MULTI-DIMENSIONAL DATA STRUCTURES AND CLASSIFICATION SCHEMES AS A TOOL FOR THE NONDESTRUCTIVE ANALYSIS OF COMPLEX MATERIALS

by
Kargakos Stefanos
A Thesis

Submitted to the
Department of Electrical & Computer Engineering
In partial fulfillment of the requirement
For the degree of
Integrated Master of Science in Electrical & Computer Engineering
at
Technical University of Crete
July, 2022

Thesis Chair: Konstantinos Balas , Ph.D.

Committee Members:
Michalis Zervakis, Ph.D.
Vasilis Samoladas, Ph.D.

Abstract

Microscopy is a critical component for the visualization of samples & objects that cannot be seen with the unaided eye, allowing scientists to have a glimpse at a world of unimaginable complexity. Different specialized microscopes deploying several imaging techniques are essential for this purpose, however this way scientists tend to consume an unrealistic amount of time to achieve accurate diagnosis. In this thesis we exploit a new advanced HTS microscope featuring Hyperspectral, Transmission, Reflectance, Fluorescence and Polarization imaging. By illustrating the morphological, molecular, electronic and crystalline structures of matter, this microscope provides unique features for automated state of the art object classification. Upon observing those unique characteristics we approach a Multi-Modal object classification method, utilizing Convolutional Neural Networks for each modality and a Fully-Connected Neural Network which combines every unique illustration of the specimens. The CNN's outputs are providing a serial unique encoding for each image and the FC-NN serve as a decoder capable of processing tabular labeled data. This thesis provides extensive analysis and results regarding the different combinations of imaging modalities with the intention to extract valuable information about their importance on the classification process. Exploiting the full power of our NNs system, by deploying every imaging modality we achieve accuracy greater than **99%**.

Table of Contents

Abstract ..... **iii**

Chapter 1: Introduction **1**

Chapter 2: Background **5**

2.1 Microscope Utilities 5

2.2 Applications of Microscopy in Forensics Science 23

2.3 Artificial Intelligence 31

2.4 State Of the Art 39

Chapter 3: System Design **41**

3.1 Convolutional Neural Networks 41

3.2 Supervised Learning 49

3.3 Dataset Acquisition 50

Chapter 4: Implementation **55**

4.1 System Configuration 55

4.2 Single Modality Classification Methods 65

4.3 Multi Modal CNN Design 73

Table of Contents (Continued)

Chapter 5: Classification Results **79**

5.1 Data Visualization 79

5.2 Proof of Concept 80

5.3 Validation 82

Chapter 6: Conclusions and Future Work **98**

6.1 Future Work 102

References **104**

Chapter 1

Introduction

Brief Description Since always,human's curiosity, intelligence, inventiveness and need for answers drived us to create tools and machinery to explain and understand everything that surrounds us, the psychical attributes ,the chemical properties and the use of them for further exploration of the universe and its components. Subjects that could not go unnoticed is the microcosmos and macrocosmos that naked eye is not capable to examine. That led to inventing specialized machinery for enhanced observation that could help our research. Microscopy as one of the main research field for enhanced observation utilizes different observation techniques and with those in hand we can further explore and evolve our understanding on objects nature and potentials uses.

Microscopy is a key component for many science fields like Biomedical science, Forensic science, determining the health of an ecosystem ,studying the role of a protein within a cell, studying atomic structures, tissue Analysis(Histology) and many others. It can be used to solve problems, assist to a conclusion even prevent diseases to plants, animals and humans.

This thesis provides extensive analysis and results regarding the different combinations of imaging modalities with the intention to extract valuable information about their importance on the classification process. The specimens we have available are hairs, skin, fibers, blood, glass and sand. Microscopy is vital to multiple areas within forensic science and there is a variety of techniques utilized.

Traditionally, forensic document experts and paleographers used chemical solution based

methods to study the extrinsic and intrinsic components of the important historic documents [1]. These chemical solution based methods helped in document analysis. But unfortunately these techniques were time consuming, sensitive to temperature changes and destructive in nature i.e. harms to the important documents were irreversible.Microscopy implements non destructive ways of examine a trace and provide us information about the object's psychical substance by applying plenty of lighting techniques and different acquiring processes.

Microscopy can be applied in the identification of trace evidence such as *fingerprints, fibers, hairs,fragments* which are left the crime scene, on a victim or suspect.

Fragments of glass may be found in incidents such as house break-ins and road traffic accidents. Microscopes can be used to compare *shattered glass* left at the scene to that found on a suspect. Different types of glass have varying compositions, allowing scientists to analyze its origin by looking closely at its structure.

When a gun is fired, it leaves behind a chemical residue, known as *gunshot residue* (GSR). This residue can help identify a weapon used at a crime scene and link a suspect to a crime if the GSR is found to match that which was left at the scene of the crime.The tiny particles that makeup GSR measure between a few hundred nanometers to a few microns.

Forensic epidemiologists investigate how disease spreads. In *epidemiology*, microscopes can be used to investigate the root of an outbreak by tracing the pathogen involved to its source. This information can be essential in preventing further infection, reducing the risk of an epidemic.In vast times like nowdays with Covid-19 still spreading it could be extremely essential, also an attempt to "catch" covid traces at early stages in human lungs has already begun , similarly disasters could be prevented.

Microscopes have many uses in the field of *forensic anthropology*, the field of identifying the factors of death. For example, bone and other body tissues can be investigated using a microscope to gain clues about the cause of death.

Scanning electron microscopes are often used to study soil samples taken from the location of a body to determine the length of time the body had been there, the bones can be investigated for sharp force trauma and the coating left on the teeth may be examined, which can also be indicative of a cause of death.

Similar to forensic anthropology, *forensic pathology* investigates and determines the cause of death. The difference between forensic anthropology and forensic pathology is that forensic pathologist aims to establish the exact cause of death, rather than indicators of potential events leading to death.

Microscopy can be used to identify the bacteria or virus which caused death, or to examine body tissue for wounds, determining cause and investigating the potential that would be fatal.

Overall, there are a number of uses of microscopy in forensics. As technology continues to advance, it is expected that there will be more and more applications over the time. Even though the information provided by the microscope is extremely rich and useful enough to determine the psychical properties,the age of the object and identify it,the operator has to be an expert in a relevant field of the object so the results of the examination can be certified.This is not only time consuming, but in some cases, like a crime scene, where time and accuracy are critical and such an expert is hard to find at the time, all this specialized technology becomes useless .

A way around this issue is the development of a automated classifier embedded in the microscope which is programmed and monitored by an expert, hence anyone can use it

and come to a conclusion of the evidence and hopefully solve a part or a whole case.Our approach to tackle this problem is with Neural Networks which are ever-growing in the microscopy area with a variety of applications , from object classification,to background extraction and noise removal.Neural Networks are vastly used in computer imaging foremost their capability of auto-extracting the strong features of an image through image processing.

Chapter 2

Background

2.1 Microscope Utilities

2.1.1 Interaction of light and matter

The interaction between light and a specimen is determined by the optical properties of the specimen and the incident light. As hyperspectral imaging measures such interac-

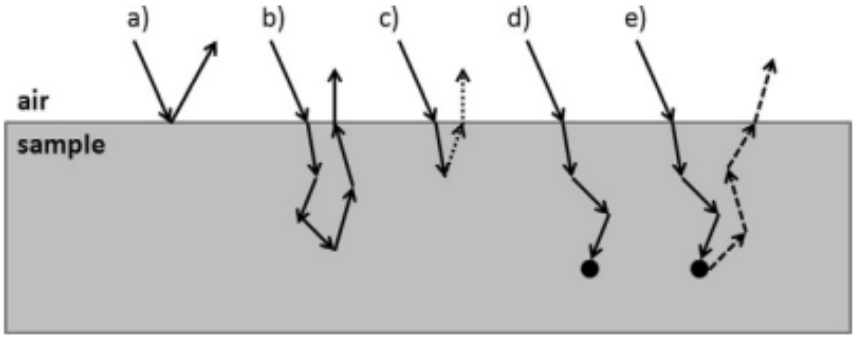


Figure 2.1: *The interaction of light with a specimen may lead to (a) specular reflection, (b) elastic scattering followed by diffuse reflection, (c) inelastic scattering followed by emission of Raman shifted light (dotted lines), (d) absorption, and (e) absorption followed by photoluminescence emission (dashed lines).*

tion, it may be used to characterize the material. In practice this involves illumination of the object under investigation. Commonly, the first interaction will be on the specimen surface where part of the light will be reflected (Fig. 1.1.a). This part contains no to little information from within the medium but is governed by the index of refraction difference between media. Upon entering the material, the light can be scattered or absorbed.

Scattering is the process by which light interacts with structures in a specimen and causes a change in direction of propagation, depending on the wavelength, size of the particle and index of refraction differences (Fig. 1.1.b). The majority of light is scattered

at the identical wavelength of the incident light, a process referred to as elastic scattering. There may also be a small fraction that will be inelastically scattered (Raman scattering) which will cause wavelength shifts corresponding to the vibrational states of the molecules in the specimen (Fig. 1.1.c).

The absorption properties of a chemical compound are wavelength dependent. Absorption in the visible wavelength range corresponds to the electronic states of the molecule, while absorption in the NIR and IR is determined by the vibrational modes. Upon relaxation, return to the ground state, the energy will be released in the form of radiation (heat or photoluminescence) or by transfer to another molecule. So both the spectral absorption and, if present,the induced photoluminescence can be measured to identify the chemical contents of a specimen using hyperspectral cameras in reflectance, or transmission mode (Fig. 1.1.d/e). Quantitative analysis, however, is complicated because the length of the path traveled by the detected light depends on the optical properties of the specimen [2].

Transmission of light through the electromagnetic spectrum, where light waves pass through a material without absorption, is affected by thickness, and type of material.Through that technique we can acquire:

2.1.2 Widefield Microscopy

Widefield microscopy, is fundamentally any technique in which the entire specimen of interest is exposed to the light source with the resulting image being viewed either by the observer or a camera.

- **Brightfield Microscopy**

  Brightfield Microscopy The most basic form of Widefield microscopy is 'brightfield microscopy' in which the entire specimen is illuminated by white light either from above (in an inverted configuration), or below (in a standard upright microscope), as shown in Figure [2.6]. This configuration has influence on the specimen which can

be investigated. Upright microscopes are commonly used for applications with fixed samples, mounted on glass slides. Inverted microscopes have been invented to watch living cells. They are commonly grown in liquid solutions. Only a configuration with the objective below and the condenser above the specimen guarantees a close enough proximity of the objective to the specimen.

Figure 2.2: *Widefield Microscopy Setups*

Construction of a microscope using the brightfield microscopy technique, is based on these components.

- **Light Source**: Trans-illumination below sample that propagates through to condenser and objective lens. Typically a broadband source such as a quartz halogen bulb or LEDs is used.
- **Condenser Lens**: Collects trans-illuminated light and focuses to sample
- **Objective Lens**: Collects light which propagates through sample and enhances details by a factor of magnification.
- **Eyepiece/Camera**: Views or records the image.

A typical brightfield illumination image has a dark sample with a white background, the darker the regions on a sample, the more absorption of light that has occurred, as shown in Figure[1.6].

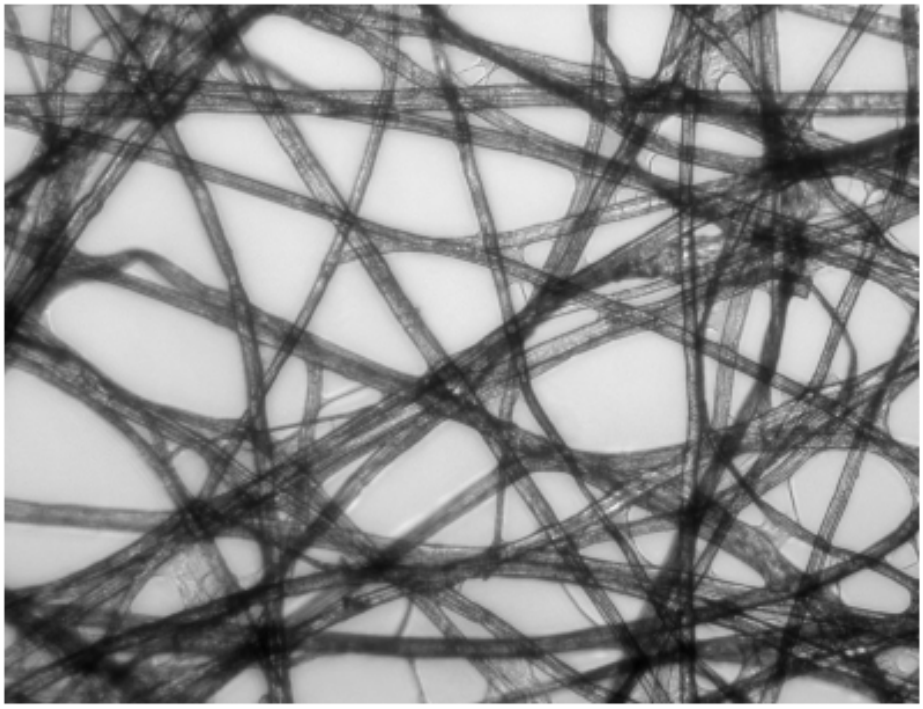


Figure 2.3: *Brightfield Illumination Image of Tissue Paper. Source:https://www.edmundoptics.com/*

2.1.2.1 Reflectance Microscopy Reflectance microscopy is capturing the light that is being reflected by the specimen, when illuminating light reaches the specimen, it may absorb some of the light and reflect some of the light, either in a specular or diffuse manner. The only restriction is that the microscope must have an operational spectral range from the deep-UV to the NIR while maintaining good image and spectral quality. Standard microscopes cannot be used as they only cover a portion of the visible spectrum due to their optical designs and light sources.

2.1.3 Hyperspectral Imaging

Hyperspectral imaging (HSI) integrates conventional imaging and spectroscopy, to obtain both **spatial and spectral** information from a specimen. This technique enables investigators to analyze the **chemical composition** of traces and simultaneously visualize their spatial distribution. HSI offers significant potential for the detection, visualization,

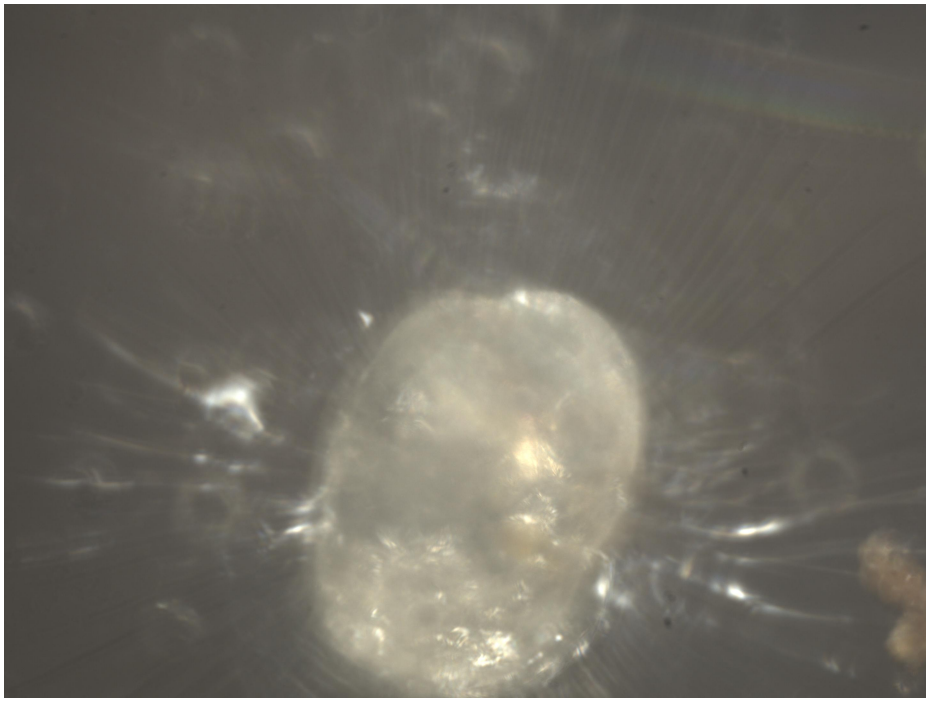


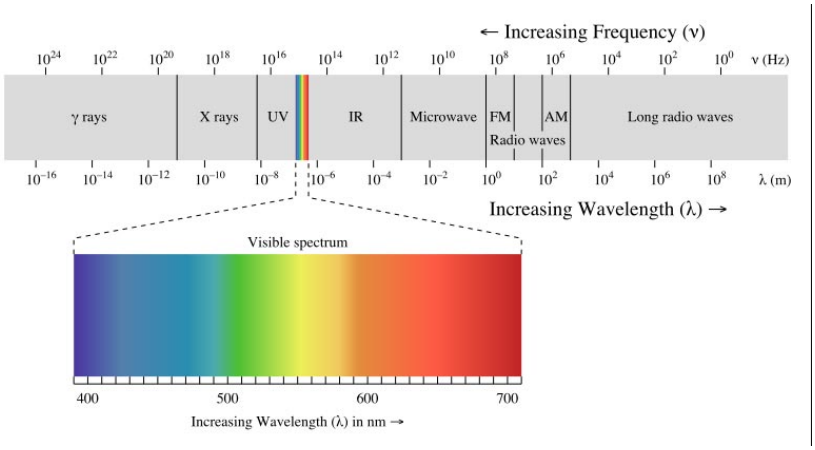
Figure 2.4: *Sand captured with Reflectance Microscopy on the visible spectrum*

identification and age estimation of forensic traces. The rapid, non-destructive and non-contact features of HSI mark its suitability as an analytical tool for forensic science.

A wide range of techniques are available when implementing a Hyperspectral Imaging system, including chemical enhancement techniques and the use of light sources with 1-10 nm bandwidths. Many of these techniques are, however, either destructive or subject to human interpretation. Hyperspectral imaging (HSI) is suitable for the **non-contact** identification of evidence, thus **minimizing the risk of contamination and destruction of traces**. HSI integrates conventional imaging and spectroscopy to obtain a three dimensional data set containing both spatial and spectral information of a specimen. In addition, analysis of the temporal behavior of spectra can give insight in the chemical changes within the specimen, which can be used for **age estimation** purposes. Estimation of the age of forensic traces provides investigators with valuable information, which can assist the reconstruction of the timeline of events.

HSI was originally developed for **remote sensing** applications utilizing **satellite imaging data of the earth** [3] but has since found application in such diverse fields as **food science** [4], **pharmaceuticals** [5] and medical diagnostics [6]. Hyperspectral images are analogous to a stack of images, each acquired at a narrow spectral band. Like spectroscopy, HSI can be applied in different parts of the electromagnetic spectrum, like :

- Ultraviolet (UV)
- Visible (VIS)
- Near infrared(NIR)
- Mid infrared (IR)
- Thermal infrared



in those regions we can capture:

- **Reflectance**
- **Transmission**
- **Photoluminescence**
- **Luminescence**

by hyperspectral cameras with a spectral resolution similar to miniature spectrographs. Spatial resolutions can be adapted to the application, which range from microscopic to landscapes. Advantages of HSI include **speed of data acquisition, reduction of human error, no destruction of traces, no specimen preparation, and the ability to illustrate the results**.

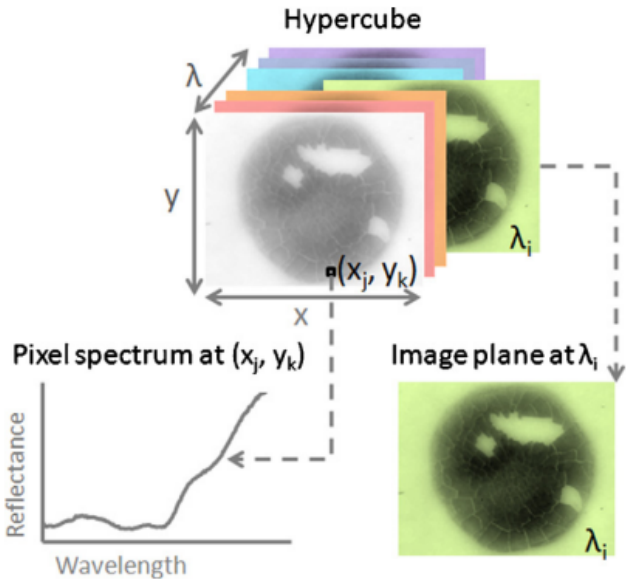


Figure 2.5: *Hypercube of a blood stain, with two spatial (x,y) and one wavelength (λ) dimension. From the hypercube an image plane is shown for one wavelength (λ_i) and a spectrum is obtained from one pixel (x_j, y_k).*

2.1.3.1 Hypercube Formation Hyperspectral images are analogous to a stack of images, each acquired at a narrow spectral band. The resulting data set is a three-dimensional block of data, the so-called hypercube, with two spatial (x,y) dimensions and one wavelength (λ) dimension (Fig. 2). This hypercube provides images for each wavelength (λ_i) and a spectrum can be obtained from each individual pixel (x_j, y_k), as depicted in Fig.1.2.

Obtaining information in all three dimensions of a hypercubes simultaneously is currently not feasible; instruments can only capture two dimensions at a time. Temporal

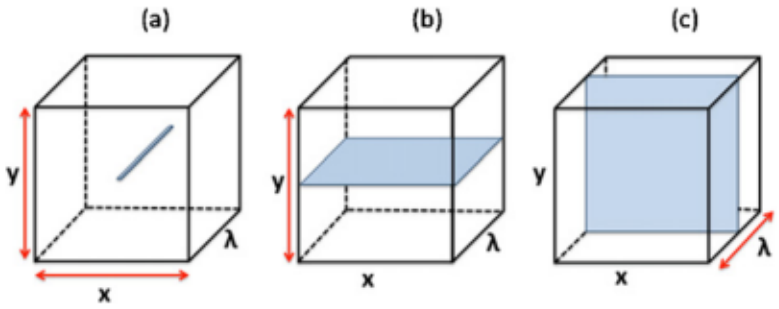


Figure 2.6: *Methods for acquiring three-dimensional hypercubes: (a) point scanning, (b) line scanning, and (c) area scanning. Hypercubes contain two spatial (x,y) and one spectral (l) dimension. Blue areas represent data acquired by one scan. Red arrows represent temporal scanning required to complete the hypercube. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)*

scanning is needed to create a three-dimensional hypercube by stacking the two dimensional data in sequence. There are three ways of acquiring a hypercube (Fig. 1.3), commonly known as point scanning (or whiskbroom), line scanning (or pushbroom), and area scanning (or staredown). These descriptive names refer to the hardware methodology used to acquire the hypercubes:

- In a point scanning system, a complete spectrum is acquired at a single point. Light originating from this point enters the objective lens and is separated into different wavelengths by a spectrometer and detected by a linear array detector. Once spectral acquisition is completed, the spectrum of another point can be recorded. Scanning has to be performed in both spatial directions to complete the hypercube.

- In the case of line scanning systems, the spectra of all pixels contained in one image line are acquired simultaneously. The light is dispersed onto a two dimensional charge coupled device (CCD) detector. This way, a two dimensional data matrix with the spectral dimension and one spatial dimension is acquired.The second spatial dimension of the hypercube is achieved by scanning across the specimen surface in

a direction perpendicular to the imaging line. This means that relative movement between the object and detector is necessary,which may be achieved either by moving the specimen (e.g. using a translation stage or a conveyor belt) and keeping the hyperspectral camera in a fixed position or by moving the camera and keeping the specimen fixed.

