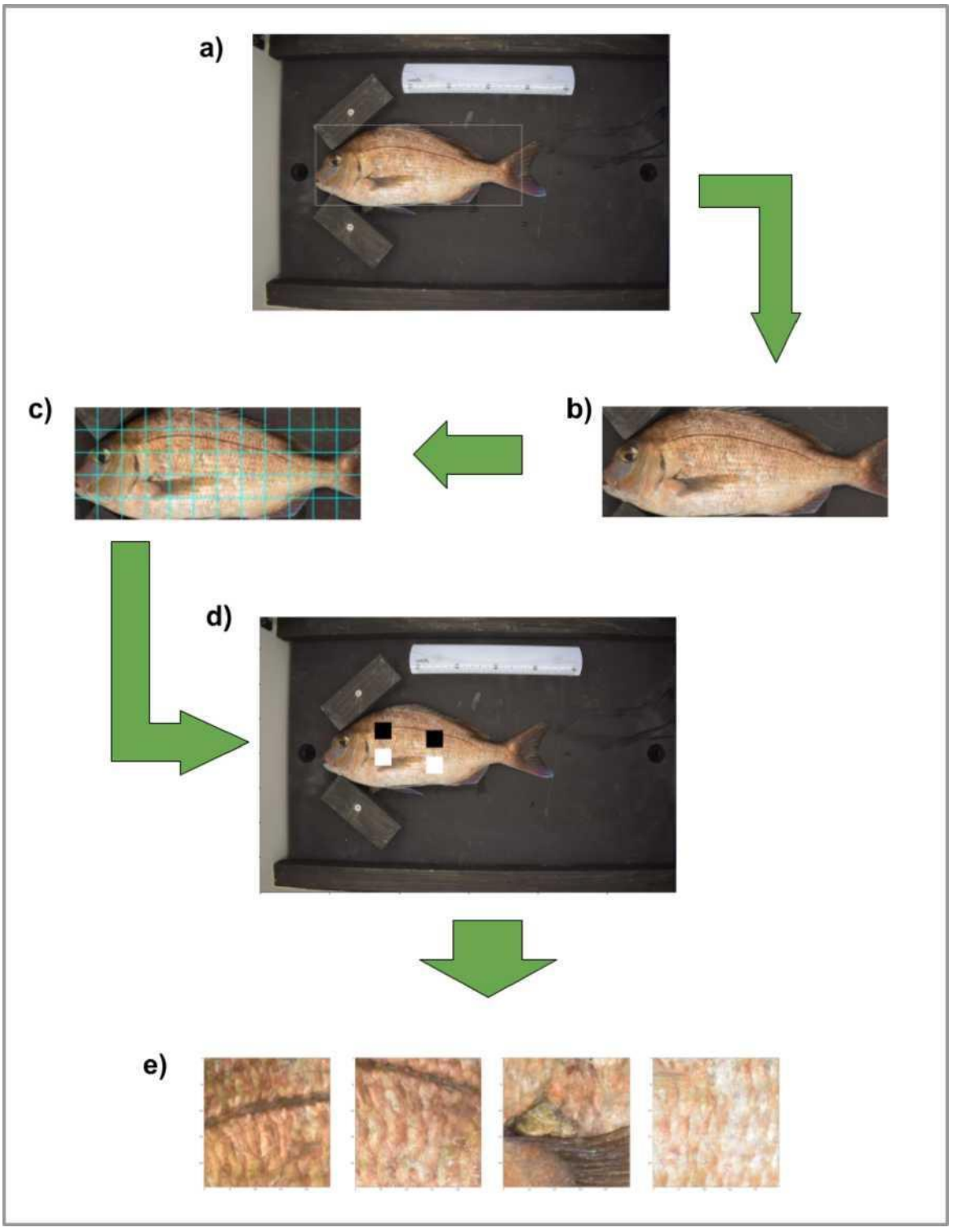
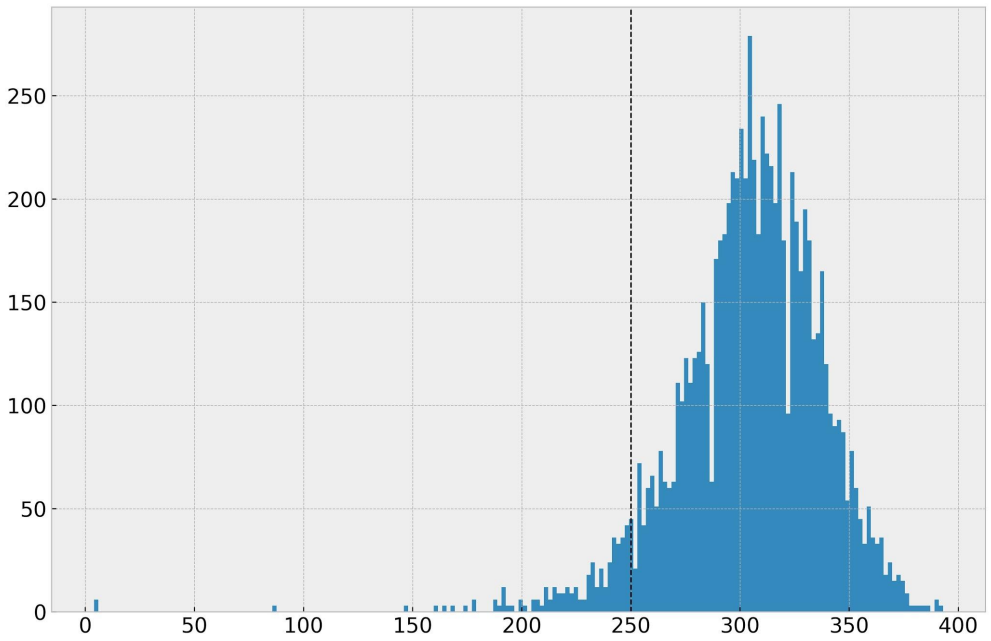


Figure 19: a) A ROI found on the example image b) The cropped ROI c) The grid on the cropped ROI d)With black squares we can see the dorsal areas automatically selected and in white the ventral areas automatically selected e) The cropped areas D1,D2,V1,V2, from left to right, automatically cropped.

3.2.2.2) Final Dataset

Now we have in our disposal 10.924 images. After discarding the V1 parts because instead of the skin they contain the fin of the fish, our dataset contains 8193 images with varying sizes. Each quadruple set of those new images are saved with names of the fish and the part it depicts for organizational purposes. (e.g. DSC_0001D1.JPG). As we mentioned before, the main problem that occurred following this procedure is that the size of the areas cropped from each fish depends on the actual size of a fish. In order to build a Convolutional Neural network the size of the input must be fixed to a specific value. A histogram with all the sizes of the dataset images were calculated and a threshold to a specific size in order to discard all the images with smaller size than the threshold. This is done in order to not lose useful information by downsampling the majority of our images due to some outliers. The larger end outliers are not as important because we can always crop a part from them and downsample. The histogram and the threshold used in the end at 250 x 250 pixels are shown in Graph2.



Graph2: On the X axis are the sizes of the cropped images from each fish. Black dashed line at 250. All the images with size smaller than 250 were discarded. On the y axis are the number of images that have a specific size.

We ended up with 7.737 images after clearing the outliers and discarding the V1 category. We crop the remaining images in our dataset to (250,250) size and 3 channels in the RGB

colorspace. We are now in position to build models in order to predict the color parameters of a fish in 4 different areas.

Finally, we splitted the remaining dataset images into 3 sub groups, train set, validation set and test set with ratio 90/10. In more detail we split the dataset of ~7000 images with ratio 90/10 to train and validation sets and then we subsplitted the train set into a new train set and a test set again with 90/10 ratio. The validation set is used for the in-train validation of the training procedure while the test set is left unseen until the final evaluation of the model.

3.2.2.3) Results

3.2.2.3.1) Fully Connected Neural Network after Dimensionality Reduction

A PCA algorithm was trained on the train set as explained in the Materials and Methods section in order to reduce our dataset in size but preserve the total variance as much as possible. To find the optimal number of principal components needed for that transformation we plot the cumulative explained variance versus the number of components. The cumulative explained variance is the sum of the variance a number of principal components have divided by the total variance. A rule of thumb for the selection of the number of principal components we are going to keep is to find where the curve has the steepest slope. In our case that is around 961 principal components.

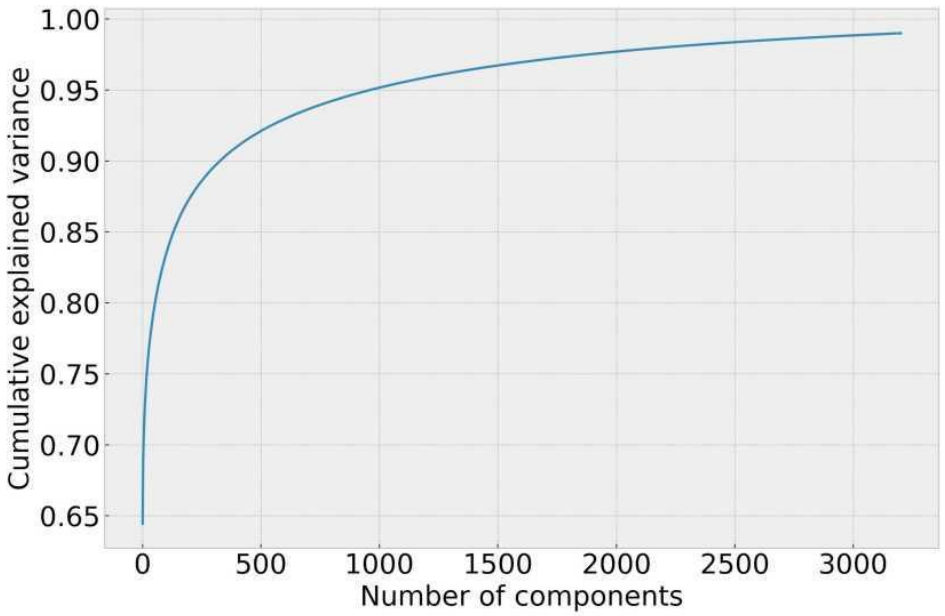


Figure 20: Cumulative explained variance of the principal components when PCA is trained on the whole

train set.