- An area scanning system also acquires a two-dimensional data matrix but in this case the data represent a more conventional image with two spatial axes. A complete hypercube is obtained by collecting a sequence of these images for one wavelength band at a time. The wavelength of incoming light in this configuration is typically modulated using a tunable filter.

2.1.3.2 System for Hyperspectral imaging Typical hyperspectral imaging systems contain the following components: objective lens, wavelength modulator, detector, illumination, and acquisition system (Fig. 1.4). All these components can be adjusted to the requirements of the application. The forensic environment of analysis may range from laboratory to field conditions, whereas the areas of interest may range from the microscopic to landscapes. As for conventional imaging, different objective lenses can be chosen to obtain the right spatial resolution for each application, e.g. macroscopic lenses, zoom lenses, wide angle lenses etc. For analysis on a microscopic scale, the HSI system can be coupled to a microscope.

2.1.4 Polarized Imaging

Visible light is actually electromagnetic radiation inside the electromagnetic band which is visible to the human eye.It has both the electric and magnetic components, and as we can see in figure 1.6 it is a transverse wave and it's electric field vectors vibrate in all planes that are perpendicular with respect to the direction of propagation and are random in space and time. If we manage to restrict the electric component into one plane then this

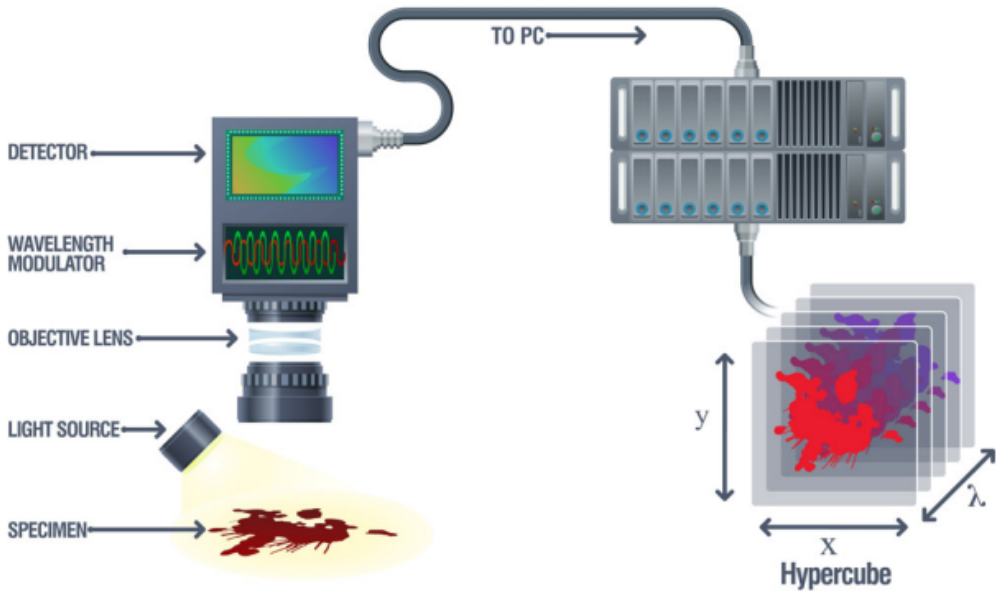


Figure 2.7: *Schematic showing components of a HSI system, resulting in a hypercube of the specimen.*

is referred to as **linearly polarized** .Light can be **linear,circular or elliptical** poralized.

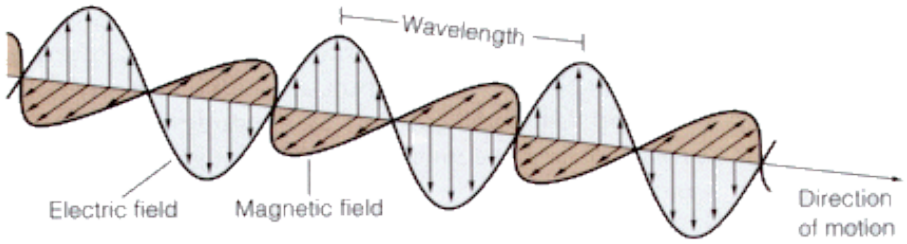


Figure 2.8: *A linearly polarized electromagnetic wave.*

2.1.4.1 Linear Polarization The essential component that we need to polarize the light is a polarizer, basically a filter that contains long-chain polymer molecules that are oriented in a single direction.And the light that vibrates in the same plane as the oriented polymer molecules will be absorbed, whereas the light that vibrates vertically to this plane with pass through and create a wave.

When the unpolarized beam propagates towards the vertical polarizer of Figure 1.7, all electric field vectors vibrate perpendicular to the direction of propagation with an

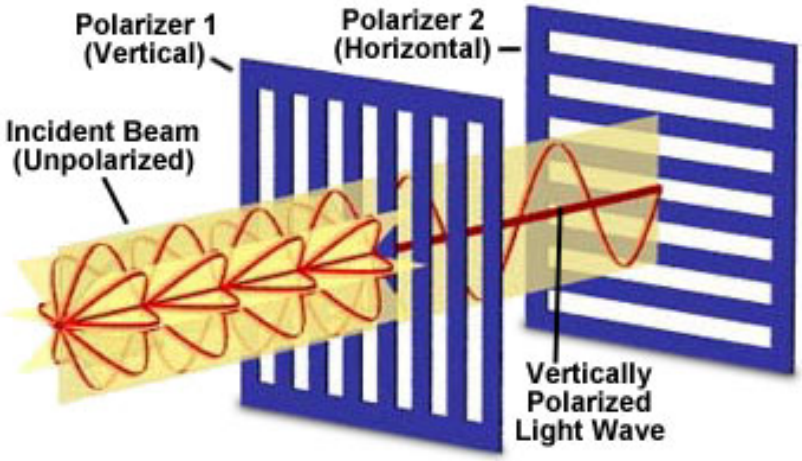


Figure 2.9: *An Incident Beam passing through a cross polarizer setup.*

equal distribution in all planes .When light meet the first polarizer, every plane except the vertical one is blocked while light after the filter creates a vertically polarized light beam.By encountering the first polarized, only the electric vector that lives on the vertical plane is allowed through. Resulting in a vertically polarized light beam.And when this beam encounter the horizontal polarizer, no light passes through because it was absorbed by the filter.

2.1.4.2 Elliptical and Circular polarization When the electric field of the wave has circle shape and its always perpendicular to the direction of the propagation its called *circular polarized light*.That effect is created by placing two linearly polarized beams with an appropriate angle with respect to each other,with the same amplitude and phase shift $\Delta\phi = n$.

When either the phase shift between the two light beams or the amplitude of one wave or both change, then the light is described as elliptically polarized.*Elliptical polarization*, which is a general state of circular polarization, has a rotational "sense" that refers to the direction of electric vector rotation around the propagation axis of the light beam.

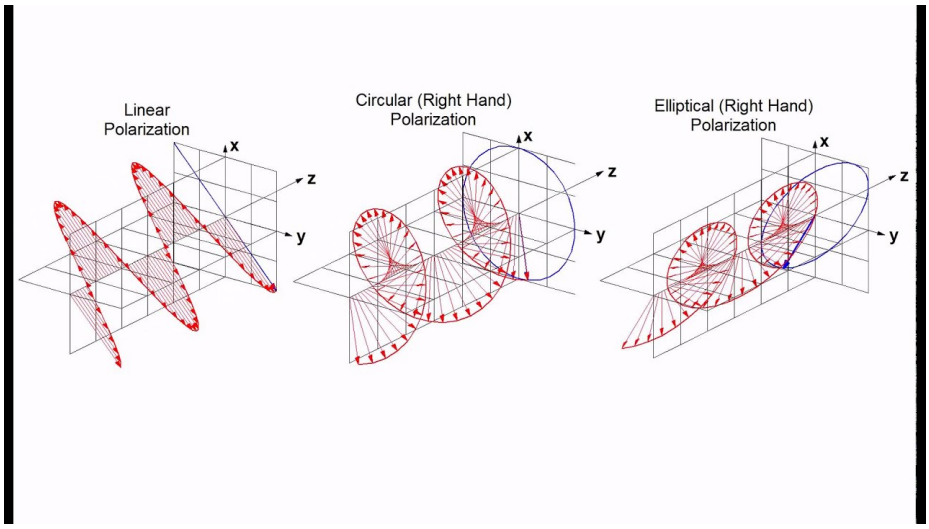


Figure 2.10: *Linear,Circular and Elliptical Polarization Graphs*

2.1.4.3 Birefrigence is the optical property of a sample having a refractive index that depends on the polarization and propagation direction of light.A refractive index is number that can give the indication of the light bending ability of the sample. These optically anisotropic materials are said to be birefringent (or birefractive). Birefringence is responsible for the phenomenon of double refraction whereby a ray of light, when incident upon a birefringent material, is split by polarization into two rays taking slightly different paths.The phase shift of this two paths can gives us plenty of information

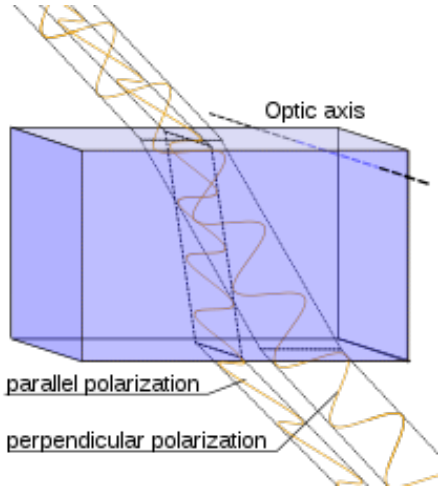


Figure 2.11

2.1.4.4 Polarized Microscopy Polarized light is a contrast-enhancing technique that improves the quality of the image obtained with birefringent materials. Polarized light microscopes have a high degree of sensitivity and can be utilized for both quantitative and qualitative studies targeted at a wide range of anisotropic specimens.It is designed to observe and photograph specimens that are visible primarily due to their optically anisotropic character. There are several ways of analysing the polarization of birefringent specimen.

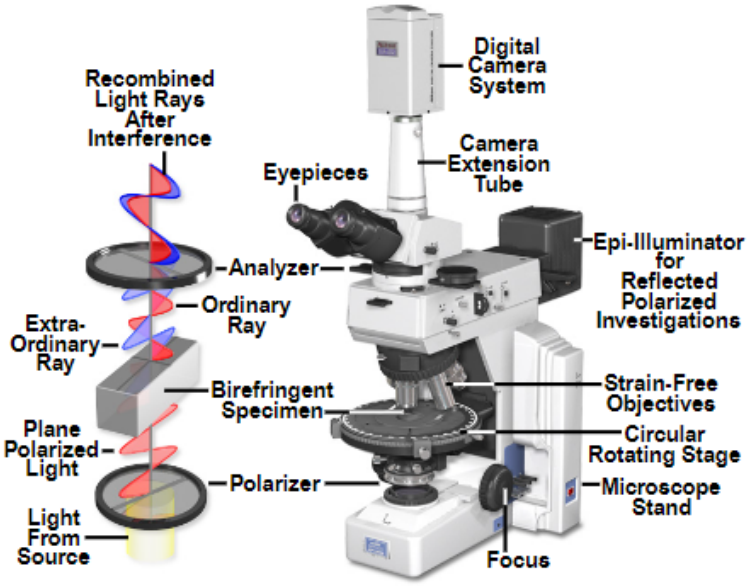


Figure 2.12: *Cross Polarized Microscope setup*

A common technique for polarimetry analysis is cross polarization microscopy. The components used in this technique are a **broadband light source** like a LED, a **polarizer**, positioned in the light path somewhere before the specimen, and an **analyzer** (a second polarizer; see Figure 1.10), placed in the optical pathway between the objective rear aperture and the observation tubes or camera port.The background is extracted by using a pair of **polarizers** in a *cross polarized* formation at a specific angle,which they work as filters.

Some other, more advanced, polarimetry imaging analysis techniques are based on the **Stokes Parameteres** which are a set of calculated values that describe the polarization state of electromagnetic radiation. These stokes parameters are obtained by measuring sev-

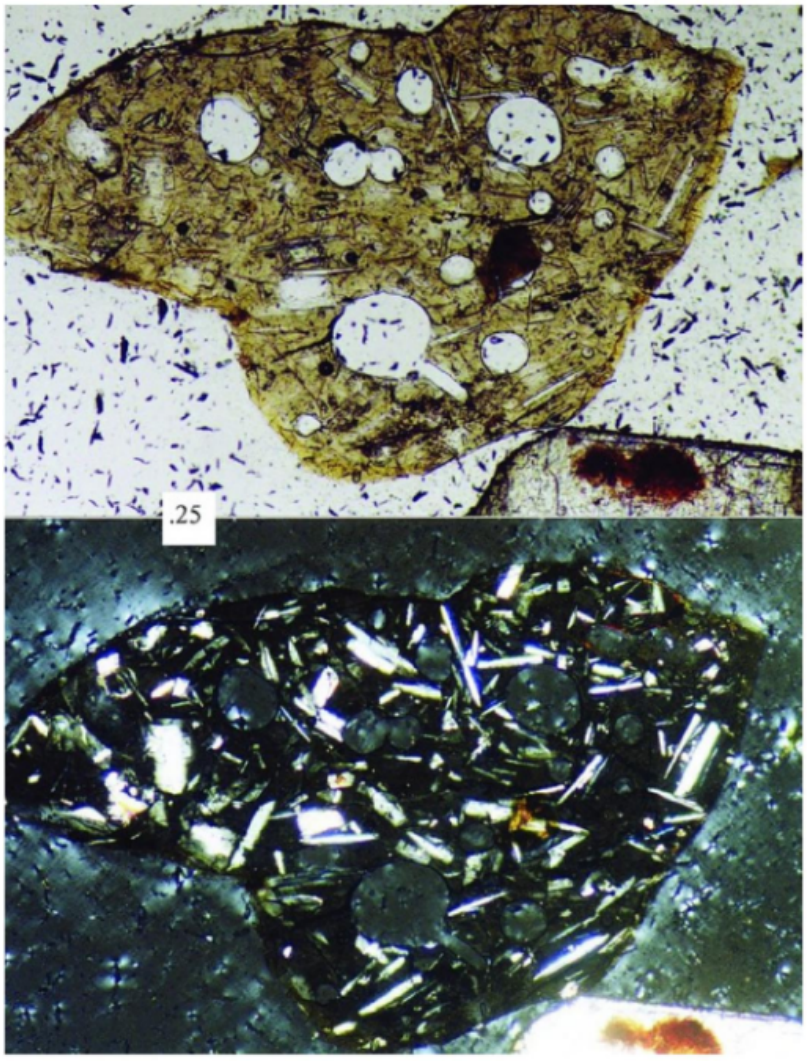


Figure 2.13: *Photomicrograph showing soil grain under plane and cross polarized light microscope*

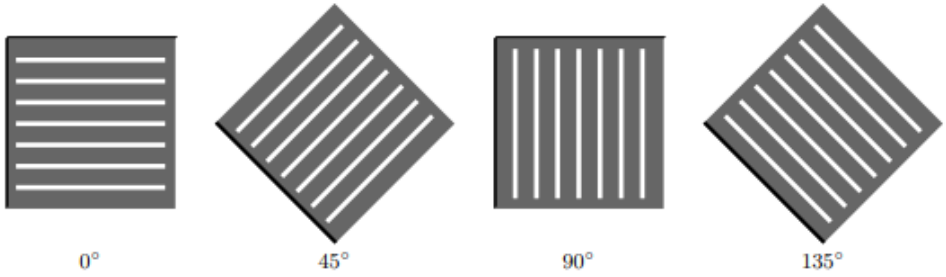


Figure 2.14: *Polarizers in four different angles.*

eral different angles under the condition of right-circular polarized incident light. **Stoke**

Parameters:

- I = F where the flux density (F) is the total flux density of the unobstructed light.
- Q = F_{0° - F_{90° where the degree denotes the flux density of the beam after passing through the filters of Fig.1.10.
- U = F_{45° - F_{135° where the degree denotes the flux density of the beam after passing through the filters of Fig.1.10.
- V = $2F_C - I$ where F_C is the flux density of clockwise circular polarized light.

The modern notation for the Stokes parameters is the Stokes vector S˜ with its components S_0, S_1, S_2, S_3 such as:

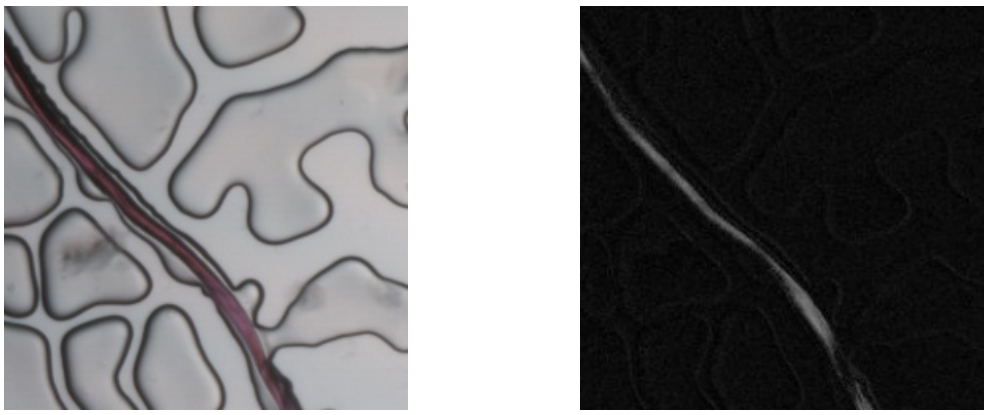
$$\overrightarrow{S} = \begin{pmatrix} \mathrm{S}_0 \\ \mathrm{S}_1 \\ \mathrm{S}_2 \\ \mathrm{S}_3 \end{pmatrix} = \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix}$$

When computed, the next step is to find the **Degree of Linear Polarization**, or DoLP.

$$DoLP = \sqrt{S_1^2 + S_2^2}/S_0 \tag{2.1}$$

Finally, by observing the DoLP Stoke image we can discriminate differences in the polarization characteristics of a specimen. This parameter is in fact very useful for discriminating similar (under common microscopy inspection) materials.

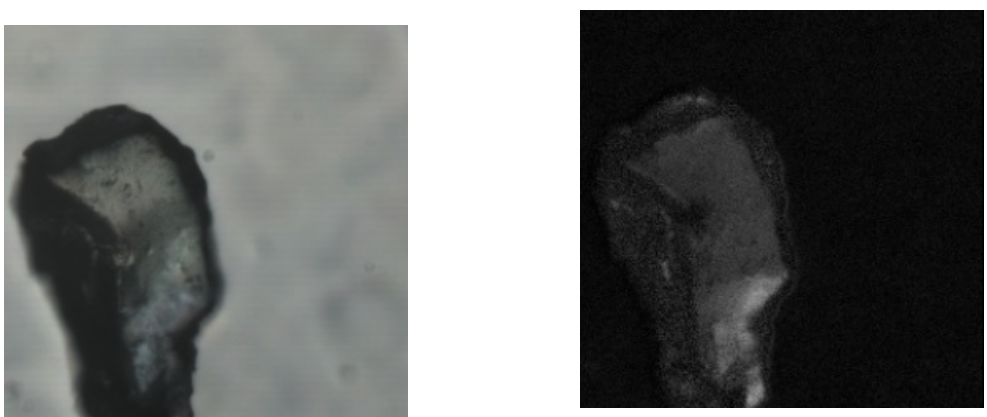
Polarized light microscopy is capable of providing information on absorption color and optical path boundaries between minerals of differing refractive indices, information



(a) *Image of a fiber captured using brightfield microscopy's technique.*

(b) *The same fiber captured using polarized microscopy and DoLP parameter.*

Figure 2.15: *Characterization of birefringent fiber using DoLP imaging.*



(a) *Sand captured using brightfield microscopy*

(b) *The same sand captured using polarized microscopy.*

Figure 2.16

about the crystallic lattice of the sample and obviously if the object is isotropic or not. Also the contrast-enhancing technique exploits the optical properties specific to anisotropy and reveals detailed information concerning the structure and composition of materials that are

invaluable for identification and diagnostic purposes.

2.1.5 Epi-Fluorescence

A fluorescence microscope is like a conventional light microscope with added features to enhance its capabilities. The conventional microscope uses visible light (400-700 nanometers) to illuminate and produce an image of a sample. On the other hand, a fluorescence microscope uses a much higher energy light source which excites a fluorescent species in a sample of interest, and this in turn emits a lower energy light of a longer wavelength. By capturing only the emitted wavelengths of light (filtering out high energy excitation radiation) we acquire an image that describes the electronic structure of the material under inspection[7].

Components that we use: A ***Light Source***, powerful LEDs or xenon arc lamp.
The filter that narrows down the wavelengths of the light that pass through to only those who excite the sample is called ***Excitation Filter.***
A ***Dichroic Mirror*** which acts like a beam splitter, reflecting the light coming from the source at the sample and transmitting only the emmited light from the sample to the detector. This component is used in order to perform illumination and detection through the same lens system.
Another *filter for emitted light* that transmits only the wavelengths of the emitted light from the sample and blocks all the light passed through the excitation filter.
Any capable ***camera*** to capture and view the image.

The basic task of the fluorescence microscope is to let **excitation light radiate the specimen** and then sort out the much weaker emitted light from the image. First, the microscope has a **emission filter** that only lets through radiation with the specific wavelength that matches your fluorescing material. The radiation collides with the atoms in your specimen and **electrons are excited to a higher energy level**. When they **relax** to a lower level, they **emit light**. To become detectable (visible to the human eye) the fluorescence emitted

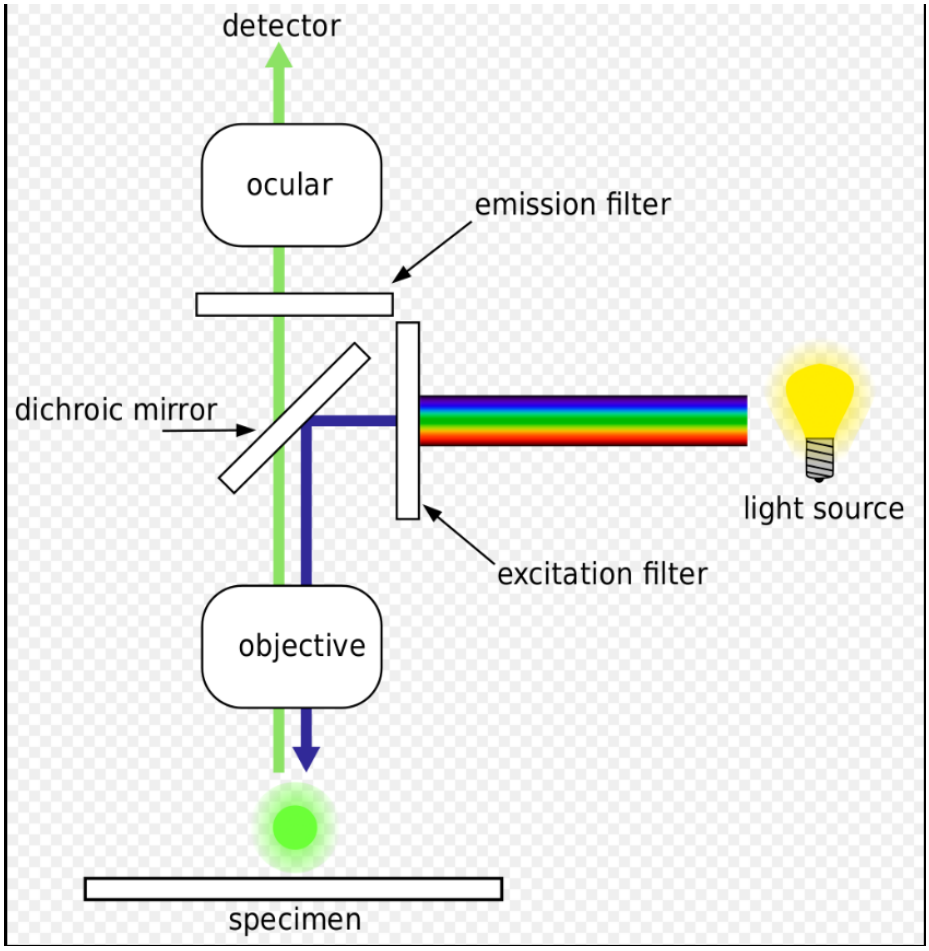


Figure 2.17: *Schematic of a fluorescence microscope(by wikipedia: Henry Mühlpfordt (talk) Fluoreszenzmikroskopie*$_2$*008 – 09 – 28.svg* : .

from the sample is separated from the much brighter excitation light in a second filter. This works because the emitted light is of lower energy and has a longer wavelength than the light that is used for illumination.

Most of the fluorescence microscopes used in biology today are epi-fluorescence microscopes, meaning that both the excitation and the observation of the fluorescence occur above the sample. The filters are assigned to capture the maximum excitation and emission wavelengths of a specific fluorophore but not the whole fluorescence.With the excitation filter we can define a band of wavelength that we let pass through, and we use it as a band-pass filter.Emission filters are usually band-pass or long-pass filters, depending on the specific fluorophore and imaging experiment.

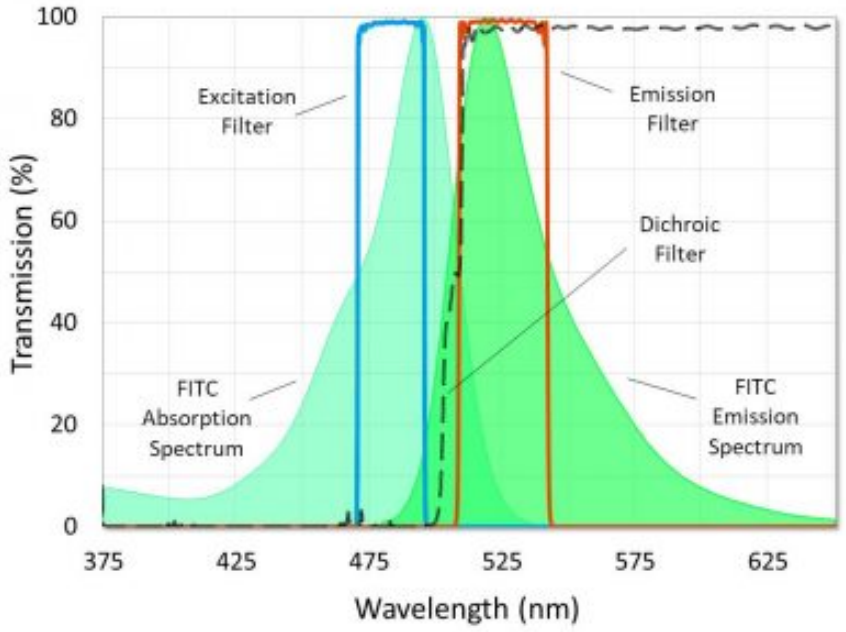


Figure 2.18: *A fluorescence filter set designed for use with FITC(fluorescence dye).Source:https://www.alluxa.com/optical-filter-applications/fluorescence-filters-microscopy-imaging/*

2.2 Applications of Microscopy in Forensics Science

2.2.1 HSI Applications in Forensics

HSI is a powerful emerging tool for the analysis of forensic traces. Latent traces can be detected and visualized by using spectral differences to obtain optimal contrast between a trace and its background. Individual spectra give information about the chemical composition of the specimen, which is useful for identification and quantification purposes, and the spatial distribution of traces is simultaneously recorded. In the last decade, HSI has proven to be a valuable technique for the imaging of **latent fingermarks** and the **detection of trace materials** within these prints. HSI is also emerging in other fields of forensic science and has shown its value in comparative research of materials including **fibers, paint chips, or inks**, where the question arises whether two traces share common origin. The possibility of viewing spectral and spatial information side by side is advantageous in these cases.

Recent developments in HSI technology added potential for forensic science investigations. Because HSI systems are becoming increasingly **portable**,they may be used at the scene of investigation, where traces can be viewed and interpreted in the original context. The development of fast scanning systems enables investigators to scan a complete scene, which reduces the workload in forensic laboratories and quickly provides investigators with valuable information which can lead the investigation.

Although hyperspectral imaging has mainly been used for the analysis of fingermarks, studies are also reported on several other forensic traces, including drugs, hair, dentin, bruises, blood stains,condoms, inks, tapes, firearm propellants, paints and fibers.

2.2.1.1 Fingermark Applications

Fingerprints are the unique identities of each individual and do not change as an individual gets older, thereby making this type of forensic evidence extremely powerful during criminal investigations. In addition to identifying individuals present at a crime scene, fingerprint analysis is also crucial to preventing and detecting counterfeiting.

Several authors recently evaluated the possibility of detecting untreated latent fingermarks using HSI. Exline et al. used visible reflectance and photoluminescence HSI to detect untreated latent fingermarks on plastic and paper[8].

Resulting images were compared to images created with a conventional forensic imaging system, in which different excitation and observation wavelengths could be chosen. While both methods succeeded in visualizing latent fingermarks on plastic, HSI showed enhanced contrast on paper surfaces. Processing tools used included background division, offset correction, normalization and PCA. In a further study Payne et al. optimized this visualization technique by using different processing tools to achieve an improved image[9].

Crane et al. demonstrated the ability of IR HSI to detect latent untreated fingermarks on various porous backgrounds (copier paper, cigarette butt paper, U.S. dollar bill,

Figure 2.19: *Two-year-old fingerprint on a white plastic garbage bag after treatment with cyanoacrylate and Basic Red/Basic Yellow stain (observation in the luminescence mode): CONDOR™ (left), Poliview (right).Picture from:[8]*

postcard) and nonporous backgrounds (trash bags, a soda can, tape)[10].

In two papers, Tahtouh et al. also described the application of infrared HSI to the visualization of untreated fingermarks [11, 12]. Results indicated that the infrared spectra of many untreated fingermarks show peaks due to C–H stretching vibrations around 3333 nm, mainly due to fatty acid residues. These peaks are common to most organic compounds, but they can be used to visualize fingermarks against some backgrounds, like metals, minerals, and ceramics, that do not contain C–H bonds. For fingermarks on other backgrounds, they stated that some type of chemical enhancement technique is required prior to hyperspectral imaging.

Bhargava et al. described an approach to use IR HSI to reveal latent fingermarks overlaid on top of one another, each made under different hand washing conditions [13]. Differences observed in the absorbance of the C–H stretching mode and other vibrational modes in the spectra indicated that the two prints had different chemical compositions. Because of this variation, linear unmixing applied to the spectral content of the data could be used to provide images revealing both superimposed fingermarks.

Figure 2.20: *Cut and flattened Dr. Pepper's soda can with fingermark deposit. (A) Soda can imaged by a document scanner. (B) Infrared image of the outlined area obtained by plotting the band intensity at 9842 nm (1016 cm1).*

2.2.1.2 Other HSI Applications Apart from the analysis of fingermarks, the benefits of HSI can be exploited for the analysis of many other traces of importance in forensic science. Latent traces can be detected and visualized by using **spectral differences**

to obtain optimal contrast between a trace and its background. Individual spectra give information about the chemical composition of the specimen, which is useful for identification, quantification, or age estimation. The possibility of viewing spectral and spatial information side by side is an advantage in comparative research of e.g. fibers, paint chips, or inks, where the question arises whether two traces share common spectral features.

Kalasinsky et al. were the first to demonstrate the value of infrared HSI for determining **drugs of abuse in hairs**[14]. By examining only the interior portion of the hair, drugs exclusively resulting from human ingestion were measured and distinguished from drugs that made contact with the outside of the hair. After microtoming the hair, IR hyperspectral images were obtained of the cortex and the medulla. Drug free hairs of different sources all correlated with standard spectra of proteins.An intensity band image at 5824 nm showed that the drug was concentrated in the center of the hair. This way, relative drug concentrations across the hair could be successfully determined and visualized. In a further study, Kalasinsky showed the distribution of drugs in human hairs, which is critical information to validate drug testing data [15].

The analysis of **bruises**, or aging of bruises in particular, can give important evidence in cases of domestic violence or child abuse. Several studies have been performed as initial steps toward the **aging** of bruises using HSI. A bruise is formed after blunt trauma, which results in blood being present in the skin. In time, hemoglobin in the blood is degraded into other products, including bilirubin. Both hemoglobin and bilirubin have typical spectral features in the visible region[16].

Payne et al. showed the possibility to use HSI to differentiate pure blood from blood with bilirubin based on these spectral features [17]. Randeberg et al. presented hyperspectral images of bruises on porcine and human skin [18]. They used minimum noise fraction transform, a statistical method similar to PCA, to classify the injuries. Stam et al. described how HSI can be used to accurately determine the areas covered by hemoglobin and bilirubin in the bruise, by fitting pixel spectra with a combination of reference spectra

of chromophores present in bruises [19].

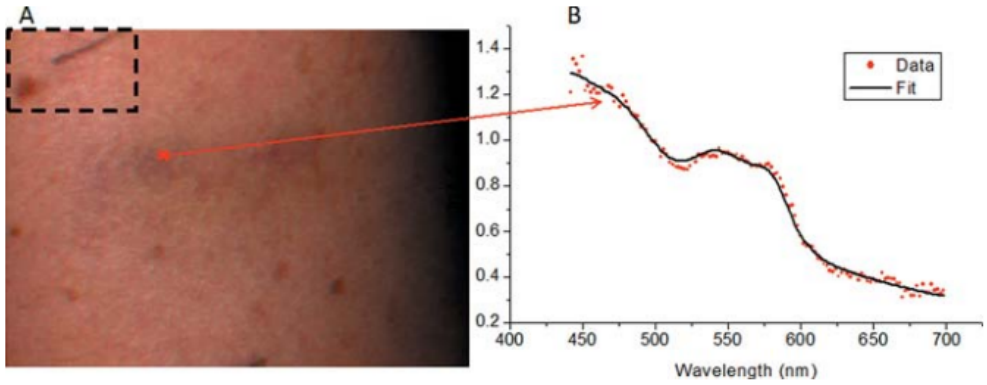


Figure 2.21: *(A) RGB image of the first bruise at 52.1 hours. The dashed square indicates the ROI chosen for alignment. The red dot indicates the pixel for which the spectrum is shown in B. (B) Spectrum of pixel (dots) and spectral fit (line).[19]*

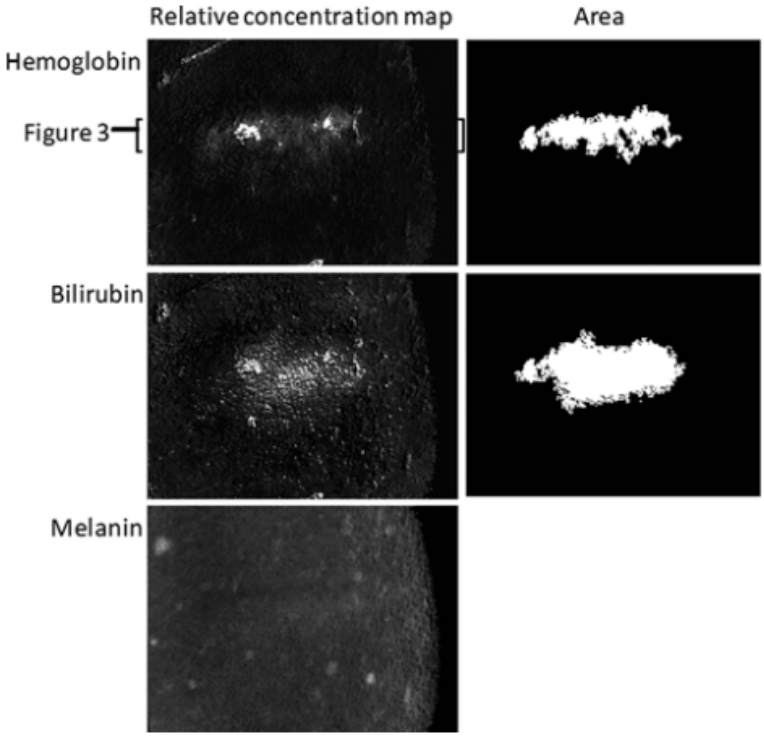


Figure 2.22: *The relative concentration maps obtained using the fit coefficients(wavelength dependent variables of bruise's chemical composure who were extracted by the RGB image.) in every pixel are show in Figure 1.17, left column. Left column: Relative concentration maps of hemoglobin, bilirubin and melanin. Right column: Binary images of hemoglobin and bilirubin areas.[19]*

2.2.2 Fluorescence Microscopy collaboration with Forensics Science

Fluorescent microscopy is associated with several advantageous properties that make it an ideal tool for forensic researchers. Fluorescent microscopy techniques have been widely used within forensic science to assist in the visualization of latent fingerprints, identification of GSR and examination of other types of trace evidence.

Fingerprints are often invinsible to the naked eye, additional efforts by forensic investigators are required to develop and ultimately allow for the visualization of latent fingerprints (LFPs). Several traditional methods that can be used for the development of LFPs include powder dusting, ninhydrin spraying and cyanoacrylate fuming, many of which are simple and effective methods.

Due to the fluorescent label potential of lanthanide phosphors, recent work on this material has led to the fabrication of red-emitting La2(MoO4)3:Eu3+ phosphor microcrystals for the visualization of LFPs deposited onto several different nonporous substrates. Some of the unique characteristics of this novel material include high contrast and low background interference, both of which contribute to the ability of this label to provide well-defined finger ridge details when analyzed by SEM[20].

And plenty more uses of fluoresence microscopy on latent fingermarks, like a research on **drugs** left on fingermarks and obtaining four week aged both in presense and absense of light fingermarks from smoked cigarretes[].

Another trace that can be indentified by fluorescence microscope is **gunshot residue(GSR)** due to the fact that it leaves behind evidence which is principally composed of burnt and unburnt particles from the detonation, as well as fragments of the bullet, cartridge case, and the firearm. This is why it has been applied to many real cases and attract so much interest and is considered a necessary tool for the researchers , as Damian K. et. al. who demonstrated the capacity of time-resolved fluorescence microscopy as a practical analytical tool in the forensic sciences via the imaging of gunshot residues that are expelled when a firearm is discharged[22, 23].

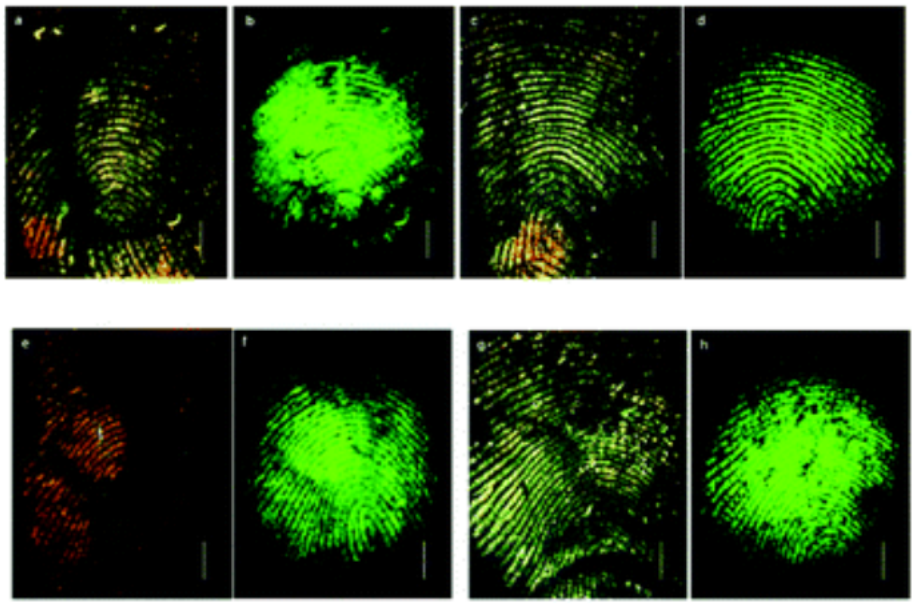


Figure 2.23: *Detection of fingermarks after ageing at room temperature in the presence and absence of light. Brightfield (a) and fluorescence (b) images of fingermarks after ageing for 3 days in the presence of light. Brightfield (c) and fluorescence (d) images after ageing for 4 weeks in the presence of light. Brightfield (e) and fluorescence (f) images of fingermarks after ageing for 3 weeks in the absence of light. Brightfield (g) and fluorescence (h) images after ageing for 4 weeks in the absence of light[21]*

Our **task** for this thesis was to try to **classify specimens** like Hair, Blood, Sand, Skin etc., objects that can be found on any **crime scene** .Those traces may be evidence or lead the investigation and must be classified correctly and quickly. If the examination of an expert is no longer needed it will save a great amount of time, and sometimes time has great importance.

A constantly increasing in popularity choice for object classification of any kind is **Neural Networks**. A great variety of applications have used Neural Networks from system identification, geomorphology or even in brain research ANNs have studied short-term behavior of individual neurons,[24] the dynamics of neural circuitry arise from interactions between individual neurons and how behavior can arise from abstract neural modules that represent complete subsystems.

But Neural Networks is also used in a big variety of forensics applications like identifying Recaptured Forensics images that are very small with Laplacian Convolutional

Neural Networks that put signal enhancement layer into Convolutional Layer and Laplacian filter is used in the signal enhancement layer[25].

Taken in account the various applications in many fields and the preference of many to choose Neural Networks to decide for them we decide to approach our thesis task also by using Neural Networks.

2.3 Artificial Intelligence

Artificial intelligence (AI) is the broadest term used to classify machines that mimic human intelligence. It is used to predict, automate, and optimize tasks that humans have historically done, such as speech and facial recognition, decision making, and translation.One of the tasks which can be achieved by AI is computer vision, which is the ability for computers to process and analyse images, aiming to mimic human vision. One of the main tasks of computer vision is image classification, which is the process of labelling images into “classes”. For example, if there are images of multiple objects, and those images need to be categorized into “classes”, for instance “car”, “plane”, “ship”, or “house”, that is image classification.One common way to execute image classification is through convolutional neural networks, a technique implementing deep learning, which is a subset of machine learning, which is in turn a subset of AI.Neural Networks make up the backbone of deep learning algorithms.In fact, it is the number of node layers, or depth, of neural networks that distinguishes a single neural network from a deep learning algorithm, which must have more than three layers.Figure 2.25

2.3.1 Artificial Neural Networks

As the name suggests the basic idea comes from the so far understanding we have on our own human brain and the way it’s processing the new information-data.So, how does our brain think: When we receive an external stimulus like vision or sound, data travels as electrical signals through a path between neurons. The specific path is determined by

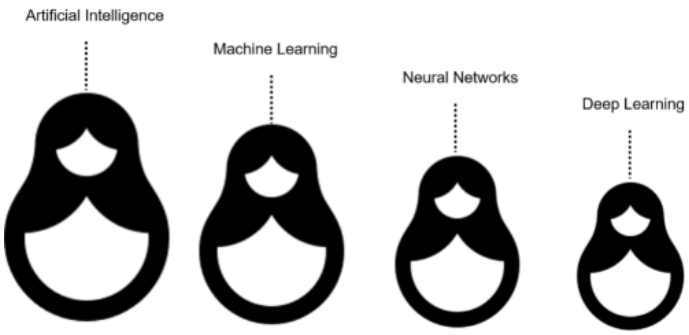


Figure 2.24: *Simple Visualisation between AI,ML,NN and DL*

the strength of inter-neuron connections, which itself is a cumulative result of all previous learning experiences[26]. A neuron may get signals from many other neurons. If the sum of all input signals crosses its activation threshold, it transmits the signal to the next connection; otherwise the signal dies at that neuron. Thinking essentially involves taking the information from input neurons, progressively abstracting it through multiple connections among 'thinking neurons', finally leading to muscle instruction by output neurons. Our ability to abstract from raw information is a key attribute of intelligence.

Now visualize the neuron as a computational unit, the strength of inter-neuron connections as a quantitative weight for signals into each computational unit, and activation threshold as a quantitative threshold for releasing information, and you have the basic building blocks of ANN. The computational unit in a neural network is called a 'node' and, unlike the 3-dimensionally packed neurons, nodes are connected to each other in layers – an input layer for getting information, an output layer for generating results and multiple 'hidden layers' in between for processing. Every node assigns a 'weight' to the connection from an incoming node, and its output is the weighted sum of data from each incoming node. If the weighted sum is above the quantitative threshold, the node fires its output to the next connected node. In the active state, each layer of ANN takes information, combines it into next level of abstraction, and then passes it to the next layer, until the information reaches the highest level of abstraction at the output layer. The larger the number of layers, the deeper the network is, and hence the phrase 'deep learning'.

Unlike the brain, ANN randomly assigns the weights and thresholds in the beginning. By running data for known outcomes, weights and thresholds are optimized to minimize the error. This 'try try again, until you succeed' approach (called supervised learning) is continued until the correct output is achieved, at which stage the ANN is deemed 'trained'.

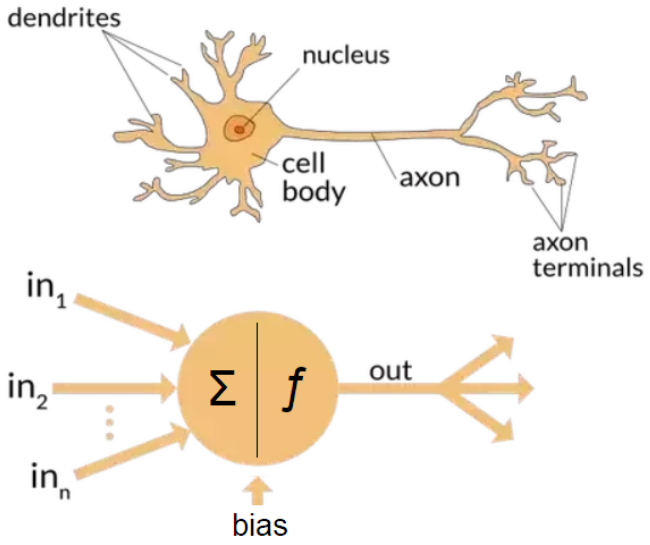


Figure 2.25: *Human Neurons and Artificial Neurons*

2.3.2 Generic Architecture for any type of Neural Network

An artificial neural network (ANN) is a set of layers of neurons (in this context they are called units or nodes). In the case of a fully connected ANN, each unit in a layer is connected to each unit in the next layer.