Setting 961 as a hyperparameter we retrain PCA on the dataset now asking it to project each instance of our full dataset to only 961 dimensions. After fitting the algorithm on the train set, we use it to transform the validation and test set. The DNN was trained on the transformed train set, updating its weights by its performance on the validation set and was finally tested on the test set. In Table3 we present the results of the model versus the Benchmark model.

Table3

Model\MSE	L	a*	b*
Benchmark	515.37	58.71	320.530
DNN	75.41	4.72	13.20

In Figure 21 a.1),b.1) and c.1) we can see in red dots the predicted values versus the true values for each colorspace value. A model with MSE = 0 would have red dots only along the y=x straight line depicted in blue. Figure 21 a.2),b.2) and c.2) shows the predicted values obtained by the FNN,the benchmark values and the true values per image for a random set of 30 test images. In both cases we observe that the FNN produces more accurate results and are in accordance with the results presented in Table3.

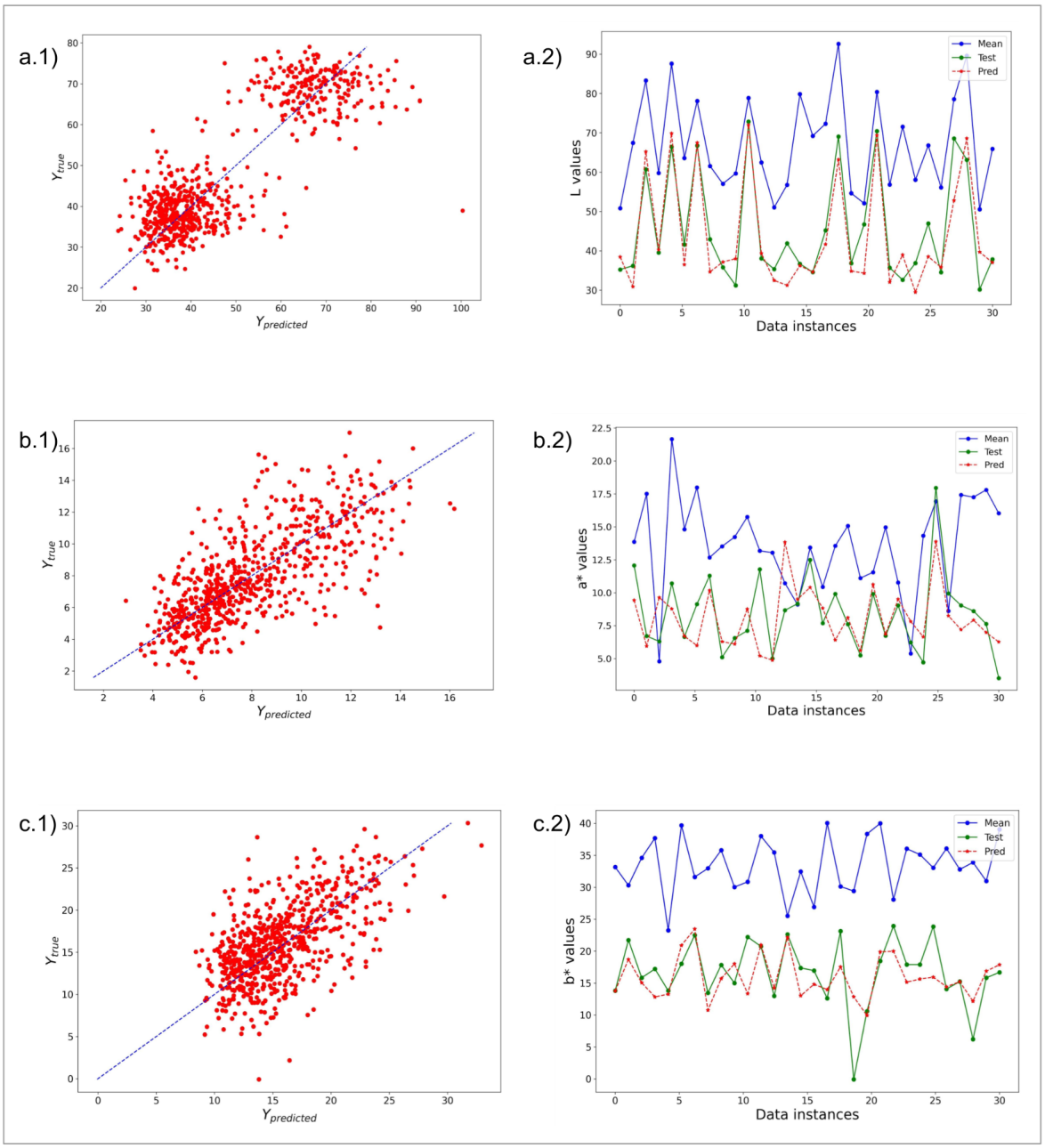


Figure 21: (a.1) Scatter plot of true versus predicted L values for all the test images of red porgy fish. The dashed blue line is the optimal result that could be achieved in this visualization. (a.2) Scatter plot of the predicted, true and benchmark model values for 30 random test images of red porgy. Blue dots show the benchmark results, green dots show the true values and red stars the predicted values from the DNN model. (b.1) and (c.1) show the respective results for a* and b* values respectively, as explained for (a.1) . (b.2) and (b.2) show the respective results for a* and b* channels respectively, as explained for (a.2).

Concluding we have now 3 trained models, each measuring a different value. All the models are producing better results than the benchmark model. Various architectures and hyper parameters were tested for each model and the presented above are the best versions achieved.

3.2.2.3.2) Convolutional Neural Network (CNN)

Training the CNN models is more straightforward. Each takes as an input an RGB image and calculates one of the three different values L,a* and b* in question. In Table4 we present the results of the FNN models for each value versus the Benchmark Model and also the CNN results.

Table4

Model\MSE	L	a*	b*
Benchmark	495.66	61.87	322.76
DNN	75.41	4.72	13.20
CNN	26.43	3.2	9.23

Again, we visualized our results, as we did with FNN. In Figure 22 we present a visualization of the predicted values of the CNN on the test set, similar to those of the FNN. Figure 22 a.1),b.1) and c.1) shows in red dots the predicted values versus the true values for each colorspace value. Figure 22 a.2),b.2) and c.3) shows the predicted values from CNN, the benchmark values and the true values per image for a random set of 30 test images. Again, the results are in accordance with the results presented in Table4.

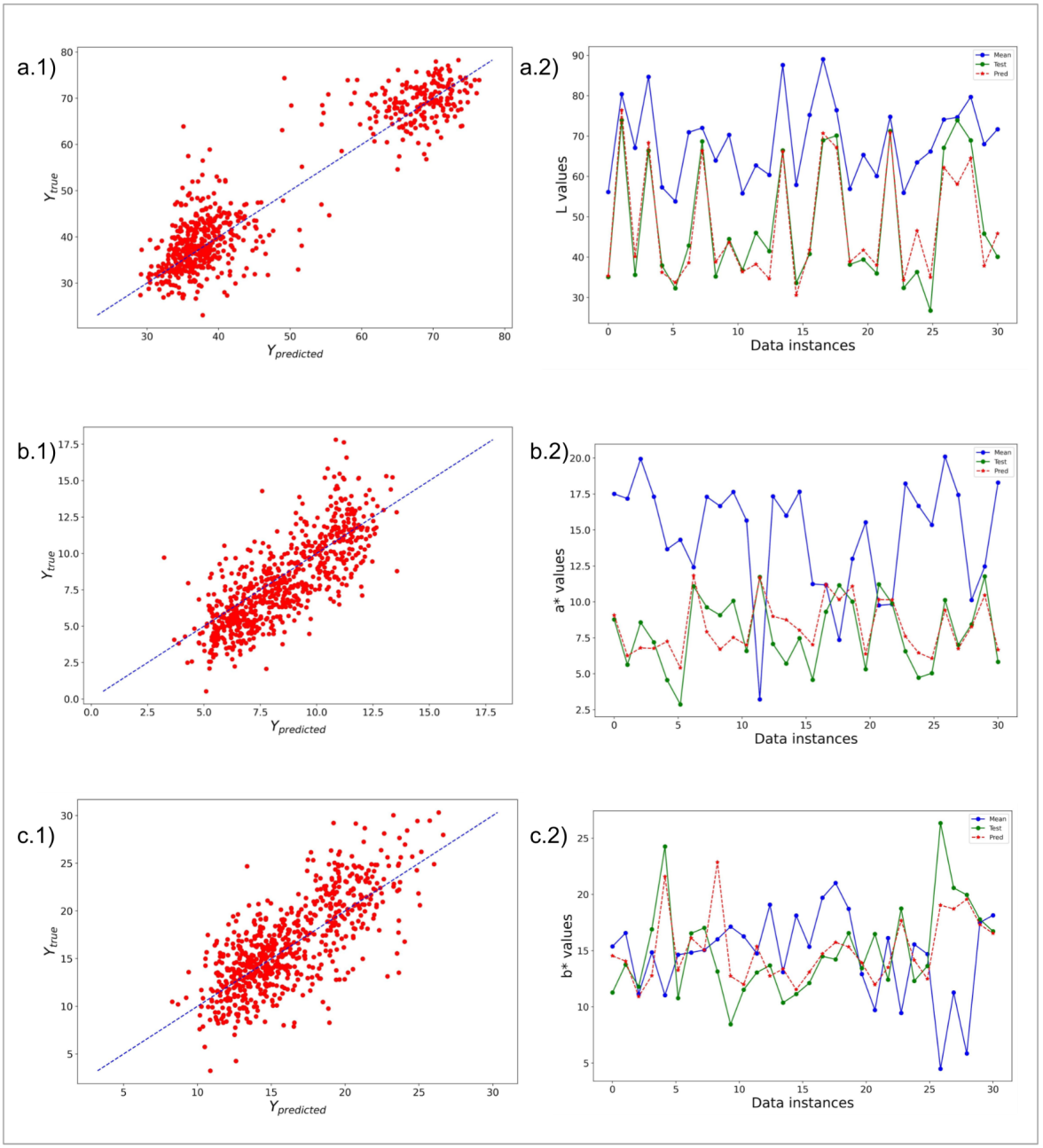


Figure 22: (a.1) Scatter plot of true versus predicted L values for all the test images of red porgy fish. The dashed blue line is the optimal result that could be achieved in this visualization. (a.2) Scatter plot of the predicted, true and benchmark model values for 30 random test images of red porgy. Blue dots show the benchmark results, green dots show the true values and red stars the predicted values from the DNN model. (b.1) and (c.1) show the respective results for a* and b* values respectively, as explained for (a.1) . (b.2) and (b.2) show the same graphs for a* and b* channels respectively, as explained for (a.2).