There is an input layer, where the network takes all the information needed, in this case the images to classify.Between the input layer and the output layer are hidden layers. Each hidden layer is used to detect a different set of features in an image, from less to more detailed. For example, the first hidden layer detects edges and lines, the second layer detects shapes, the third layer detects certain image elements, for example a face or a wheel.The output layer is where the network makes predictions. The predicted image

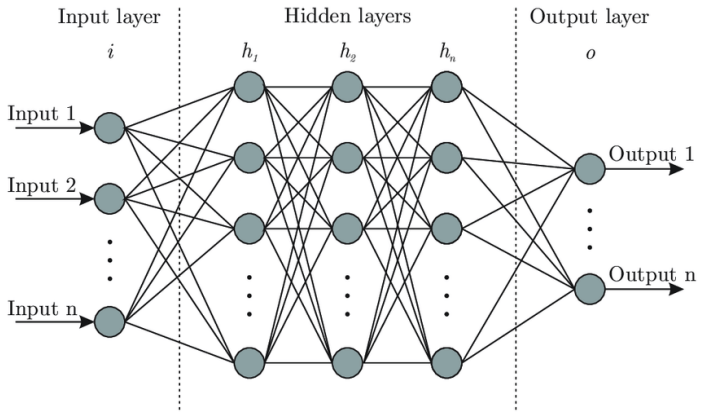


Figure 2.26: *An example of a ANN(Fully-Connected Feedforward NN)*

categories are compared to the labels provided by humans. If they are incorrect,the network uses a technique called backpropagation to correct its learning, so it can make guesses more correctly in the next iteration. After enough learning, a network can make classifications automatically without human help.

2.3.3 Artificial Neuron(Perceptron)

An artificial neuron is a connection point (unit or node) in an artificial neural network and has the capability to receive and process input signals and forward an output signal to another node.

2.3.4 Weights ,Bias and Activation Functions

2.3.4.1 Weights The connections between the units in a neural network are weighted, meaning that the weight indicates how much influence the input from a previous unit has on the output of the next unit. To mathematically compute an artificial neuron, all the products of all the inputs ($x_1 - x_n$) and their corresponding weights ($w_1 - w_n$) are added, then a bias (b) is added to that sum, then the resulting value is fed into an activation function (f) to form the output

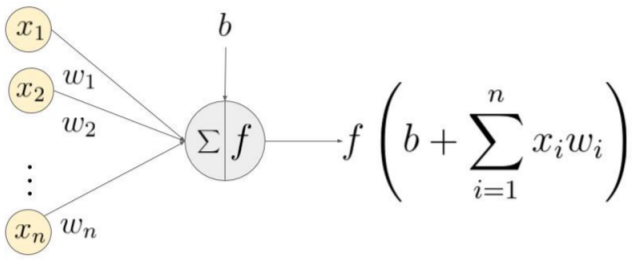


Figure 2.27: *Weights,Bias and Activation Function*

2.3.4.2 Biases A bias is an extra input to a neuron and it is technically the number 1 multiplied by a weight .The bias makes it possible to move the activation function curve left or right on the coordinate graph, enabling the neuron to create the required output value .

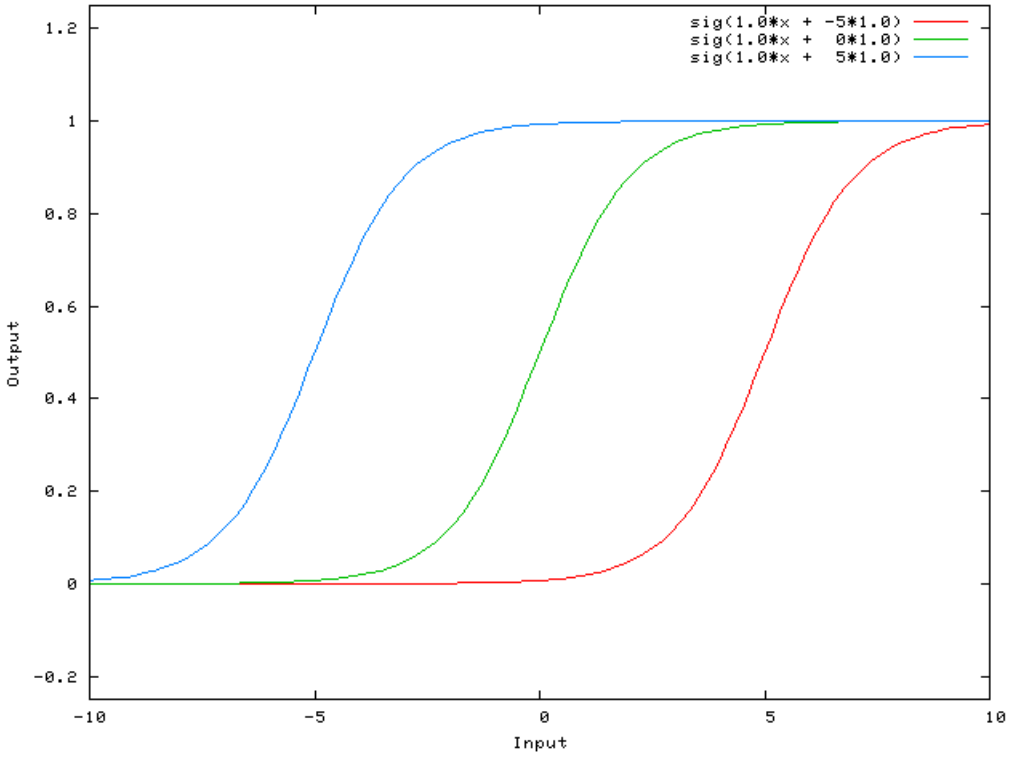


Figure 2.28: *Bias value , shift the activation function left or right*

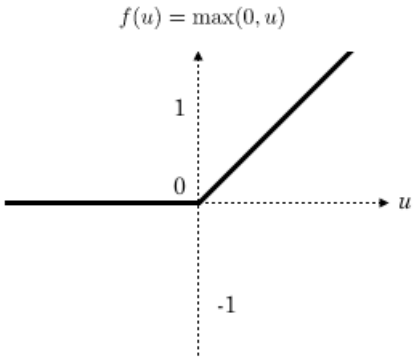


Figure 2.29: *Relu*

2.3.4.3 Activation Functions By definition, an activation function decides if a neuron should be activated ("fired") or not.It introduces non-linearity to the output of a neuron. A neural network without activation functions is just a linear regression model.The most common activation functions for CNNs are the Relu, Tanh and Sigmoid of which we choose ReLu : $f(u) = max(0,x)$. For values below zero the output is always 0 and for $x > 0$ the output is x.ReLu is less computationally expensive than the other activation functions because the mathematical operation is simpler and the activation is sparser. Since the function outputs 0 when x 0, there is a considerable chance that a given unit does not activate at all.Sparsity also means more concise models with more predictive power and less noise or overfitting. In a sparse network, neurons are more likely to process meaningful information. For example, a neuron which can identify human faces should not be activated if the image is actually about a building. One more advantage, which the ReLu possesses over the others, is that it converges faster. Linearity (when $x > 0$) means that the slope of the line does not plateau when x increases. Therefore ReLu does not have the vanishing gradient problem suffered by some other activation functions, such as Sigmoid or tanh.

2.3.4.4 Backpropagation Backpropagation is an algorithm which helps feedforward neural networks to train and learn their parameters , mostly from errors in predictions. This chapter is going to explain backpropagation using a gradient descent

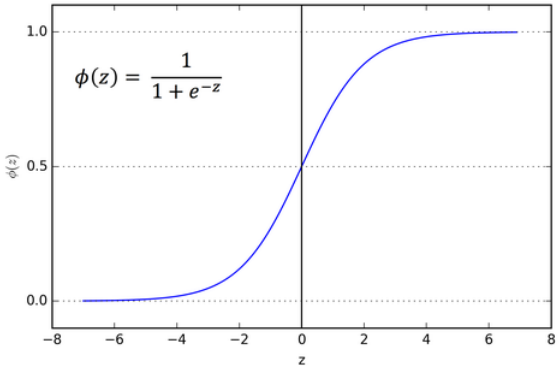


Figure 2.30: *Sigmoid*

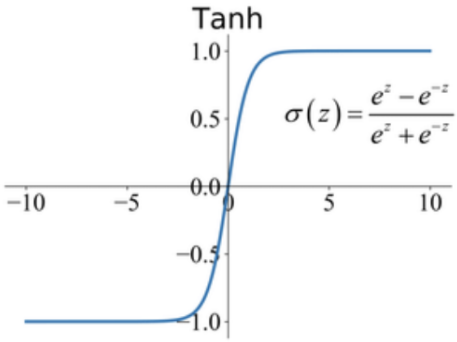


Figure 2.31: *Tanh*

2.3.4.5 Loss Function A loss function is an error metric, a way to calculate the inaccuracy of predictions. The aim of deep learning models is to minimize this loss function value, and the process of minimizing the loss function value is called optimization.

2.3.4.6 Gradient Descent A gradient descent is an optimization algorithm which modifies the internal weights of the neural network to minimize the loss function value. After each iteration, the gradient descent algorithm attempts to decrease the loss function value by tweaking weights, until the point where further tweaks produce little or no change to the loss function value, also called convergence .

2.3.4.7 Learning Rate and Adam Optimizer A learning rate is the step size of each iteration in the gradient descent or other optimization algorithms.A common value

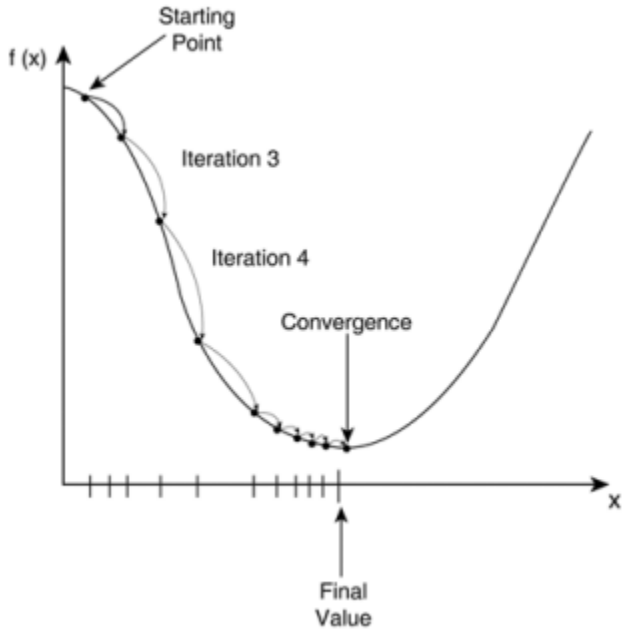


Figure 2.32: *Loss Function = f(x) ,x = output of Activation Function. We continuously tweak the weights by the "learning rate" and check the result.If the Loss function become smaller we continue until we find the minimum extremum.*

is 0.01 but LR is probably the most important hyperparameter of a NN so fine-tuning has to be done because every NN process different data.The LR values can vary from 10^{-6} to 1 .If the learning rate is too small, convergence will take a long time to happen, but if the learning rate is too large, there might be no convergence at all because we missed the extremum. There are many optimization algorithms, one of which is the Adam.The name Adam is derived from adaptive moment estimation , and takes the advantages of 2 others optimization algorithms(Root Mean Square propagation and Adaptive Gradient Algorithm) and combines them .A learning rate is maintained for each network weight (parameter) and separately adapted as learning unfolds. It's one of the best optimizers and great for computer vision , and the one we use for our CNNs.

2.4 State Of the Art

The constant Microscopy evolution has reached such a level that we can acquire images real time and with different imaging modalities. That great advance is used in many Science fields ,like Pathology, Biomedical, Forensics and many others.

A segmentation method was introduced using Magnetic Resonance imaging and X-ray computer tomography images of the human brain[27] and Convolutional Neural Networks were utilized to develop an adaptive learning scheme able to overcome inter-slice intensity variations typical of MR images. Multi-Modal imaging and CNNs were also combined for image Restoration using RGB guided depth imaging SR [28], a RGB guided multi-spectral imaging SR [29] and flash guided non-flash imaging for denoising purposes.[30]. Also at the same research they used techniques like Multi-Focusing imaging, Multi-Exposure imaging and Medical imaging and for Image Fusion with a network that is called the Common and Unique information splitting network (CU-Net) and it is designed to solve both problems together[31].

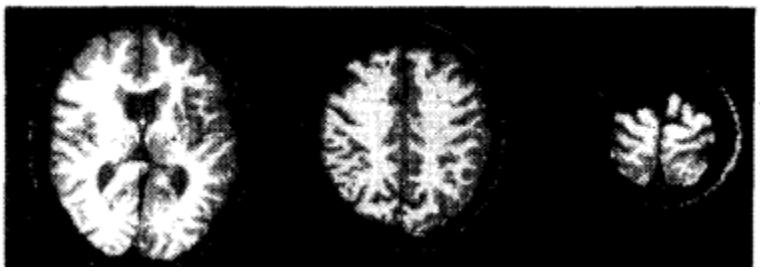


Figure 2.33: *Neural network segmentation of intensity slice 6 (left), slice 10 (middle), and slice 14 (right) of MR images.Source : [31]*

Convolutional Neural Networks also used in detecting recaptured images or distorted images, that was a difficult task so far with common techniques. One application for detecting recaptured images, consist of 1 camera capturing the real image and 8 other cameras used for recapturing. In the recapturing process, the composition, light, shaking, and other variables were strictly controlled[32].

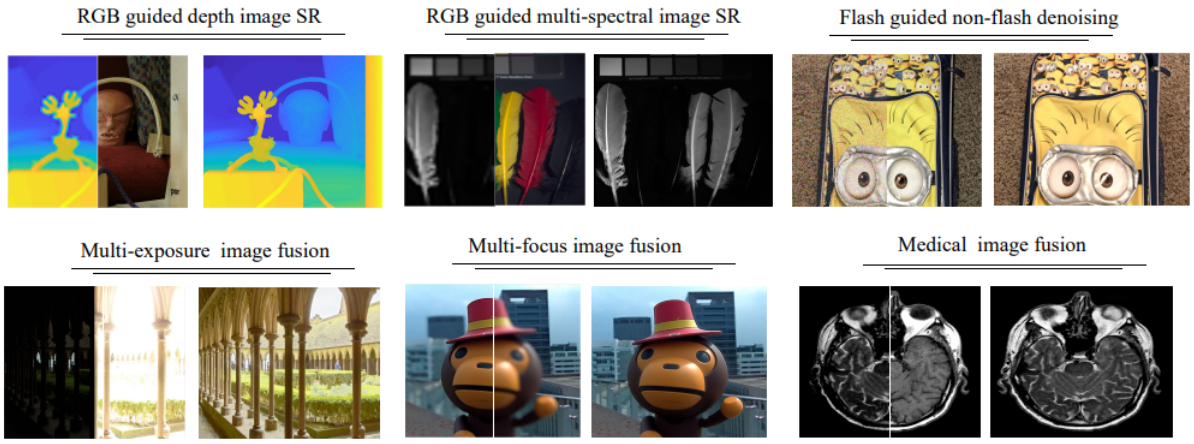


Figure 2.34: *Examples of different multi-modal image restoration and fusion tasks. The first row shows the MIR related applications, including RGB guided depth image SR, RGB guided multi-spectral image SR, and flash guided non-flash image denoising. The second row shows the MIF related applications, including multi-exposure image fusion, multi-focus image fusion, and medical image fusion.Source : [32]*

Chapter 3

System Design

3.1 Convolutional Neural Networks

An image is comprised of pixels. In the RGB model that displays the image with color, each pixel has three color elements, red, green and blue.In the grayscale model each pixel has only one color element that describes the ”grayness” of the image.Bit depth refers to how many bits of information are used to store color information in an image file.A colored image uses 8bits x 3channels = 24 bits for each pixel , and a grayscale image uses 8bits x 1channel = 8 bits.Most commonly each element can range from 0 (no color) to 255 (full saturation) in value.In digital image processing, a colored image is a three-layered matrix of pixels,where each layer is a two-dimensional matrix representing red, green or blue pixel values .The width and the height of an image is also counted in pixels e.g. 600x400 .

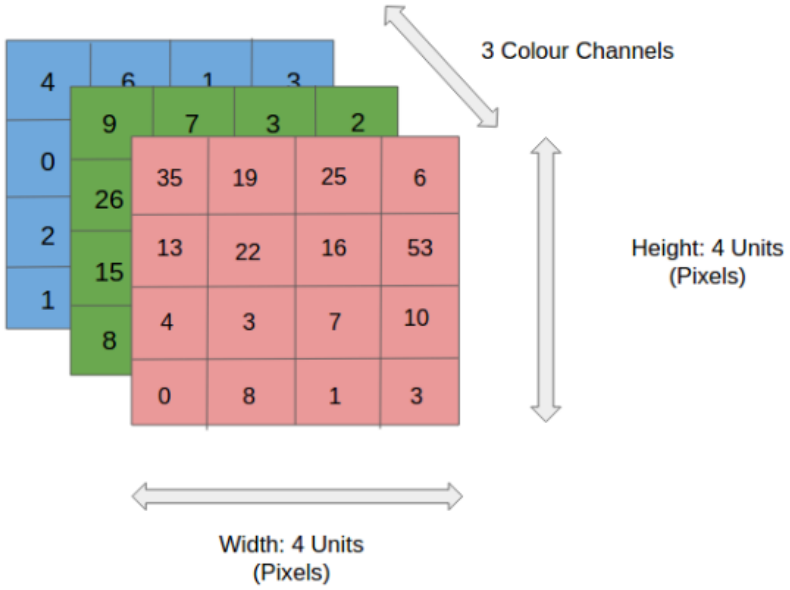


Figure 3.1: *The RGB matrixes that combine to produce a colored image*

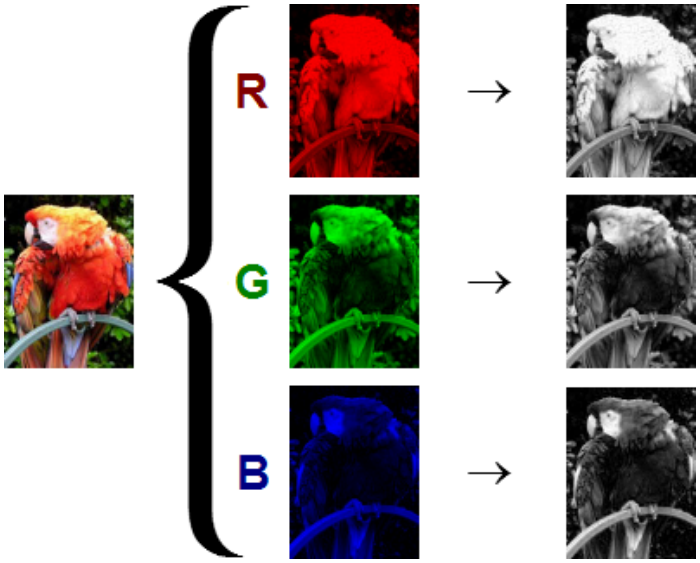


Figure 3.2: *A more eye friendly way to present the structure of a colored image*

3.1.1 CNN Architecture

A convolutional neural network (CNN) is a type of artificial neural network used primarily for image recognition and processing, due to its ability to recognize patterns in images.They also have many applications outside of image recognition and analysis, including image classification, natural language processing, drug discovery, and health risk assessments

All CNNs share a very similar architecture which is the input layer , then a number of convolutional layers depending on the difficulty of the task , pooling layers after some or every convolutional layer , then some Dense or Fully Connected layers that flatten the matrixes produced by the convolutional layers and at last a classification function depending on the number of the classes(objects) we wish to discrete. In simple word what CNN does is, it extract the feature of image and convert it into lower dimension without loosing its characteristics.

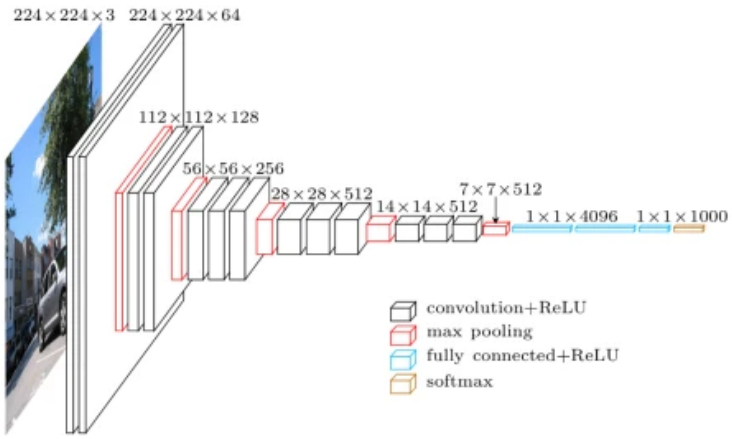


Figure 3.3: *VGG-16 Architecture Model.A very "deep" model that supports 16 Convolutional Layers,5 Max Pooling layers , 3 Fully Connected layers and a Softmax Activation.Visualizing the flow and the process of an image after each layer.*

3.1.2 Convolutional Layers

The convolution layer is the core building block of the CNN. It carries the main portion of the network's computational load. The term "convolution" refers to the mathematical combination of two functions to form a third function. When that happens, two sets of information are merged.Practically convolution uses a filter(or kernel) for edge detection , it detects the parts of the image with high contrast and creates a new 2-D matrix with those "edges" saved on it and we call it feature map.Another name for Convolutional layer is Feature Extractor.The number of filters and the size of them are defined by us in a suitable way depending on the size of the image.

The first Convolutional layer is responsible for capturing the Low-Level features such as edges, color, gradient orientation, etc. With added layers, the architecture adapts to the High-Level features as well, giving us a network which has the wholesome understanding of images in the dataset, similar to how we would.

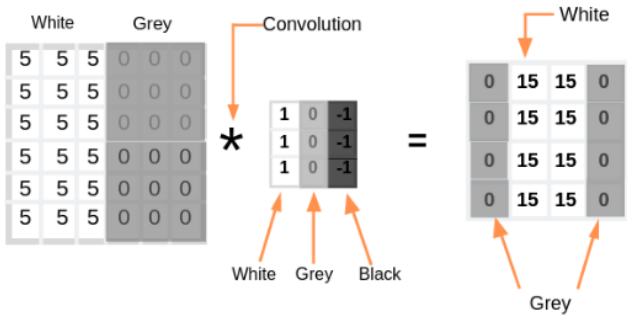


Figure 3.4: *Convolution between a 6x6 image which is half white and half gray and a 3x3 filter that amplifies the brightness and reducing the output image dimensions.Output may seem distorted but we only need to detect that there is an "edge" and save it to the feature map.*

The filter slides repeatedly through the input image , and creates a new feature map by calculating the dot product between receptive field(it is a local region of the input image that has the same size as that of filter) and the filter. Result of the operation is single integer of the output volume. Multiple filters are used for one input and the resulting feature maps are joined together for the final output of one convolutional layer.

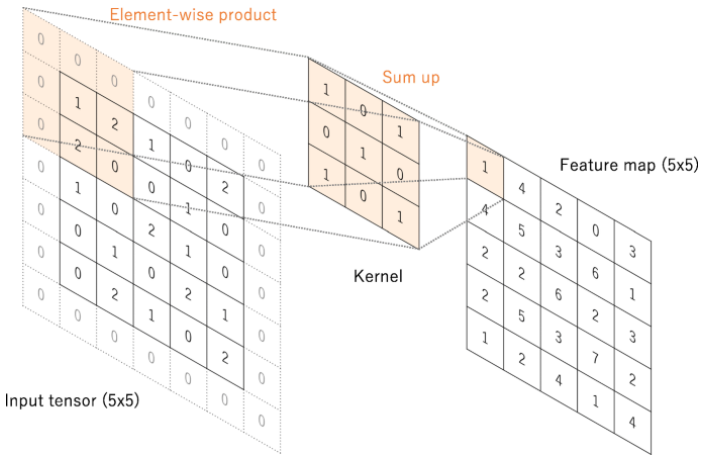


Figure 3.5: *Convolution process between Input image and kernel.*

3.1.3 Padding and Strides

There are two types of results to the convolution operation — one in which the convolved feature is reduced in dimensionality as compared to the input, and the other in which the dimensionality is either increased or remains the same. This is done by applying Valid Padding in case of the former, or Same Padding in the case of the latter.In our case we apply Same Padding. Another hyperparameter in every CNN model is the stride size of the kernel ,which by default is one , and this is also the stride we chose.