4) Discussion and Conclusions

Fish coloration patterns have been a study subject in many research areas for their various functions. Photoprotection, temperature control, behavior and economical qualities (in the case of farmed fish) are some of the many aspects fish coloration patterns have been studied for. As a result, fish coloration patterns can be a fundamental feature of a more complex welfare and economic index. In this work we created robust tools for biologists and physicists with practical applications on the field (fish farms) that will hopefully further the study on the field or make their work more efficient. Our work focuses on two levels separately. The first part of our work revolves around melanin analysis on a microscopic level, studying the melanophores on a fish scale while the second part of our work is approached in more a straightforward manner, by analysing high resolution photographs and calculating the chromatic parameters of specific areas on red porgy skin.

Microscopic Analysis with Hybrid Microscopy and Machine Learning

On a microscopic level, chromatic parameters of a fish can be explained by studying the chromatophores on their skin and scales. To do so, we propose a novel approach, the use of hybrid microscopy. A hybrid photoacoustic and fluorescence microscope was used to image the melanin on fish scales taken from three fish categories, red porgy, sea bream and sea bass. As shown in Figure 5 our configurations' output is in spatial coherence with the images a biologist could acquire from a stereoscope up until now. The difference in intensity on some parts is explained by the fact that the hybrid microscopy method will give us a lot more information on a scanned sample than looking at it through a stereoscope. Probably the most useful modality of our configuration is its ability to return an image with two channels per measurement, each channel containing different information, complementary with each other. In addition, the photoacoustic channel gives us, robustly, more information than a stereoscope. That is due to the fact that the photoacoustic signal is proportional to the concentration of the melanin that is radiated per pixel. Due to those features of our hybrid configuration we are able to measure different quantities on fish scales, such as the local melanophore concentration on a scale, the percentage of coverage of the total melanin on a scale etc that were not easy to be measured before. Finally, the results of our configuration are not highly depended on external conditions in order to be accurate. (e.g. external light conditions when using a stereoscope). On Figure 6 we can see the complementary features of our configuration, as well as, obvious differences between the three different fish categories we worked on. Examining Figure 6 (b), (e) ,(h). red porgy scales return the least signal, close to zero. Sea bream has an intermediate signal output which is dispersed across the base of the scale. Sea bass has the most signal output

which is still towards the base of the scale but not that dispersed across it. Later in the discussion we will also describe the precise quantified information our configuration provides.

Proceeding, we measured 300 fish scales evenly separated per class. This was done first and foremost to prove the intra class statistical differences but also to build a robust toolset for their analysis in general. We propose 5 parameters that statistically differ on the three fish categories. With the use of computer vision and unsupervised machine learning the analysis is done robustly with no human interaction required. As of now, the area of a fish scale was done manually with a large human error by circling the area around a scale in ImageJ. K means and computer vision was used as image segmentation tools to determine the surface of a fish scale returning optimal results. As for the melanin, an empirical threshold on the outputs of our configuration was used in order to clear the noise. K means was also used on the melanin channels to determine a mathematical threshold with no success. The threshold K means returned on that application distorted the images when applied and was not used for the analysis. It is valid to point out that k-means is not the most “intelligent” algorithm when it comes to clustering and is safe to state that another unsupervised algorithm could result in a proper mathematical threshold value. Another, more sophisticated clustering algorithm could solve that problem and it is in the sphere of future goals.

A statistical analysis was conducted on the information we acquired from the 300 fish scales. As shown in Table1 and Figure 15, there are differences between each class and in fact all the features have statistical important differences with each other after multiple t-tests were conducted for each class pair. The p-value on which the classes have statistical important differences is 0.0001. It is important to mention that being able to extract correct (as the statistical analysis states) information from single fish scales is a novelty by itself.

The calibration line acquired from measuring different concentrations of cuttlefish melanin, diluted in agarose in order to create scanning samples, was an attempt to transform the values of our configuration (Volts) to milligrams of melanin per pixel and also to prove that the photoacoustic signal is indeed proportional to the concentration of the radiated melanin. Although the photoacoustic signal has a positive, linear relationship, with the melanin concentration, it is a difficult task to transform the concentrations of melanin to absolute mg of melanin. To do so, we have to know the exact volume from which the photoacoustic signal came from. This volume is not easily defined as the signal comes from various axial levels of the sample.

An Artificial Neural Network was created and trained in order to automatically classify a given image to its respective class. The results were above 80% on all metrics, except F1 score which was above 70%, as shownin Table1. From them, two discussion points emerge. First and foremost, its immediate application is that it can classify a fish by its scale. Its utility value can be seen in scenarios where we want to determine if a fish has its supposed qualities on its scales. If this model statistically misclassifies scales from a fish batch of red porgys as sea bream then this is a good signal that something is not right with that fish batch. A second point

that can be made for our model is that although we can not be certain that the neural network is classifying the data only on the melanin levels because although it is trained only on the melanin channels the spatial difference on the scale shape and size is depicted on the melanin channel as well, we are certain that it can accurately find and classify differences on raw measurements obtained by our configuration. This opens up a road to future dataset creations in order to build really useful tools when it comes to melanin on scales.

The model's results had some intricacies as well. During the hyperparameter optimization a confusion matrix gave us great insights as for its classification capabilities. Their results denoted clearly that the model is confused mostly on the separation of the intermediate class (sea bream) from the other two and vice versa. The Accuracy of a binary classification between labels red porgus and sea bass, if the same architecture was trained only on images of those two classes, is 100%. It is not uncommon that the intermediate labels in a machine learning project are more hard to work with but further investigations are required in order to determine why exactly this is the case in our work.

Concluding on this part of the discussion we will refer to some future goals on the subject. The hybrid configuration could be improved in order to obtain images at a greater speed and resolution. Greater resolutions could be achieved with our current configuration but the time of a measurement is too long and can not be used for a dataset creation. Galvanometric mirrors could be used to improve scanning speed and thus resolve that problem. Furthermore, multi-spectral imaging could be achieved on both the fluorescence as well as the photoacoustic modality of our configuration. With that we could image other chromatophores in addition to melanophores. Finally, the algorithmic part of our work could be useful in fast image analysis in in-vivo measurements. If a portable measuring device was created using a similar hybrid microscope, faster means of analysis would be necessary in order to not only return the images of fish scales but also the quantifying values regarding fish scales and their melanin levels. Finally, many machine learning projects could be implemented on the correct datasets to further this work. For example, a model that will be trained on scales of the same fish category with different melanin levels could produce an anomaly detection model that would notify a researcher of a problematic fish.

Macroscopic Color Regression

On a macroscopic level, chromatic parameters on a fish are most commonly depicted by the parameters of various color spaces. The most common color space is probably RGB as it is the default colorspace for many image formats like PNG, JPEG etc. In fish coloration research bibliography though, when the color is measured by digital means it is best described by a nonlinear transformation of RGB, that of LAB color space. The ground truth of color measurement is the LAB parameters a colorimeter will return from a surface. Our red porgy's (also known as Japanese sea bream with the “inspired” Greek translation «βραχύπτερο φαγκρί») photograph dataset has as dependent variables, measurements of a colorimeter on

3 specific areas per fish. We created a model that could replace the colorimeter for red porgy measurements.

The first step of our work entails the object detection on the images, in order to find and extract a fish from the background image robustly as well as to automatically navigate and crop 4 specific areas from its body. The main drawback of our procedures is that our automatic object detection algorithm is rotational sensitive. Most of the fish in our dataset had the same positional orientation in an image (head to the left and almost horizontal body orientation). Given that, even though the object detection part of the algorithm has the same accuracy no matter the orientation, the proportional navigation on the fish's body does not. If a fish was positioned in an image such that its head is looking upwards then the D1, D2, V1, V2 that our algorithm would crop from the image would be wrong. That said, it is not a difficult task to generalize in such a manner and is the top priority in our future work ideas.

The D1, D2, V1, V2 areas have size proportional to the original size of the fish in each picture. The models though, can only be trained on images identical in size. For that reason we discarded all the images with size smaller than (250 x 250).

Furthermore, as we can observe from Figure 19,e) the V1 part of the final results of our preprocessing pipelines is not good for our models, mainly because the fin is covering the area that a human would measure on with a colorimeter. For that reason all the V1 images were discarded from the dataset the final model was trained on.

With the remaining images, we trained 2 models. A Fully Connected Neural Network (FNN) and a Convolutional Neural Network (CNN). In order to decrease the number of weights the FNN had to work with, we added as a pre-train step a principal component analysis model. With that, we found and transformed the images into vectors that have the minimum length (dimensions) that still preserves 95% of the variance our original dataset had. The number of principal components that preserve 95% of the variance is 961.

The results of the models are presented in Table 3 and Table 4. The Convolutional Neural Networks are exceptionally better than the Fully Connected Neural Networks, although both models test with errors of a degree of magnitude lower than the benchmark model. The regressions percentage error for CNN is around 5-10% while for the DNN is approximately 10-15%.

In Figure 21 and Figure 22 we can see an optical representation of the results Tables 3 and 4 respectively. There, we can see that the red dots in the scatter plots in a.1),b.1),c.1) of Figure 21 and 22, are trying to match the y=x dashed line. A model with MSE = 0 would have produced an image with all the red dots lying on the blue dashed line. In a.2),b.2),c.2) of Figure 21 and 22 show how well each specific prediction matches the true value. The models do an exceptionally better job at that, with CNN surpassing the DNN. Furthermore in a.1) in both Figure 21 and 22, we can see 2 clusters of red dots which shows us that indeed the dorsal and ventral area have differences in brightness.