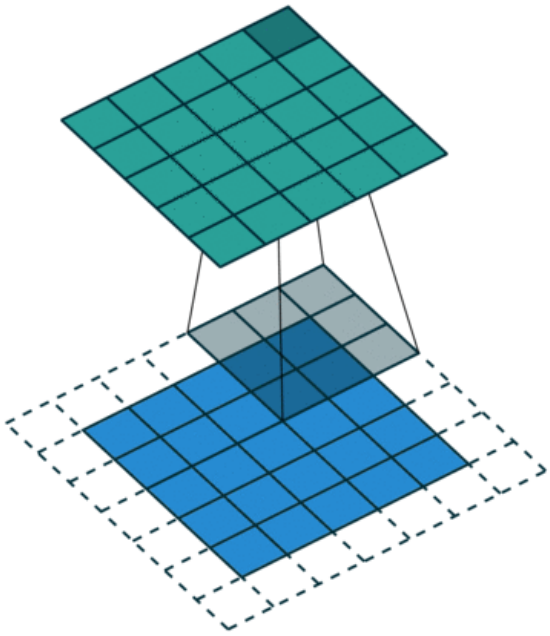


Figure 3.6: *Example of Same Padding with zeroes that preserve the width and height between input and output image*

3.1.4 Batch Normalization Layer

We place the Batch Normalization layer between Conv. Layers and Activation function.Batch normalization applies a transformation that maintains the mean output close to 0 and the output standard deviation close to 1.Standardizing the activations of the prior

layer means that assumptions the subsequent layer makes about the spread and distribution of inputs during the weight update will not change, at least not dramatically. This has the effect of stabilizing and speeding-up the training process of deep neural networks.

3.1.5 Dropout Layer

The Dropout layer randomly sets input units to 0 with a frequency that we choose, at each step during training time, which helps prevent overfitting and generalization of the model.That is because when important neurons are dropped out ,model has to find new strong features to base the classification to.Inputs not set to 0 are scaled up by 1/(1 - rate) such that the sum over all inputs is unchanged.In our model we apply Dropout layers only after FC layers with rate=0.3.

3.1.6 Leguralizers L1,L2

Regularizers allow you to apply penalties on layer parameters or layer activity during optimization. These penalties are summed into the loss function that the network optimizes.Regularization controls the models complexity and can prevent a Neural Network from overfitting or underfitting,and thus improve the accuracy of a Deep Learning model when facing completely new data from the problem domain.
Regression equation:

$$y = W * x + b \tag{3.1}$$

where x is the input,W are the weights and b is the bias.

- Kernel Reguralizer :Tries to reduce the weights W (excluding bias).
- Bias Regularizer: Tries to reduce the bias b.
- Activity Regularizer: Tries to reduce the layer's output y, thus will reduce the weights and adjust bias so Wx+b is smallest.

L1 versus L2 loss for the weight decay.

- L2 loss is defined as w^2.
- L1 loss is defined as $|w|$, where w is a component of the matrix W.
- The gradient of **L2** will be 2w.
- The gradient of **L1** will be sign(w).

Thus, for each gradient update with a learning rate a, in L2 loss, the weights will be subtracted by $a * W$, while in L1 loss they will be subtracted by $a * sign(W)$.The effect of L2 loss on the weights is a reduction of large components in the matrix W, while L1 loss will make the weights matrix sparse, with many zero values.

3.1.7 Pooling Layers

In most cases, a Convolutional Layer is followed by a Pooling Layer.Pooling layers are responsible for decreasing the dimentionality of feature maps and reducing the computational costs, specifically they are decreasing the height and width, preserving the depth.Doing so is beneficial because it decreases the required computational power to process the data, while extracting the dominant features in feature maps. There are two types of pooling layers: max pooling and average pooling .Max pooling is better at extracting dominant features and therefore considered more perfomant.

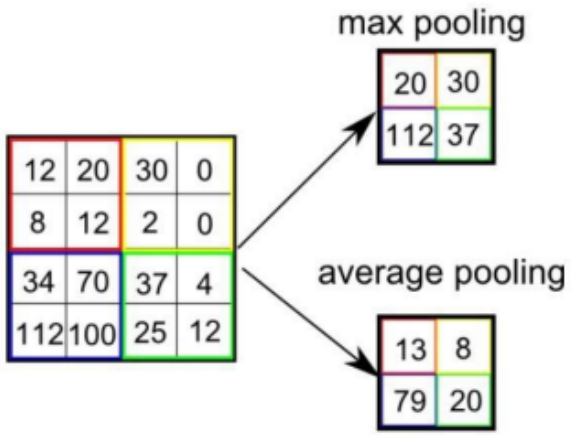


Figure 3.7: *Pooling layers.Max Pooling outputs the highest value of the elements in the portion of the image covered by the filter* $out = max(x_1, x_2, x_3, x_4)$*, while average pooling takes the average value of the four(* $out = (x_1 + x_2 + x_3 + x_4)/4$.

3.1.8 Fully Connected Layers

Fully connected layers are where classification actually happens. The input matrix is flattened into a column vector and is fed into a set of fully connected layers which are the same as the fully connected ANN architecture that was previously discussed in chapter. Each fully connected layer (called Dense layer) is passed through an activation function (e.g. tanh or ReLu), but the output Dense layer is passed through Softmax. In the Softmax multiclass classification, the loss function used is Cross Entropy (categorical crossentropy in Keras).

The output of the Softmax function is an N-dimensional vector, where N is the number of classes the CNN has to choose from. Each number in this N-dimensional vector represents the probability that the image belongs to each certain class. For example, if the output vector is [0 .1 .1 .75 0 0 0 0 0 .05], then there is a 10 percent probability that this image belongs to the class 2, 10 percent probability that it belongs to the class 3, 75 percent probability that this image belongs to the class 4, and 5 percent probability that it belongs to the class 10.

3.2 Supervised Learning

Supervised learning uses a training set to teach models to yield the desired output. This training dataset includes inputs and correct outputs, which allow the model to learn over time.There are many machine learning algorithms we can choose from.We chose Neural Networks ,which are suitable for the task, as they function like feature extractors and reduce dimensionality of the images.Thus,dataset is probably the most important part behind a CNN and great results.It has to be accurate labeled,every image that is fed to our network must contain only the object we assigned to it for proper training.It may contain only a part of the object,or contain noise for better generalization,for this job we insert a data augmentation layer provided by Keras,our framework tool,to create new images slightly or 90 degrees rotated for our model to be able to recognise the same object even if its upside down or in a random position.

Creating a capable and big enough dataset for object classification with Neural Networks, is extremely time consuming, and depending on what you want to classify, special equipment is required.For each different object,for your CNN to be robust needs multiple images from the same object from many angles, different parts and the whole image, blurred and with noise etc,approximately at least 200 images for each object. In this thesis, we have 5 discrete objects :

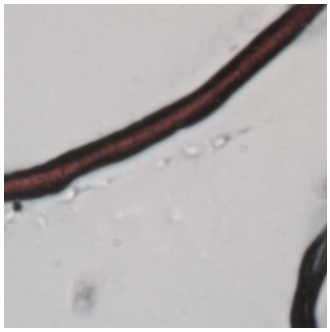
- **Blood**
- **Fiber**
- **Sand**
- **Glass**
- **Skin**

3.3 Dataset Acquisition

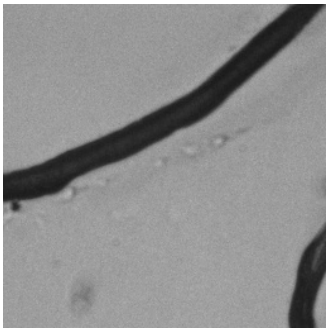
Firstly we have to gather the data that we need, luckily for this research on CNNs we have available a very large and unique dataset which was produced by the microscope's data acquisition process consists of :

- **Transmission.(RGB,Spectral Cube13 Bands)**

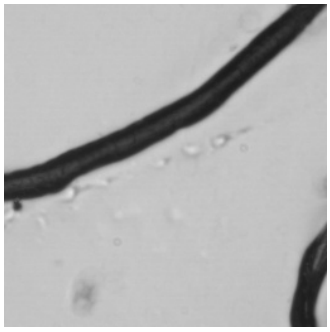
Transmitted light microscopy is the general term used for any type of microscopy where the light is transmitted from a source on the opposite side of the specimen from the objective.Besides colored(RGB) display,with optical band-pass filters we can capture the same image in different wavelengths in the Electromagnetic spectrum with HSI.



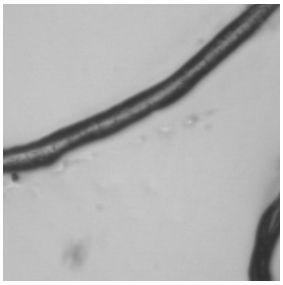
(a) *Image of a fiber captured using brightfield microscopy.*



(b) *The same fiber captured using HSI microscopy at 325nm.*



(c) *The same fiber captured using HSI microscopy at 515nm.*

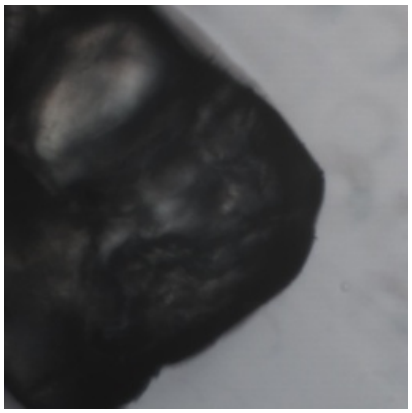


(d) *The same fiber captured using HSI microscopy at 980nm.*

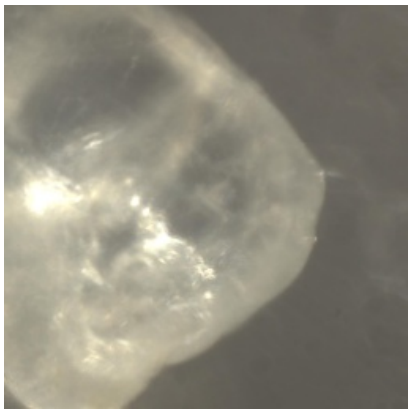
- **Reflection(UV,Visible)**

In reflected light microscopy, illuminating light reaches the specimen, which may

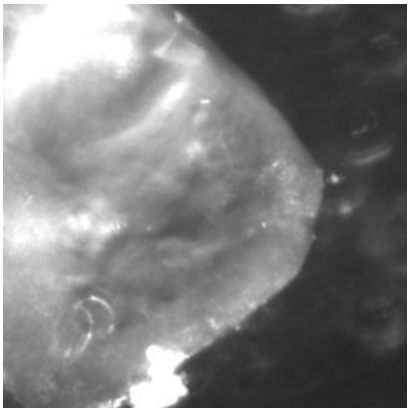
absorb some of the light and reflect some of the light, either in a specular or diffuse manner.



(a) *Image of a sand captured using brightfield microscopy.*



(b) *The same fiber captured using HSI microscopy at 325nm.*

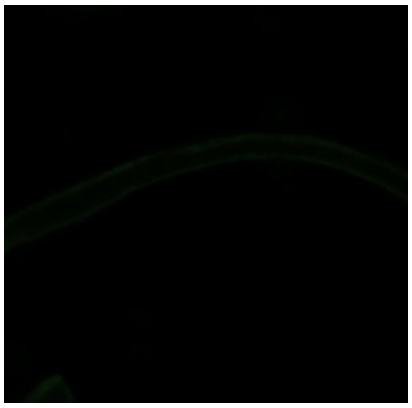


(c) *The same fiber captured using HSI microscopy at 515nm.*

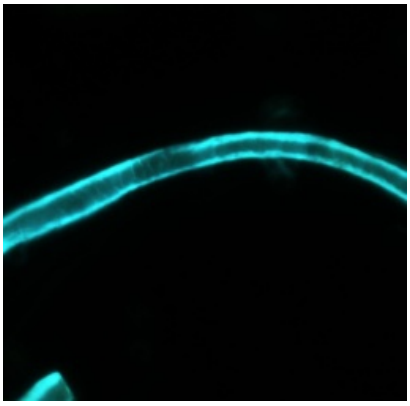
- **Fluorescence(spectrum bands:365,405,450)**.

Fluorescence microscopy is an imaging technique used in light microscopes that allows the excitation of fluorophores and subsequent detection of the fluorescence signal. Fluorescence is produced when light excites or moves an electron to a higher energy state, immediately generating light of a longer wavelength, lower energy and different color to the original light absorbed.

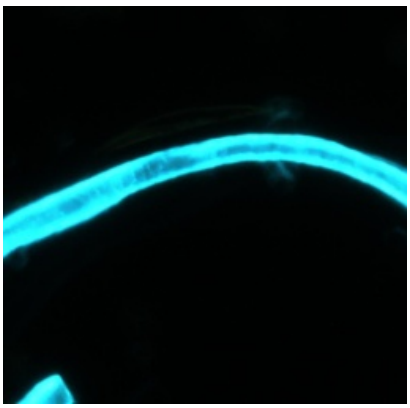
(a) *Image of a sand captured using brightfield microscopy.*



(b) *The same fiber captured using Fluorescence microscopy with a light source at 365nm.*



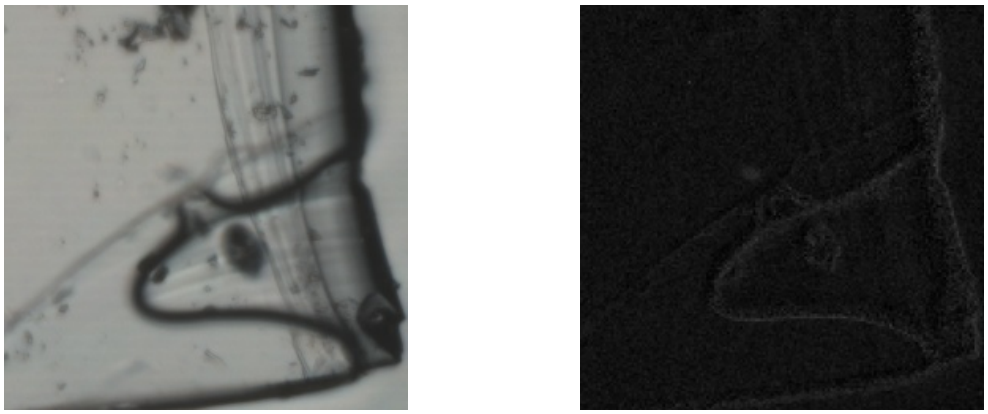
(c) *The same fiber captured using Fluorescence microscopy with light source at 405nm*



(d) *The same fiber captured using Fluorescence microscopy with light source at 450nm.*

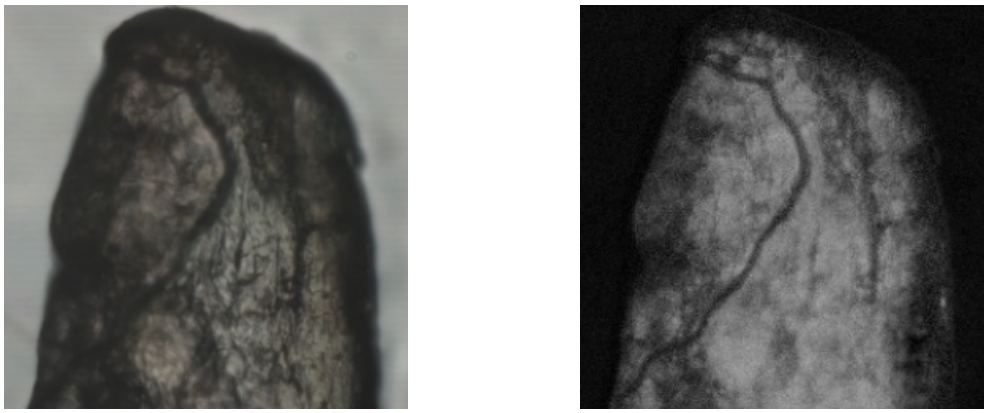
- **Polarization(Degree of linear Polarization)**.

Polarized light is a contrast-enhancing technique that improves the quality of the image obtained with birefringent materials.



(a) *Image of glass captured using brightfield microscopy.*

(b) *The same sand image captured using Polarized Microscopy.*



(a) *Image of a skin captured using brightfield microscopy.*

(b) *The same skin part captured using Polarized Microscopy.*

Figure 3.12: *Glass and skin under light and polarized micorscopy.As we can see polarized microscopy works like a contrast enhancing technique.*

In total there are 20 different displays of the same image and even though we can perform classification with Convolutional Neural Network in each display individually, our desired outcome would take in account all the valuable information of each different

display.We want to achieve this multi-layer classification with only the use of Neural Networks, which means that the output of one or more CNNs will be the input to another NN, and in the next chapters we will elaborate how the final design works and why some approaches do not work.

Chapter 4

Implementation

4.1 System Configuration

Training a CNN is computationally heavy, so the **hardware(GPU)** our system has,is important to determine the numbers of layers and the number of convolutions we can apply, as well as the batch size which we define in the training process.Our initial state attempts to create Convolutional Neural Network were processed by:

- GPU:NVIDIA RTX 750 TI 2GB , max memory capacity 1.7GB

But later we evolved our GPU to:

- NVIDIA RTX 2010 TI 4GB , max memory capacity 3.7GB

and so we could also develop stronger networks with more convolution in each layer and increase the batch size to fasten the whole process.

4.1.1 Tool for dataset extraction

The dataset that we will feed on our Neural Networks has to be accurate, which means that the same image but in different mode will be fed to each CNN. Our CNN's output is a **matrix of possibilities** as many as the quantity of the objects we want to classify, and the biggest possibility of this matrix is the decision of the CNN about what object does it sees.So we created a tool that acquire all images simultaneously and saves them in different folders.

For this tool, we convert the image into grayscale and then we apply a threshold between [1-255] which are the range of the pixel's value ,and any pixel above threshold

we ”paint” it white,and the rest of them black, so we can distinguish them easily.The white pixels that are left on the image, are our object, and our region of interest for this task.

Training any NN is expensive computationally , so we crop our images into smaller tiles that we can examine and process easier. The size that we crop our dataset for training is the same size with the images we are going to feed to our CNN for the predictions and it is **256x256** pixels.We use the RGB image to extract the contour's coordinates, and with the same coordinates we crop the images in every different mode, then we label the images before saving them.