The training graphs with validation and train errors per epoch, even though not provided in this work as we believe that doing so would be divergent from our work's purpose, show a

smooth training with minimal fluctuations in validation and training error differences from epoch to epoch. Overfitting, if existent, is avoided with the use of Model Checkpoint and Early Stopping.

Concluding on the discussion of the macroscopic approach, it is important to mention some applications of our work and possible future goals. An important application of such a neural network is that makes the monitoring of color in fish farms more effortless. It is as well as close as possible (thus far) to an actual colorimeter measurement. As a future goal an in-vivo, underwater monitoring experiment could be derived where the color on specific areas on fish bodies is measured without them even leaving water. Furthermore, similar techniques could be used in order to classify other areas of the fish into Good/Bad categories. Head dimorfies, eye diseases or scale loss could be monitored similarly with Neural Networks and computer vision techniques. Finally, a model that automatically determines the welfare index of a fish just by its photograph is a future goal not far from reach.

5) References

[1] P. Andreas Svensson and Helen Nilsson Skold ,"Skin Biopsies as Tools to Measure Fish Coloration and Colour Change", DOI: 10.5772/19913

[2] M. Pavlidis, M. Karkana, E. Fanouraki, N. Papandroulakis, Environmental control of skin colour in the red porgy, Pagrus pagrus, Aquac. Res., 39 (2008), pp. 837-849, 10.1111/j.1365-2109.2008.01937.x

[3] Amundsen, T. & Forsgren, E. (2001). Male mate choice selects for female coloration in a fish. Proc. Natl. Acad. Sci. U. S. A. Vol. 98, pp. 13155-13160.

[4] The acute stress response of red porgy, Pagrus pagrus, kept on a red or white background A.L. Van der Salm a,b,n , M. Pavlidis c,d, G. Flik b , S.E. Wendelaar Bonga a

[5] Pavlidis, M., Papandroulakis, N. and P. Divanach, 2006. A method for the comparison of chromaticity parameters in fish skin: Preliminray data on the coloration pattern of red skin Sparidae ("red porgies"). Aquaculture, 258, 211-219.

[6] Song-Hee Choi, Byeong-Hoon Kim, Chi-Hoon Lee, Young-Don Lee, Response of body color change rearing under different light intensity conditions in farmed red spotted grouper, Epinephelus akaara, Fisheries and Aquatic Sciences, 10.1186/s41240-020-00173-8, 23, 1, (2020).

[7] Roulin, Alexandre. "Melanin-Based Colour Polymorphism Responding to Climate Change." Global Change Biology, vol. 20, no. 11, 2014, pp. 3344-3350., doi:10.1111/gcb.12594.

[8] Cote, J., et al. "Melanin-Based Coloration and Host-Parasite Interactions under Global

Change." Proceedings of the Royal Society B: Biological Sciences, vol. 285, no. 1879, 2018, p. 20180285., doi:10.1098/rspb.2018.0285.

[9] Ducrest, A. L., Keller, L., & Roulin, R. (2008). Pleiotropy in the melanocortin system, coloration and behavioral syndromes. Trends in Ecology and Evolution, 23, 502-510.

[10] Fujii R. The regulation of motile activity in fish chromatophores. Pigment Cell Res. 2000 Oct;13(5):300-19. doi: 10.1034/j.1600-0749.2000.130502.x. PMID: 11041206.

[11] Kittilsen, Silje, et al. "Pigments, Parasites and Personalitiy: Towards a Unifying Role for Steroid Hormones?" PLoS ONE, vol. 7, no. 4, 2012, doi:10.1371/journal.pone.0034281.

[12] Tserevelakis, George J., et al. "Hybrid Photoacoustic and Optical Imaging of Pigments in Vegetative Tissues." Journal of Microscopy, vol. 263, no. 3, 2016, pp. 300-306., doi:10.1111/jmi. 12396.

[13] Wang, L.V. & Wu, H.-I. (2007) Biomedical Optics: Principles and Imaging, 287 pp. Wiley, Hoboken, NJ, USA

[14] *Wang, L. V., & Yao, J. (2016). A practical guide to photoacoustic tomography in the life sciences. Nature Methods, 13(8), 627-638. doi:10.1038/nmeth.3925*

[15] Fred Rost, "Fluorescence Microscopy, Applications", Encyclopedia of Spectroscopy and Spectrometry (Third Edition), 2017, pp. 627-631, https://doi.org/10.1016/B978-0-12-803224-4.00147-3.

[16] Hegyi, J., and V. Hegyi. "New Developments in Fluorescence Diagnostics." Imaging in Dermatology,2016, pp. 89-94., doi:10.1016/b978-0-12-802838-4.00009-1.

[17] Zhang, Fengxiang, et al. "Preparation and Characterisation of Collagen from Freshwater Fish Scales." Food and Nutrition Sciences, vol. 02, no. 08, 2011, pp. 818-823., doi:10.4236/fns.2011.28112.