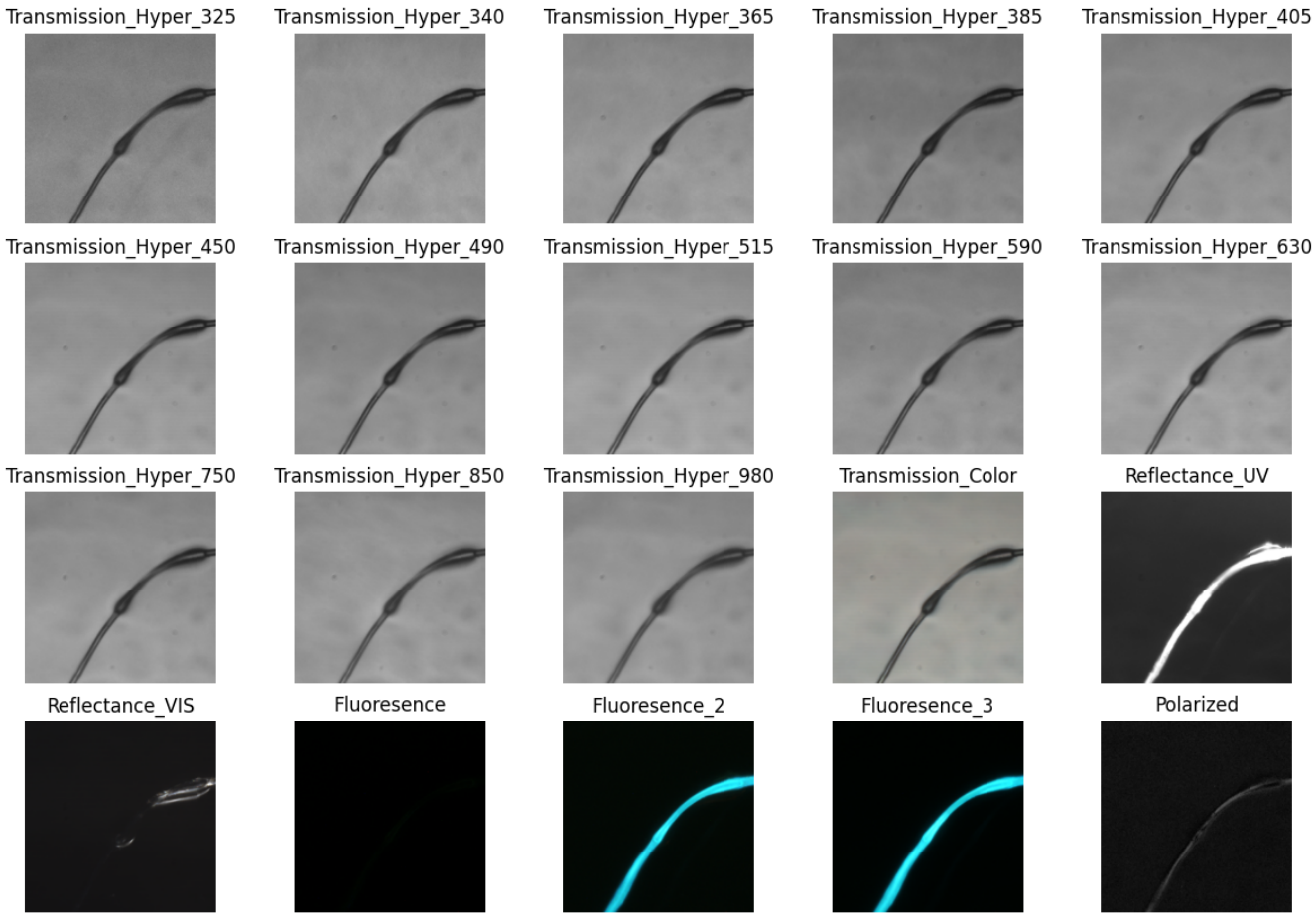


Figure 4.1: *A fiber under all different microscopy techniques.*

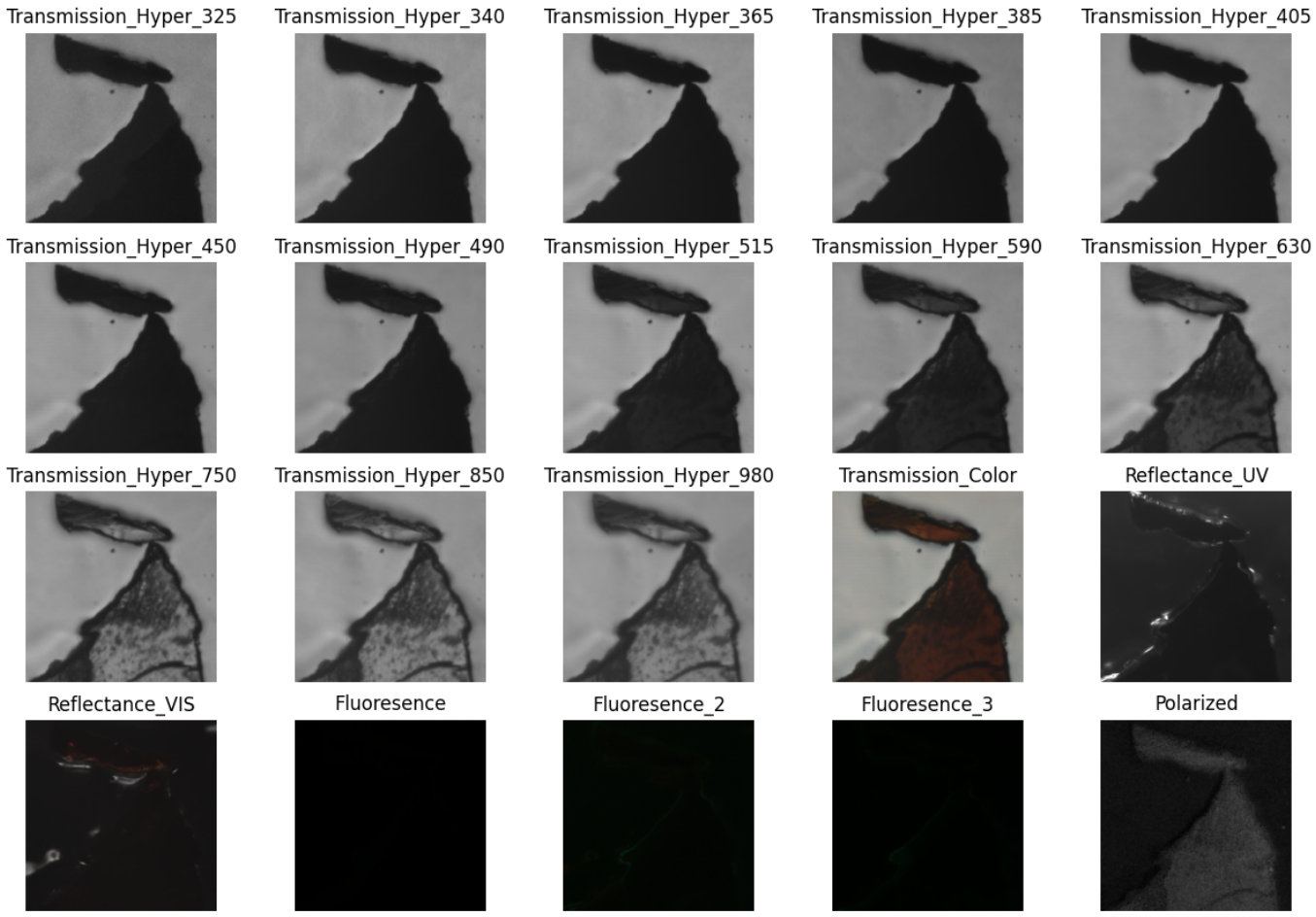


Figure 4.2: *Blood under 20 different image enhancing modes*

In figure Figure **??** we can see the output of that extraction tool, and we can notice that there are some very unique features, for example fiber is extremely fluoresce and so, very distinguishable and blood almost does not fluoresce at all.Nevertheless the major distinctions between the variety of displays we can acquire from the objects have to be taken account for our final prediction, because it will drastically affect the accuracy of our system.

As we previously mentioned, the architecture of CNN lookalike, the differences are found in the number of convolutional layers, the convolutions that take place in each layer, the kernel size for each convolution, the padding that is used, which pooling layer(Max, Average) we choose,and some features that can be inserted between layers, like Batch

Normalization, the use of Regularizers(L1,L2 or both), the Dropout Layers and the number of neurons we use for our Dense Layer. Of course, there are many other features that can be examined and implemented in a CNN but this is how far i searched for this thesis and assume its enough for a satisfying output.

4.1.2 Hyperparameters

Hyperparameters are a set of parameters that organize the whole process and every neural network that has a different task, will need to ”tune” the parameters to fit the task.They exist in the convolutional layers as the **number and size** of the kernels, the **size of the stride** and the **size of the kernels** in the pooling layer.**Pooling type and size** are also two important parameters, as they define the way and how much the image will be downsampled[33].Our kernel size is 3x3, and number of kernels is defined in each layer, size of our stride is 1.Pooling kernel size is 2x2 and we use the max pool technique that keeps the maximum value between the four pixels.

Also, **learning rate** that tunes the weights after backpropagation ,even though it is not considered a hyperparameter it needs tuning because as we mentioned before,if learning rate is too large can cause the model to converge too quickly to a suboptimal solution, whereas a learning rate that is too small can cause the process to get stuck[34].Sometimes if learning rate is too big then every few epochs evaluation accuracy after training will drop significantly because network assumes it has find the maximum extremum.An error like this occured while training with a learning rate of $10^{-}4$ and the graph with the train accuracy, validation accuracy, validation loss and train loss is presented in (a).

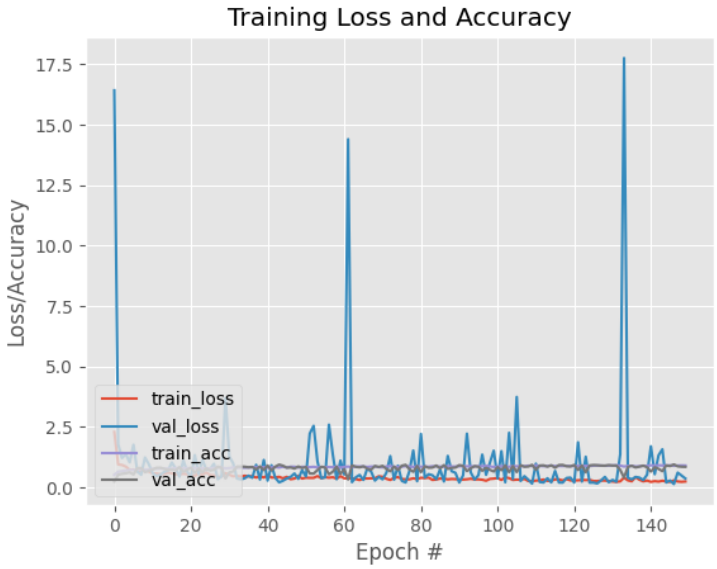


Figure 4.3: *Metrics graph for our CNN with learning rate = 0.001.There are huge spikes on evaluation accuracy.*

Figure 4.4: *Metrics graph for our CNN with learning rate set to 0.0001.Validation accuracy and loss, respond to a 15% of our total dataset which we split before the training phase.We do this split before the training phase because we want to keep those images for validation process, so the CNN will never train on those images to check if the classification works.TrainLoss:Loss function's output,in our case sparse categorical crossentropy which Computes the crossentropy loss between the labels and predictions.Maximum value is one which indicated that there is no loss,Accucary:Correctly classified images/All images*

4.1.3 Preprocessing

Before the images reach the first convolutional layer we place an rescaling layer that the whole image goes through so all the pixels get a value between 0 and 1 to reduce the computational cost, this technique is also called **normalizing** the data.Afterwards we place another couple of layers and each one will create a new image from the original image, but rotated by 90 degrees vertical or horizontal and another image rotated by 30 degrees.This way we produce three times the quantity of the original dataset, and we make our model even more robust.

Also in the preprocessing field is the Batch Normalization function that we apply after each convolutional layers and that we explained in chapter 2.

4.1.4 CNN Features Between Layers

Regularizers are placed also after convolutional layers.We use L1 and L2 regularizers but with a very small value of 0.0001.As we mentioned before, leguralizers are placed on the weights to make the really small values zero with L1 regularizer, and also reduce the values that are so high who may affect the decision of the network,With regularizer L2, by doing so we prevent overfitting and help the generalization of our model.

As we mentioned, after all convolutional layers we need the activation function, in our case we use rectified linear activation function or ReLU.

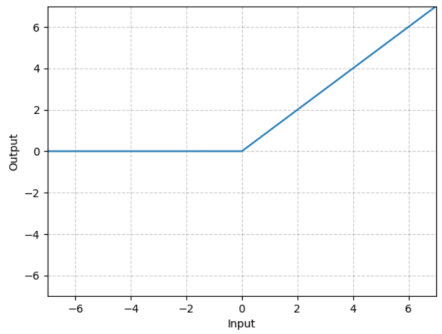


Figure 4.5: *Rectified Linear Activation*

After we pass through all convolutions and flatten the **feature map** from the image we apply a dropout layer of 20%.Dropout means that we shut down randomly the 20% of our neurons, in order to prevent overfitting, by eliminating the most dominant feature the model will find other strong features to classify the objects so we making the model independent from only one feature.

Our architecture for the initial classification we performed before the final structure looks like this:

Layers	Output Shape(width,height, convolutions)	Parameters
Input Image	(256,256,3{channels})	-
Conv2D	(256,256,128)	3584
MaxPooling	(128,128,128)	0
Conv2D	(128,128,128)	147.584
MaxPooling	(64,64,128)	0
Conv2D	(64,64,256)	295.168
MaxPooling	(32,32,256)	0
Conv2D	(32,32,256)	590.082
MaxPooling	(16,16,256)	
Flatten	(65.536)	0
Dense	(512)	33.554.944
Dense	(Number of Classes=11)	5643
Total Parameters	8.767.560	-

Figure 4.6: *The architecture of the CNN model one step before the final structure.The number of parameters for any Conv. layer are calculated like this : (filter shape =3*3, stride=1) layer is: ((shape of width of filter * shape of height filter * number of filters in the previous layer+1) * number of filters) ,e.g parameters for second Conv2D : (((3*3*128)+1)*128) = 147.584.For Dense(FC) layer are calculated like this:((current layer c*previous layer p)+1*c),e.g first Dense layer: 512*(16*16*64)+1*512 = 8.389.120.*

Every layer of filters is there to capture patterns. For example, the first layer of filters captures patterns like edges, corners, dots etc. Subsequent layers combine those patterns to make bigger patterns (like combining edges to make squares, circles, etc.).
Now as we move forward in the layers, the patterns get more complex, hence there are larger combinations of patterns to capture. That's why we **increase the filter number** in subsequent layers to capture as many combinations as possible.

The **number of parameters** in a given layer is the count of “learnable” elements for a filter aka parameters for the filter for that layer, genuinely any layer that contain weights is considered ”learnable”.Parameters in general are weights that are learnt during training. They are weight matrices that contribute to model’s predictive power, changed during back-propagation process.The changes are depended to the training algorithm we choose.

Our final architecture of CNN after the Hardware development on our system is shown in Figure Figure 4.7:

Layers	Output Shape(width,height, convolutions)	Parameters
Input Image	(256,256,3)	-
Conv2D	(256,256,128)	3584
MaxPooling	(128,128,128)	0
Conv2D	(128,128,128)	147584
MaxPooling	(64,64,128)	0
Conv2D	(64,64,128)	147584
MaxPooling	(32,32,128)	0
Conv2D	(32,32,64)	73792
MaxPooling	(16,16,64)	
Flatten	(16384)	0
Dense	(512)	8389120
Dense	(Number of Classes=8)	4104
Total Parameters	8,767,560	-

Figure 4.7: *Final architecture of CNN models.*

4.1.5 Splitting the Dataset

Before we start the training process we have to split the dataset into 3 parts.

- **Training Dataset 70%**
- **Validation Dataset 15%**
- **Test Dataset 15%**

Training dataset is utilized in the training process we described above,convolution,weight assignment and update,backpropagation etc. **Validation** dataset is used after the training in each epoch to measure the accuracy with unknown images, the only issue is that this validation is happening in the background of the training process, and we cannot see actually the labeling prediction but only a probability for accuracy and the results of our loss function.This is why we need **Test** dataset, which also is unknown for the CNN. After we save the model's architecture and final weights and load them again we use the predict function with given input the test dataset and we evaluate the result with real images.

We gathered and used two main datasets because the first dataset was not complete with all our final enhancing techniques and the number of images in each class was a little unbalanced, so both dataset consistency is shown below on Table subsection 4.1.5. Thirty percent of this dataset is not used in the training phase, and even though we randomly select the images we are going to use for test and validation practically 30$ of those numbers does not go to the training process.So dataset is even more unbalanced, and we can not be sure about the results on classifying the four last classes.

Our dataset consists of different specimens we captured through the microscope, and some of this specimens belong to the areas we are going to classify.So, except from individually images from the microscope, we acquired approximately **50%** of the objects from the classifying area and fed it into our CNNs, but 30% of it will be distributed in validation and test dataset, so finally approximately **35% of the overall specimens** exist in our first classification area will be processed by the CNNs.

First Dataset	Second Dataset
• Blood 47 images	• Blood 320 images
• Fiber Cotton 1027 images	• Fiber 406 images
• Fiber Highly Fluorescic 417 images	• Glass 282 images
• Fiber Polyester 336 images	• Hair 345 images
• Glass 107 images	• Sand 759 images
• Hair 389 images	• Skin 308 images
• Sand 80 images	• Bubbles 444 images
• No Object 420	• No Object 154 images
• Fiber Unknown 13 images	
• Multi-Object 17 images	
• Bubbles 139 images	

4.2 Single Modality Classification Methods

4.2.1 Transmission Color Approach

The most common thought around designing a CNN classifier, is to train the network with the most appealing to the human eye imaging method. To reach our first practical results we trained a CNN model to classify objects acquired utilizing **Transmission Color Imaging(Whitefield)**. This CNN model was trained with 11 classes of different objects:

- Blood
- Fiber Polyester
- Fiber Highly Fluorescic
- Fiber Cotton
- Hair
- Glass
- Skin
- No Object
- Bubbles
- Fiber Unknown
- Multi-Object

The **No object** and the **Bubbles** classes are needed for the final construction of the images we will classify.When our higher prediction is for any of those two classes we can skip the process of drawing the contours on the image.Also, bubble class functions as a noise extractor.When an image contains both bubbles and an object the CNN can identify the differences between that image and a image with plane bubbles.

As we can see from figures Figure 4.8 and Figure 4.9 the results are promising, as we achieve average confidence of **93%** with only a few mistakes and small confidence predictions that we assume its from the unbalanced dataset.

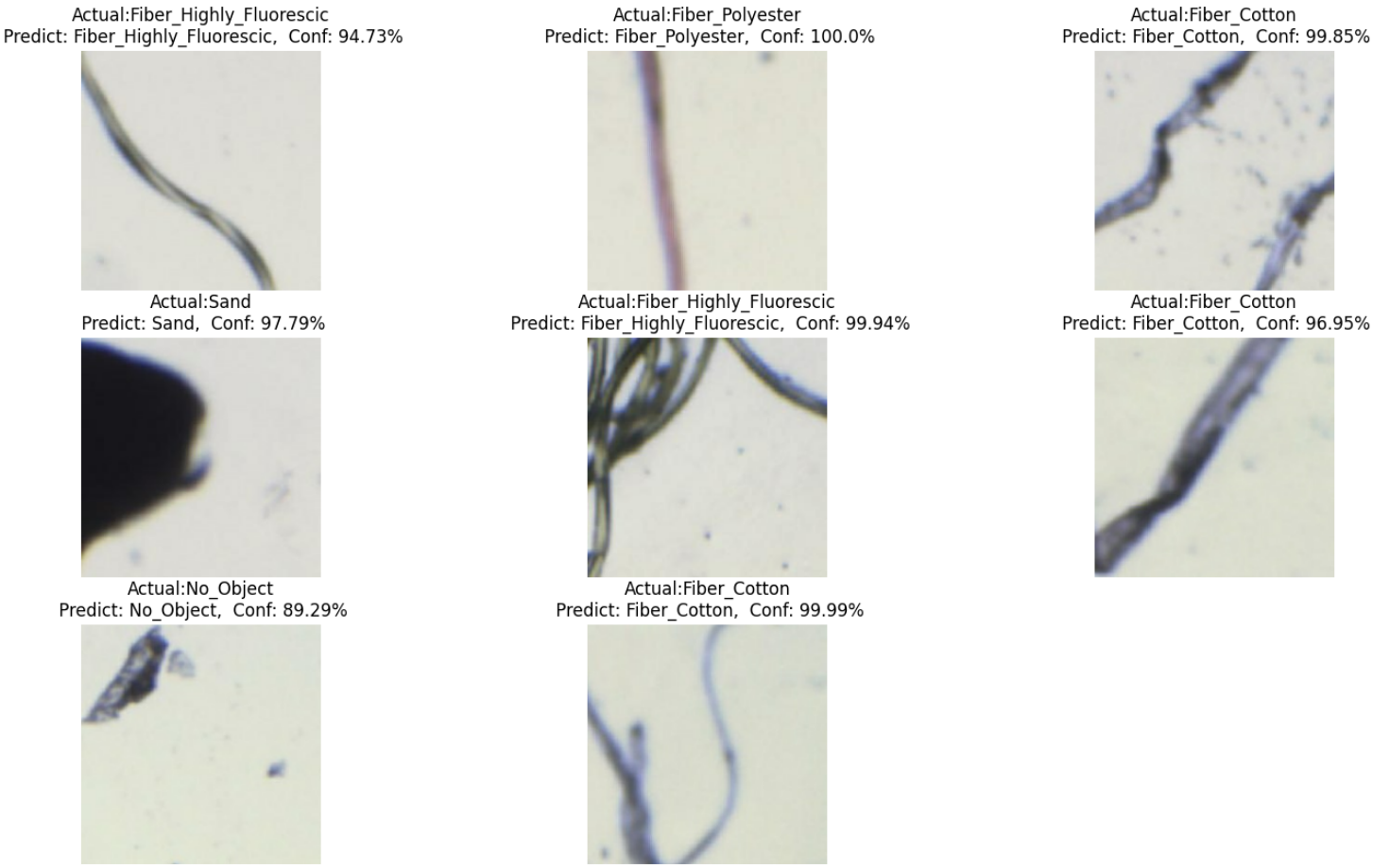


Figure 4.8: *Conf = Confidence.Testing the accuracy with images the CNN haven't seen before. Every prediction was correct with excellent accuracy.Even though the shape of the different kind of fibers are the same, and the black color that cotton and highly fluorescic fibers have are the sane, CNN can actually distinguish them!*

The actual label for this objects has been given when we were extracting the dataset with our tool, and when we took the testing dataset we also kept the label for each image. The predictions we get from our model are for each image individually and it is a matrix with n = number of classes elements, with the probability for each element.Some values are very small e.g. $10^{-}4$ so we skip them,and the **output of our trained CNN** with the

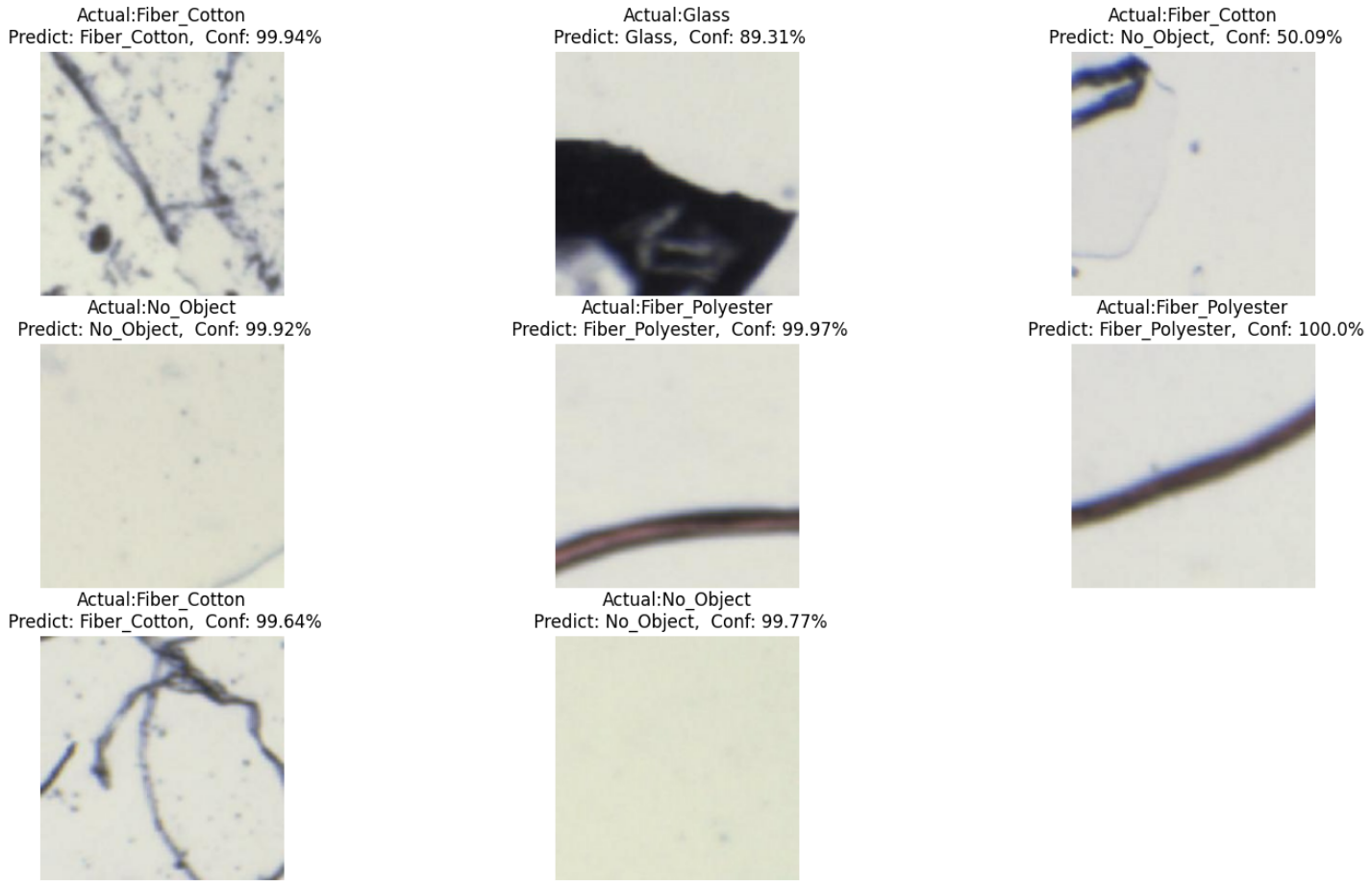


Figure 4.9: *Even though the prediction in the top-right corner is wrong,the confidence it has for that label is really low,so we can simply not acknowledge that prediction, until our Neural Network is more confident about its predictions by correcting and expanding that dataset or optimizing the hyperparameters.*

colored image dataset when we feed it a new unseen image for prediction looks likes this:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0.99 & 0.01 & 0 & 0 & 0 & 0 \end{pmatrix}$$

We can decode the matrixes we take from the CNN by knowing that the classes are distributed in **alphabetical** order from their names, first element is the probability for the image to be classified as ”Blood”, which is zero and as we continue we find that the fourth element corresponds to ”Fiber Polyester” class with 99% prediction probability. The metrics graph for this trained model is given below:

Figure 4.10: *Graph for the training process of the colored dataset.*

We can see that training and validation accuracy approximately 93% and training and validation loss is slowly stabilizing around one point and despite a few spikes there is not over or under fitting.