[18] Spring, Kenneth R., and Michael W. Davidson. "Introduction to Fluorescence Microscopy." Nikon's MicroscopyU, www.microscopyu.com/techniques/fluorescence/introduction-to-fluorescencemicroscopy.

[19] Bayguinov, Peter O., et al. "Modern Laser Scanning Confocal Microscopy." Current

Protocols in Cytometry, vol. 85, no. 1, 2018, doi:10.1002/cpcy.39.

[20] Combs, Christian A., and Hari Shroff. “Fluorescence Microscopy: A Concise Guide to Current Imaging Methods.” Current Protocols in Neuroscience, vol. 79, no. 1, 2017, doi:10.1002/cpns.29.

[21] “IF Imaging: Widefield versus Confocal Microscopy.” Proteintech Group, www.ptglab.com/news/blog/if-imaging-widefield-versus-confocal-microscopy/.

[22] Riyadh M. Al-saleem, Baraa M. Al-Hilali, Izz K. Abboud, "Mathematical Representation of Color Spaces and Its Role in Communication Systems", *Journal of Applied Mathematics*, vol. 2020, Article ID 4640175, 7 pages, 2020. https://doi.org/10.1155/2020/4640175

[23] Tserevelakis, George J., et al. “Delineating the Anatomy of the Ciliary Body Using Hybrid Optical and Photoacoustic Imaging.” Journal of Biomedical Optics, vol. 22, no. 6, 2017, p. 060501., doi: 10.1117/1 .jbo.22.6.060501.

[24] Tserevelakis, George J, et al. “Hybrid Autofluorescence and Photoacoustic Label-Free Microscopy for the Investigation and Identification of Malignancies in Ocular Biopsies.” Optics Letters, vol. 45, no. 20, 2020, pp. 5748-5751., doi:10.1364/0L.403435.

[25] Harrison, Tyler, et al. “Combined Photoacoustic and Ultrasound Biomicroscopy.” Optics Express, vol. 17, no. 24, 2009, p. 22041., doi:10.1364/oe.17.022041.

[26] Nuster, R., Schmitner, N., Wurzinger, G., Gratt, S., Salvenmoser, W., Meyer, D., & Paltauf, G. “Hybrid Photoacoustic and Ultrasound Section Imaging with Optical Ultrasound Detection.” Journal of Biophotonics, vol. 6, no. 6-7, 2013, pp. 549-559., doi:10.1002/jbio.201200223.

[27] Stephen V. Stehman,Selecting and interpreting measures of thematic classification accuracy,Remote Sensing of Environment,Volume 62, Issue 1,1997,Pages 77-89,ISSN 0034-4257, https://doi .org/10.1016/S0034-4257(97)00083-7.

[23] Geron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition,O'Reilly Media, Inc.

[24] Weijie Wang & Yanmin Lu,March 2018, “Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model”,IOP Conference Series Materials Science and Engineering 324(1):012049, DOI: 10.1088/1757-899X/324/1/012049

[25] https://commons.wikimedia.org/wiki/File:Melanophores with dispersed or

aggregated melanosomes.jpg

[26] http://molenergetics.fc.ul.pt/PAC 0055.html

[27] https://en.wikipedia.org/wiki/File:Jablonski Diagram of Fluorescence Only.png

[28] https://en.wikipedia.org/wiki/Stokes shift#/media/File:Stokes shift- Rh6G.png

[29] https://www.ptglab.com/news/blog/if-imaging-widefield-versus-confocal-microscopy/

[30] https://www.nbshare.io/notebook/626706996/Learn-And-Code-Confusion-Matrix-With-Py thon/

[31] Lloyd, Stuart P. "Least squares quantization in PCM." Information Theory, IEEE Transactions on 28.2 (1982): 129-137.

[33] Pedregosa, F., Varoquaux, Ga"el, Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... others. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12(Oct), 2825-2830.

[34] Abadi, Mart'in, Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... others. (2016). Tensorflow: A system for large-scale machine learning. In 12th $USENIX$ Symposium on Operating Systems Design and Implementation ($OSDI$ 16) (pp. 265-283).

[35] Chollet, F., & others. (2015). Keras. GitHub. Retrieved from https://github.com/fchollet/keras

[36] Kingma, Diederik P. and Jimmy Ba. “Adam: A Method for Stochastic Optimization.” CoRR abs/1412.6980 (2015): n. pag.

[37] Xu, Bing & Wang, Naiyan & Chen, Tianqi & Li, Mu. (2015). Empirical Evaluation of Rectified Activations in Convolutional Network.

[38] Mohammad Naved Qureshi, Mohd Vasim Ahamad, An Improved Method for Image Segmentation Using K-Means Clustering with Neutrosophic Logic, Procedia Computer Science, Volume 132, 2018, Pages 534-540, ISSN 1877-0509

[39] Shan, P. Image segmentation method based on K-mean algorithm. *J Image Video Proc.* 2018, 81 (2018). https://doi.org/10.1186/s13640-018-0322-6

[40] M. Usama *et al.*, "Unsupervised Machine Learning for Networking: Techniques, Applications and Research Challenges," in *IEEE Access*, vol. 7, pp. 65579-65615, 2019, doi: 10.1109/ACCESS.2019.2916648.