4.2.2 Training with Different Modalities

We have mentioned before the tool for extracting the dataset l which can crop and save all 20 images of the same object, each image captured with a different technique or with the same technique in a different wavelength.So we have 20 unique datasets, with a total number of number of images in the colored dataset*20 = 57.060 images.

Before we find the way to combine all those images and predictions into a single and final prediction we **train 20 individual CNN** each one trained with images from different technique.The following categories are all the different displays we can acquire:

- Color RGB image.
- **Hyperspectral images** with different wavelength bands:
  - from 325nm to 339nm
  - from 339nm to 364nm
  - from 365nm to 384nm
  - from 385nm to 404nm
  - from 405nm to 449nm
  - from 450nm to 489nm
  - from 490nm to 514nm
  - from 515nm to 589nm
  - from 590nm to 629nm
  - from 630nm to 749nm
  - from 750nm to 849nm
  - from 850nm to 979nm
  - from 980nm to 1100nm

- Fluorescence images with the emission wavelength at:
  - 365nm
  - 405nm
  - 450nm
- Image captured under **polarized** microscopy
- Image captured under **reflectance** microscopy on the **visible** spectrum
- Image captured under reflectance microscopy on the **ultraviolet** spectrum

Below on figures Figure **??**we can see the results from our CNNs:

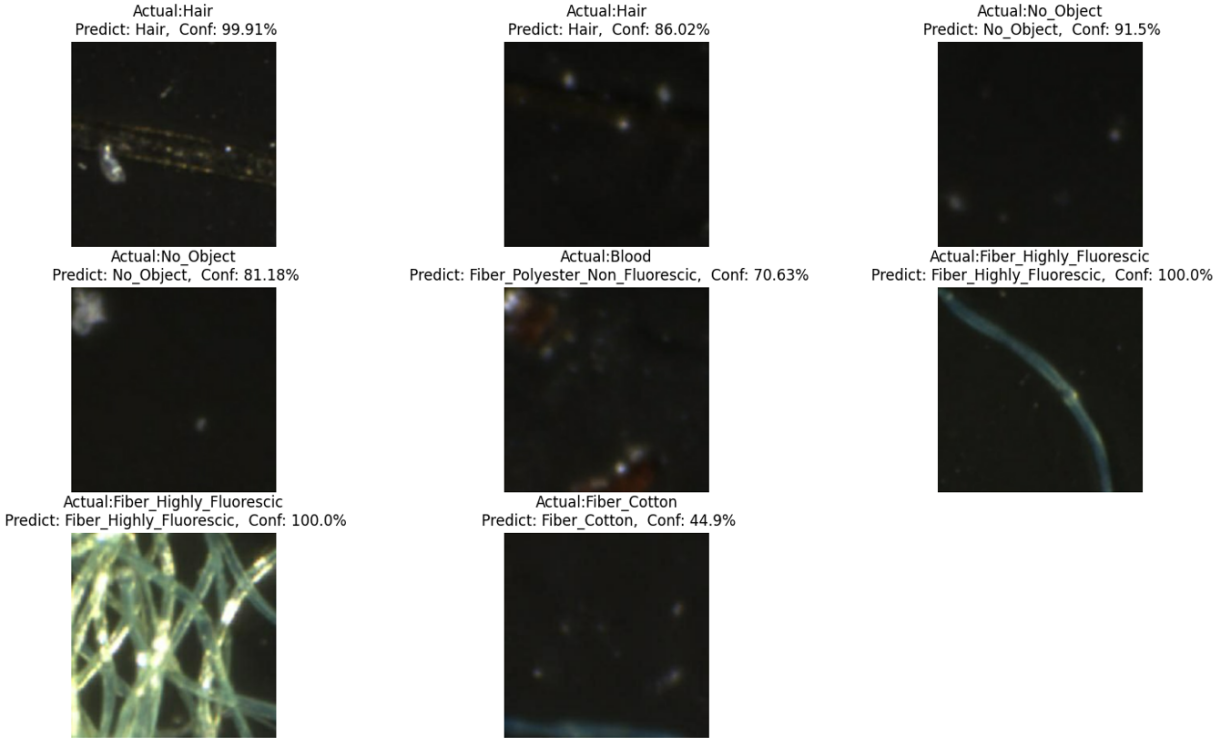


Figure 4.11: *8 Predicted images under reflectance microscopy.*

Even though the accuracy on some easily distinguished objects is high, with **reflectance** image dataset we get an **average accuracy of 86%** and some decrease spikes

on our metrics, due to the fact that many objects do not reflect any light or they do not fluoresce if they are being exposed to fluorescence light ,or a specimen may do not have birefrigence properties.In these cases the images are completely black or extremely fainted, and the network get confused because the No Object class contain only plane background, hence black images and many predictions were falsely classified as ”No object”.

As we can see fiber types are easily distinguishable on reflectance display, some human hairs do not reflect the light transmitted to them and many other features to help us categorize the objects us easy us possible.

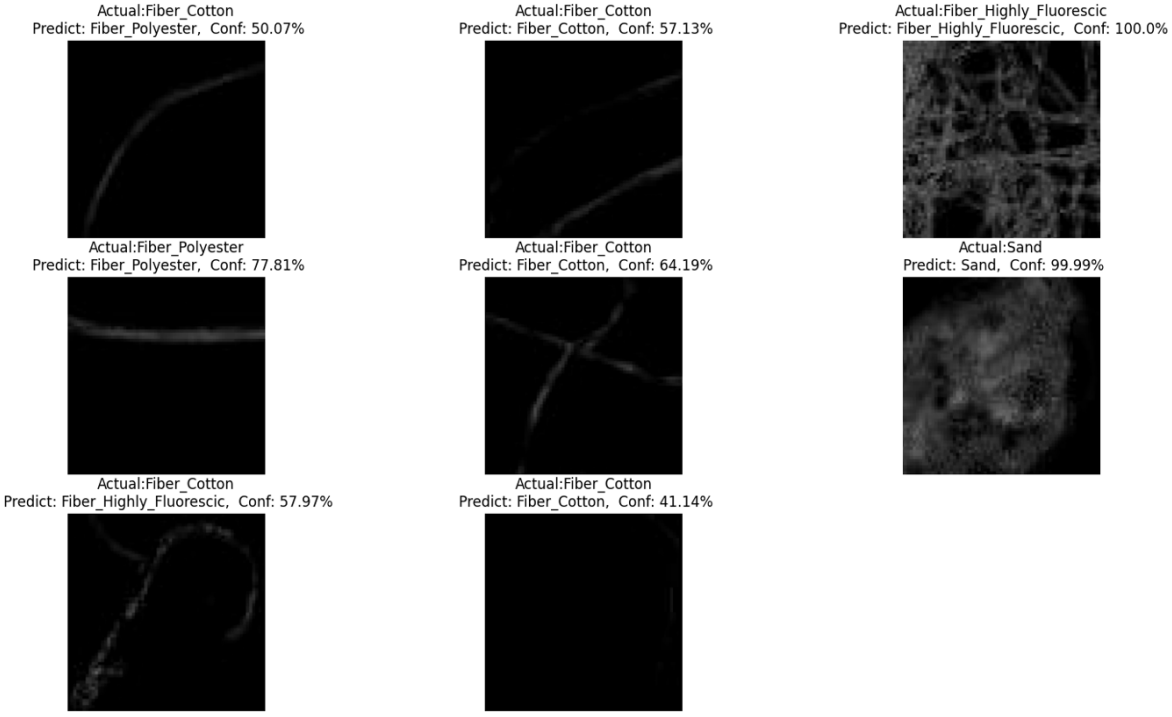


Figure 4.12: *8 Predicted images under* ***polarized microscopy.***

In **polarized** microscopy we get an **average accuracy of 83%**, it provides as monochromatic images but the birefrigence properties can really differ between objects.For example sand and glass can be hardly distinguished in some cases, but with polar images we can notice the differences based on their crystalline structures.Those unique differences

we find on specimens with those techniques are valuable information and with the right processes we can exploit them.

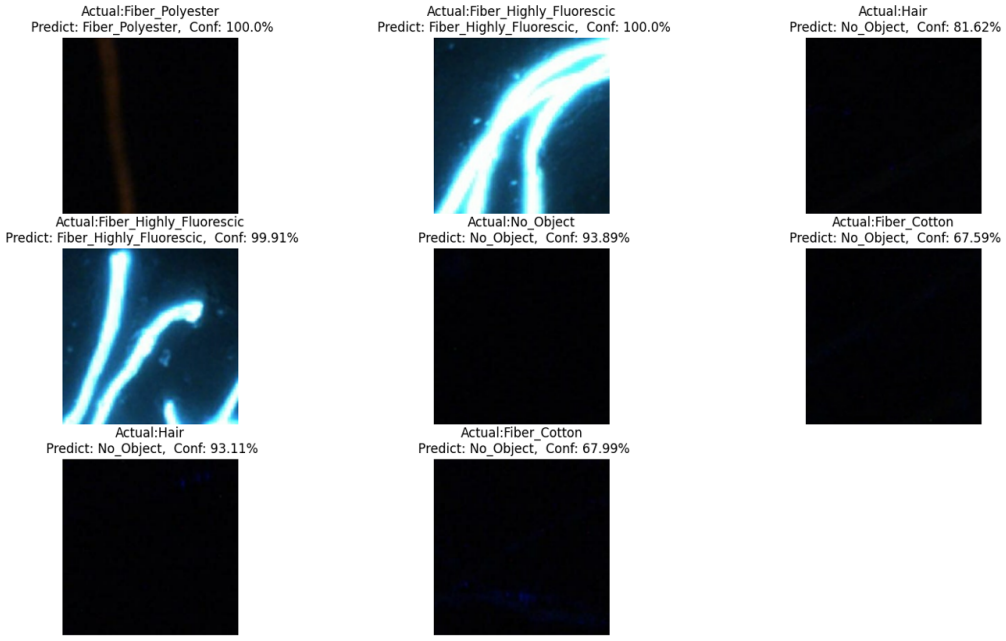


Figure 4.13: *8 Predicted images under fluorescence microscopy with a light source at 405nm.*

With **Fluorescence** imaging we get an **average** accuracy of **78%** because many specimens do not fluoresce at all, like some kind of hairs or glass.But some other specimens fluoresce enough to be easily distinguished by all the other objects like polyester and highly fluorescent fibers.

4.3 Multi Modal CNN Design

After the training process we save the model that can classify any given image based on the saved weights of our model. But it will classify correctly images only similar to those images that we fed and labeled to our network.We saw the **produced** data of a single **CNN** in a previous chapter , and they look like this:

$$F_{probabilities} = \begin{pmatrix} 0 & 0.1 & 0 & 0 & 0.77 & 0.03 & 0 & 0.1 & 0 & 0 \end{pmatrix}$$

The matrix F has n elements where n=*number of classes* we trained our network with, and the value of a element is the **prediction our model gave** us and the names of the classes are decoded alphabetically.

So we have 20 matrices each one consisted of 11 elements, in total 220 probabilities and if we want those 20 matrices to respond at the same image we just have to make sure that images are distributed to the right CNN.

4.3.1 Utilizing the data

We created a **new dataset from 2-D arrays** with size 1x221, first **220** values are the **probabilities** and the **last value** is the **actual label** of the specimen, in this case label is 1 which corresponds to "Blood".The first 11 elements are the predictions for one image by the CNN trained with the RGB images, the next 11 elements are the predictions of the next model and it goes on until we reach the last CNN.

0.99	0	0	0	0	0	0.01	0	0	0	0	0.97	.....	0	0	0.96	1

⟵ 11 ⟶

Matrix Size 1x221

Practically we acquire a matrix with probabilities but we can also see this matrix as a **code for a label**. Many of those probabilities due to the great accuracy of our individual CNNs have, are very close to 1, making the 10 rest probabilities **zero** or almost zero.That creates an evenly distributed array of elements and easy to process.Additionally, e.g. two objects may look very similar under Brightfield microscopy like fibers and hairs, but the fluorescence imaging can easily distinguish those two objects.So, if the prediction of the fluorescence model is almost 1 for fiber then for hair will be 0 and vise versa.This way we are creating a **unique encoding** for our class classification which will take advantage of the **unique properties** different objects have under different microscopy techniques.

4.3.2 Fully Connect Neural Network as final component

The fact that we create a new unique encoding ,which includes 220 values for each image on the first dataset and 160 values at the second allow us to combine our data with a Fully Connected NN due to the way a network like that functions.

A **Fully Connected NN** as we mentioned in chapter 2, consists of a series of fully connected layers that connect every neuron in one layer to every neuron in the other layer.Which means that as we go ”deeper” in the FC layers and the neurons we use decrease, only the high values will remain.Every **neuron** has its own weight that is added to its input and passed through a **RelU** activation function.As last layer we have **softmax activation** function which will give us the final probabilities.Softmax function compresses values between 0 and 1, the Softmax neurons allow the prediction of outputs to certain classes.

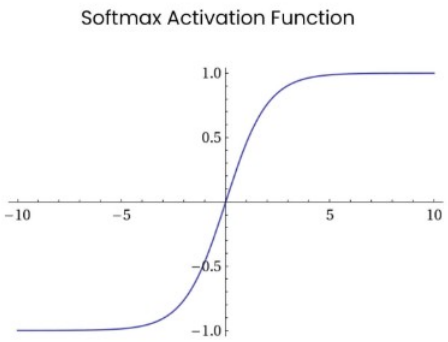


Figure 4.14: *Softmax Activation Function.Typically placed in output layers of networks used for classification.*

$$softmax(y_i) = e^{y_i} / \sum_{n=j} y_j$$

Where y_i is the **neuron's output**.

The major advantage of fully connected networks is that they are **“structure agnostic”** i.e. there are no special assumptions needed to be made about the input.And we are going to try and use it as a **decoder** for our 220 element arrays.

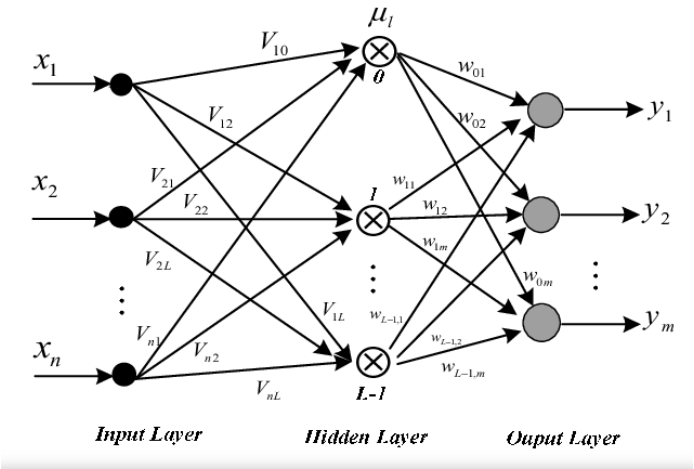


Figure 4.15: *A generic architecture of a FC NN.*

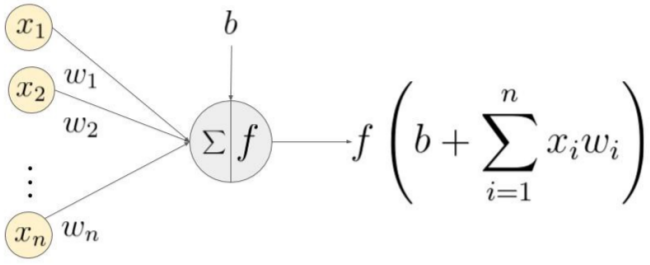


Figure 4.16: *Neuron's process.All the previous neurons pass their output to each neuron in the next layer, which adds its own bias(if exists) and weight and then the sum goes through the Softmax activation function f.*

Every neuron's mathematical equation is the same with different weights, inputs and biases:

$$y_i = f(x_1 * w_1 + x_2 * w_2 ... x_n * w_n + b)$$

,where y_i is the output of the i-th neuron,w are the weights and f is the RelU activation function.

4.3.3 Architecture of the Fully Connected NN

The architecture for this NN is simpler than CNN's architecture, and it consists of the input layer, 2 hidden FC layers and for multi-classification we use the Softmax activa-

tion with neurons equal to the number of our classes.**First layer** input shape has the shape of our array(1x221 and 1x161), **first hidden layer** has 1024 neurons and **second hidden layer** has 512 neurons. Also we apply a **Dropout layer** between Dense layers which shut down 30% of each layer's output, for the purpose of increasing the generalization of the model and prevent overfitting.This NN in term of python coding is shown below:

```
model = Sequential([
    Dense(1024, activation='relu', input_shape=input_shape),
    Dropout(0.3),
    Dense(512, activation='relu'),
    Dropout(0.3),
    Dense(number_of_classes, activation='softmax'),
])
```

Figure 4.17: *Python Code for our FC NN.*

Now that we have all the components we need for multi-classification and a solution to combine and exploit the psychical properties of objects through the variety of microscope's utility techniques our final design from start to end is shown below:

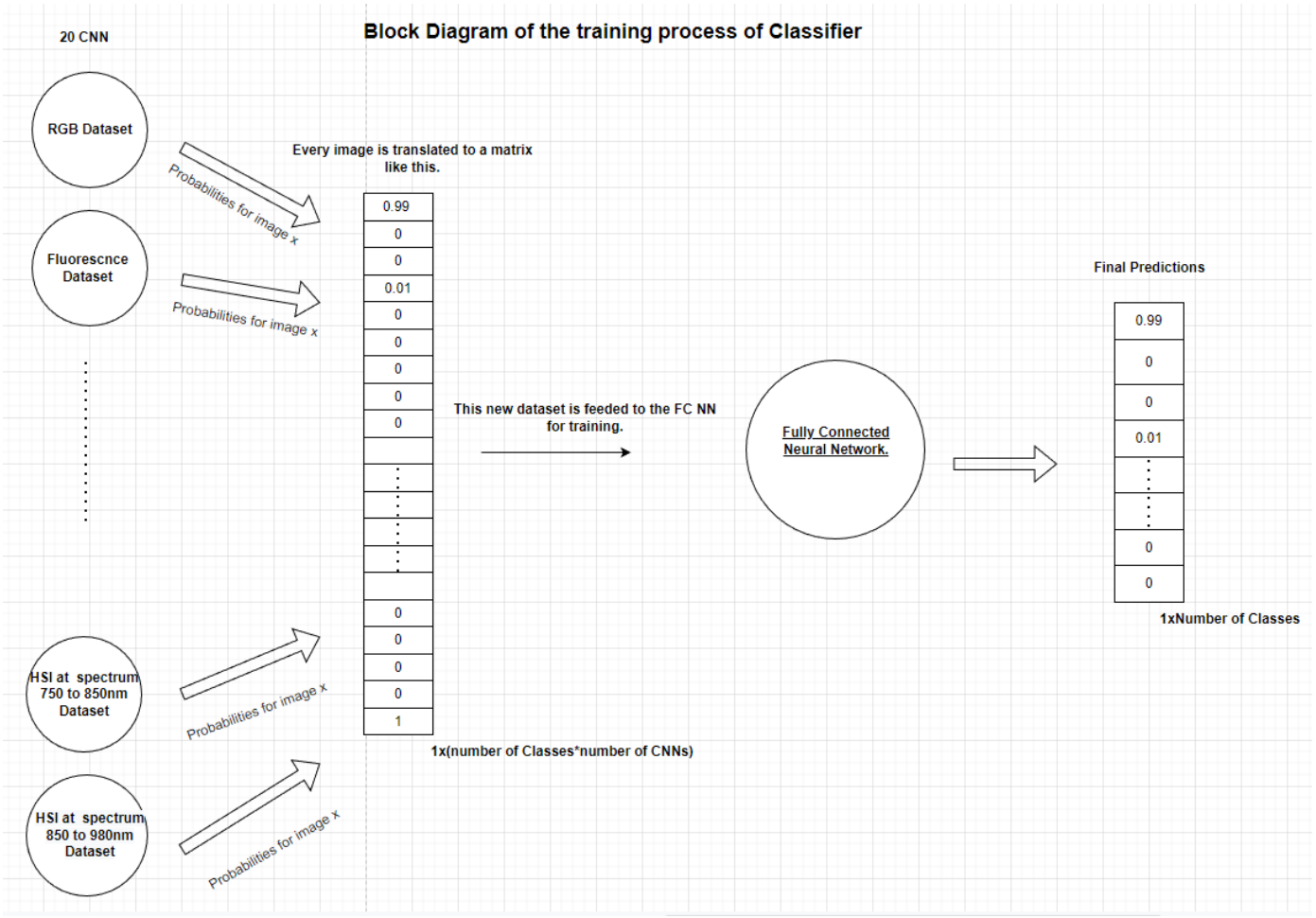


Figure 4.18: *Process behind training our FC NN and creating the Classifier Model.*

The same process will take place when we need to predict a new image, that means that every time we want to classify an image, we pass the same 256x256 image in 20 different modes, respectfully to the NN each image belongs, we take the outputs of all CNNs and create a matrix that fits the saved Classifier model and call the model to give us its prediction.

Chapter 5

Classification Results

5.1 Data Visualization

We applied classification into areas captured by the microscope, which we concatenate to produce the final overview image that consists of many smaller images with 640 pixels for height and for width 477 pixels.

When we find an object through the process we discussed for extracting the dataset, we crop the area the object is, to smaller parts that fit our **Neural Network's defined input size** which is 256x256 pixels.The pixels the areas that we are interested at more likely will not be a multiple of 256 so we will take extra pixels to fit that specific size if our ROI is smaller, or crop the image into smaller parts and individually predict the image it contains.

On our first try we draw the contours around the object as **rectangles** as we saw on figure 5.1 that contain the whole object and the prediction can be shown inside the rectangle.

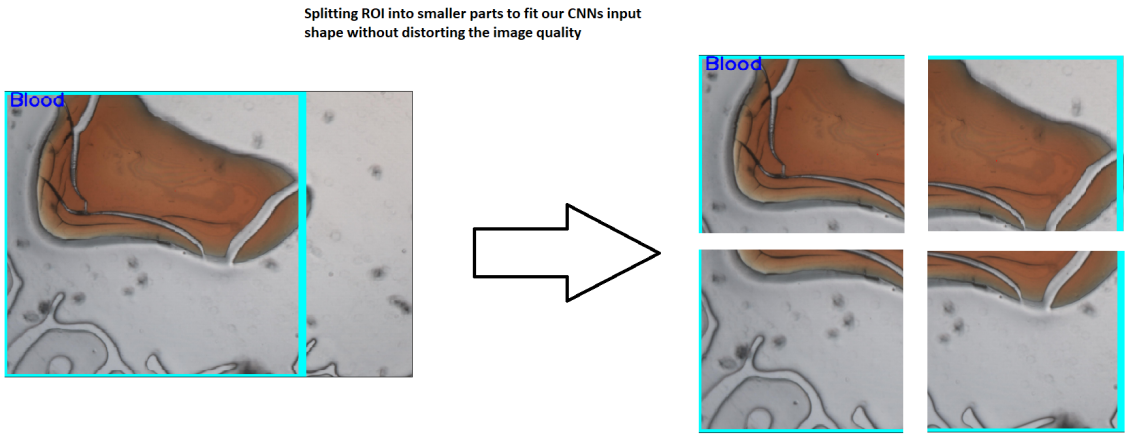


Figure 5.1: *Splitting process into right size with no distortion of the pixels.*

5.2 Proof of Concept

Our initial results were produced based on our first dataset and the **5 modes** that were taken account for are shown below:

- **Transmission** mode
- **Reflectance** mode
- **Fluorescence** mode with emmision light wavelength at 405nm
- **Fluorescence** mode with emmision light wavelength at 450nm
- **Polarized** mode

Fully Connected NN gave us 98.4% validation accuracy which is a very promising percentage and approximately **5%** higher than CNN trained with the Colored Dataset which had the best validation accuracy so far, but we have to test the power of our classifier also in unseen images and display the results.Some parts of the reconstructed image are shown below:

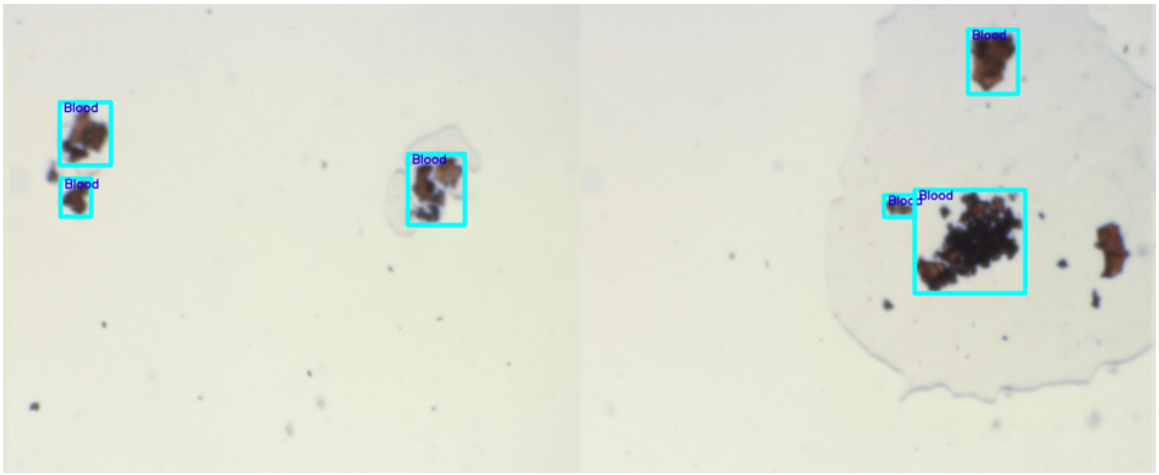


Figure 5.2: *Blood Area*

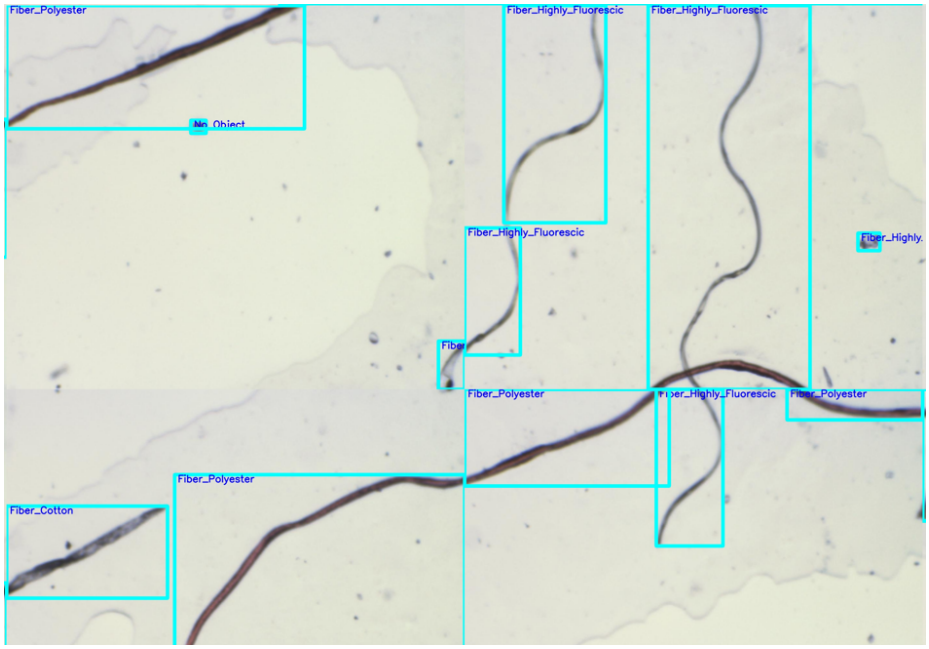


Figure 5.3: *Fiber Polyester and Highly Fluorescic Area*

As we can see on figures 5.2,5.3 and 5.4 all specimens are classified correctly except the smaller objects, which the networks were not trained with images containing very small objects. But these 3 figures are only a part of the final **overview image** which consists of **1024 tiles** of smaller 640x480 pixels images and is really huge to be shown.So by counting the successfully classified contours that contain an object, we measured the **success rate** of the classifier to be (447 contours containing object) / (431 contours correctly classified) **=96.42%**.

By verifying the success of our system we are ready to go deeper on further research to understand how each different display mode affect our accuracy we are going to **gradually add modes and check how success rate behave** on those changes and conclude on how each mode affect the results.

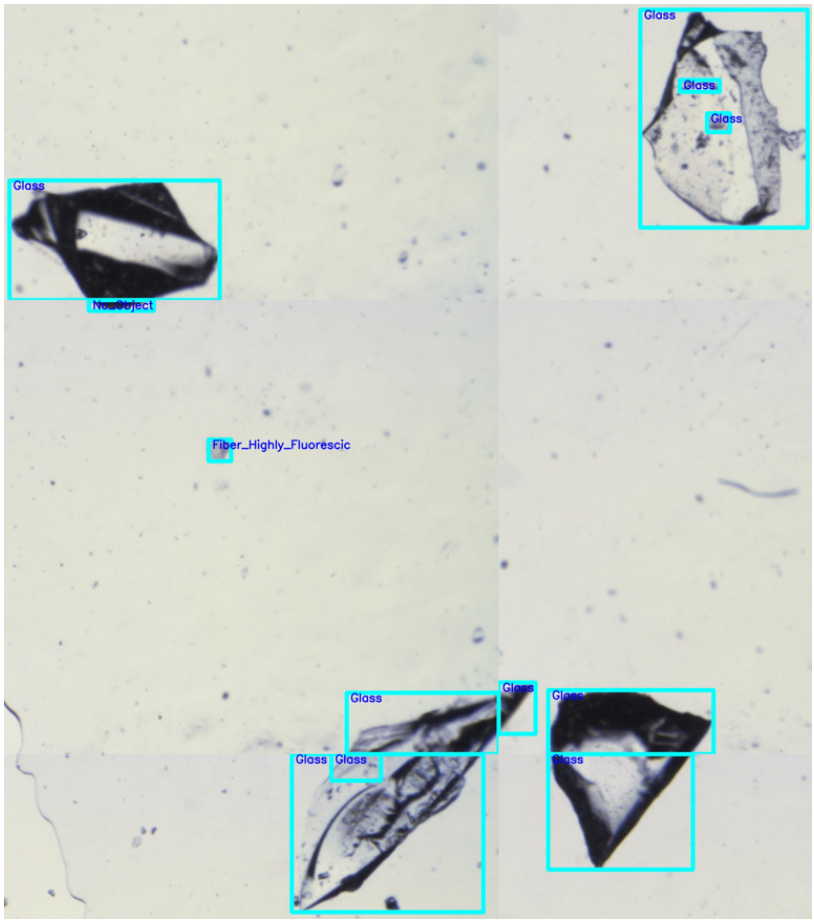
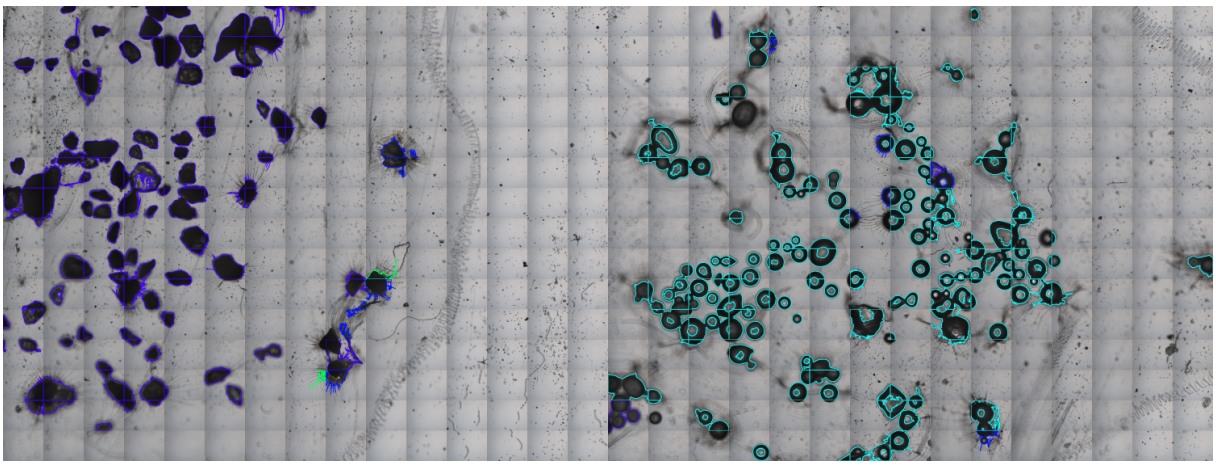


Figure 5.4: *Glass Area*

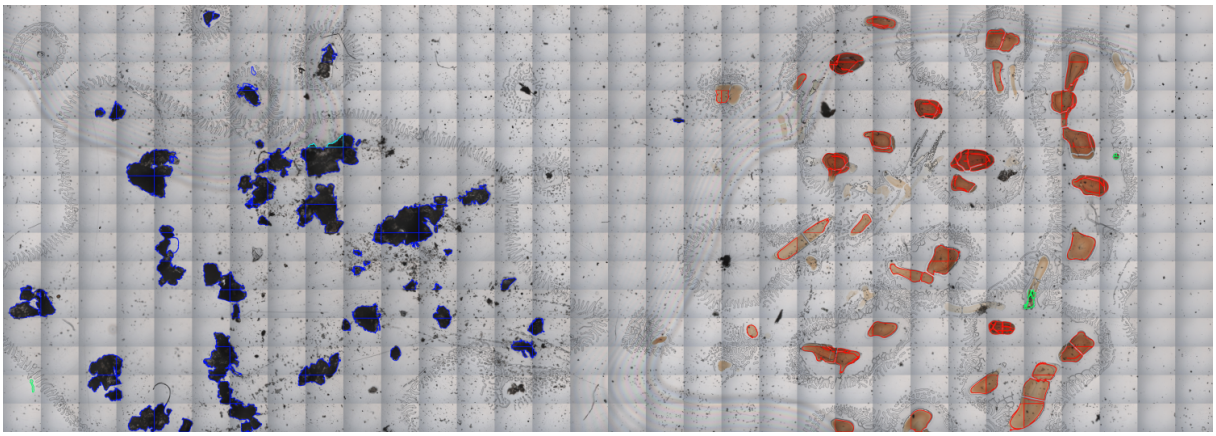
5.3 Validation

The next step was to create a series of combinations with the new available display modes, HSI at 13 different bands,another Fluorescence mode that captures the emmited light from the specimen at 365nm and another Reflectance mode with the light source transmitting at Ultraviolet spectrum. A **common classification mode** due to its ability to display objects in unique appearances, thus easily distinguable, is **Reflectance mode** with light source wavelength at the visible electromagnetic spectrum(380-700nm), so we began our research by using only the Reflectance mode.The results are shown in Figure Figure 5.5 and at **Success Rate table** :

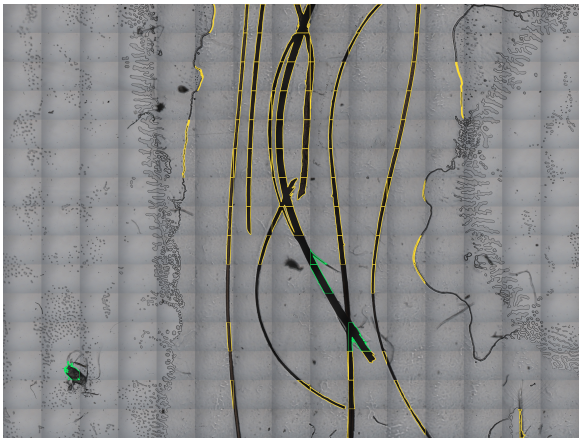
Our way of detecting the objects by dividing the overview image into smaller tiles



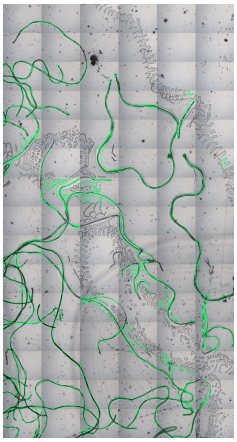
(a) *Glass and Sand Area*



(b) *Skin and Blood Area*



(c) *Hair Area*



(d) *Fiber Area*

Figure 5.5: *Classified Areas using Reflectance mode with the light source on the visible electromagnetic spectrum.*

can lead to finding an object into more than one tile, so we do not count the objects but the contours that exist in our classifying areas.So in our final areas we find:

- **Blood**: 81 contours
- **Glass**: 109 contours
- **Skin**: 104 contours
- **Hair**: 94 contours
- **Sand**: 124 contours
- **Fiber**: 81 contours

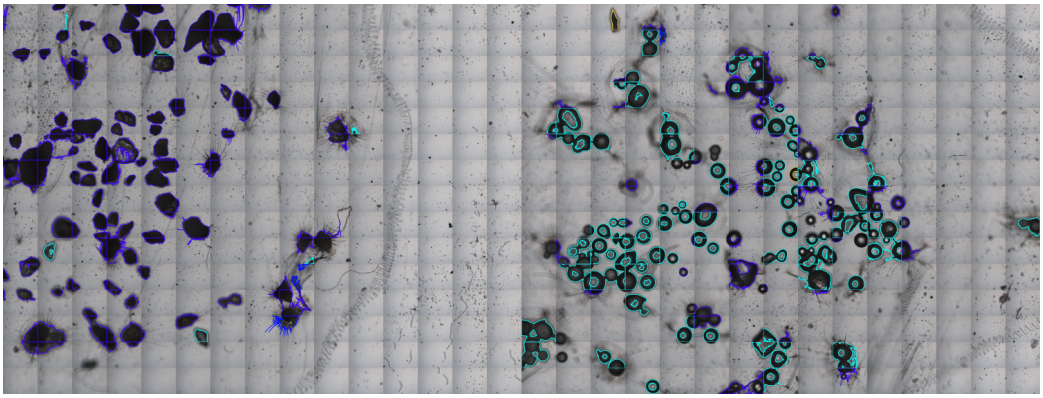
Success Rate is calculated like this: (numbers of successfully classified contours / number of contours)*100.Below is the success rate table for the classified contours only under the **Reflectance Visible** mode.

Mode	Blood	Glass	Skin	Hair	Fiber	Sand
Reflectance Visible	98%	93.57%	98.07%	89.36%	100%	95.16%

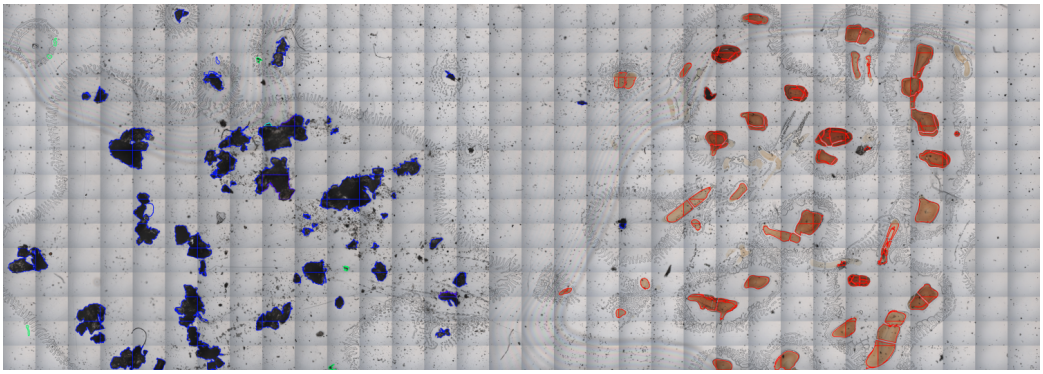
The **average time for each prediction** is 35ms and time is independent from which mode we use and its dependent only on how much information is on the image being classified, and with **2357** predictions in total, our overall time for loading the images and predicting them is **230.70 seconds** or 3.845 minutes. Even though our contours are 593 our total predictions are almost 5 times that number because one contour may need many predictions and also some invalid contours that the prediction about what object the image contains is less than 90% are left out, and also any contours with bubbles from the tape are also left undrawned.

Our next step is to apply classification with only the **Transmission Color** mode which gives us around **93%** Train and Validation accuracy and see the results:

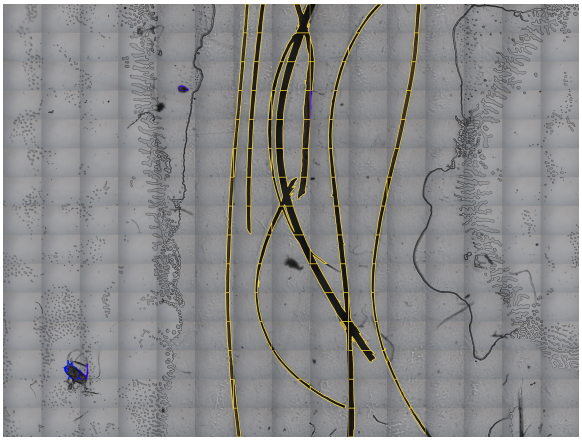
Mode	Blood	Glass	Skin	Hair	Fiber	Sand
Transmission Color	98.76%	68.8%	99%	94.6%	98.7%	95.1%



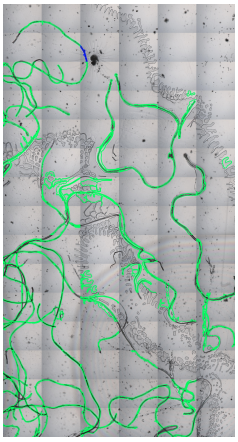
(a) *Glass and Sand Area*



(b) *Skin and Blood Area*



(c) *Hair Area*



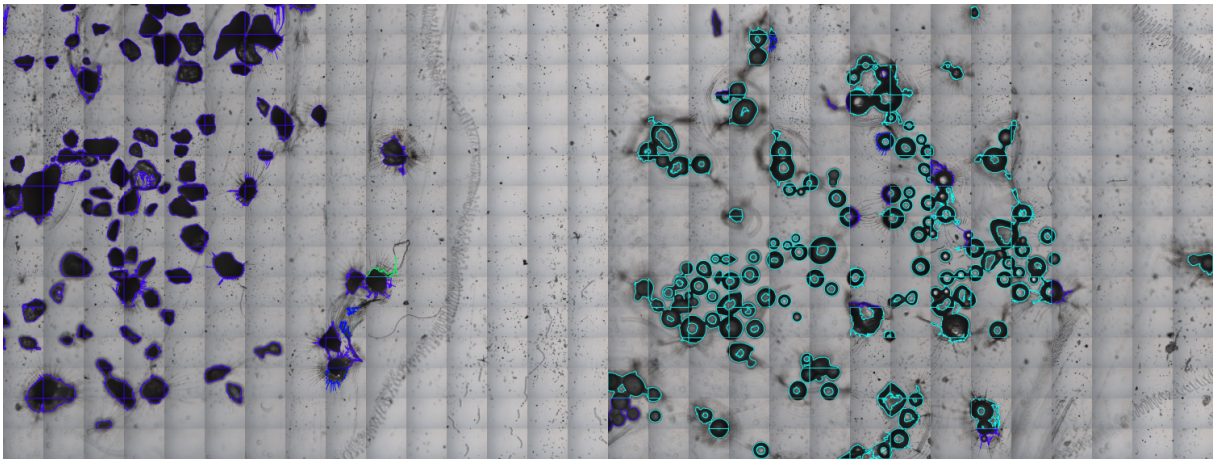
(d) *Fiber Area*

Figure 5.6: *Classified Areas using Transmission Color mode.*

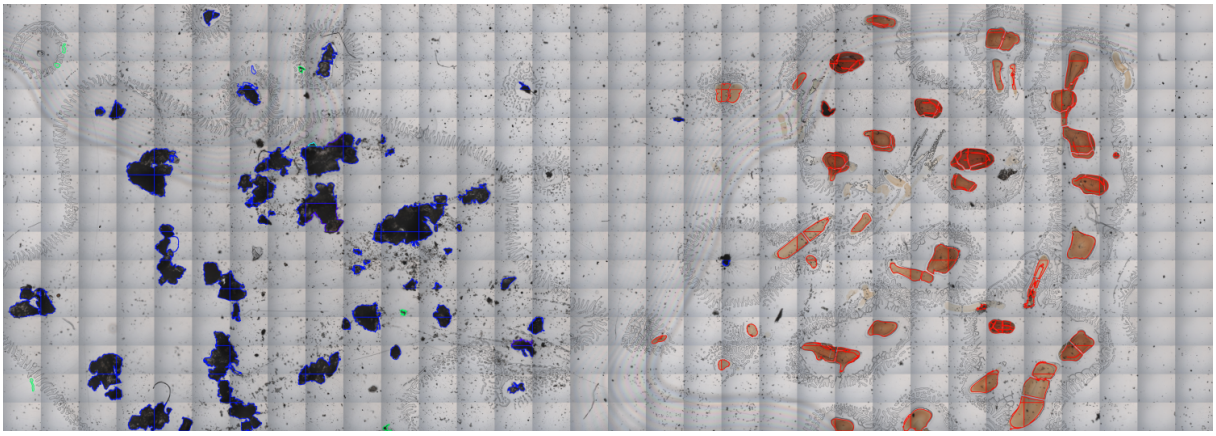
We continue our research by gradually adding mode by mode so we can deduce the value of every mode in classifying specimens, and which mode help us distinguish each specimen.So below are the results after classifying the same areas by **adding** the **Transmission Color mode**:

Success Rate table for Transmission Color and Reflectance mode:

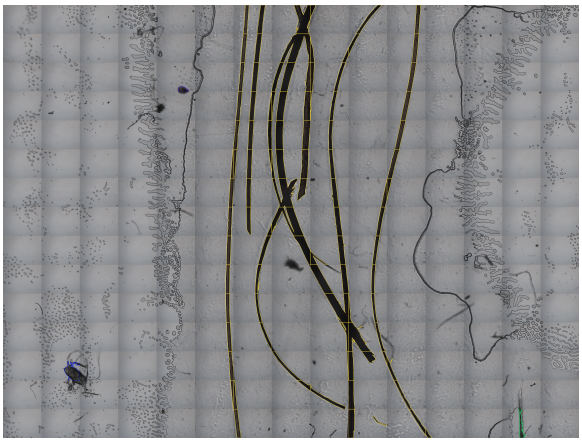
Mode	Blood	Glass	Skin	Hair	Fiber	Sand
Color and Reflectance	98.76%	91.74%	99%	100%	98.7%	97.58%



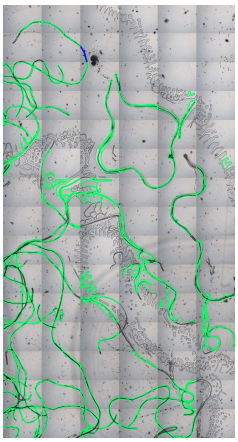
(a) *Glass and Sand Area*



(b) *Skin and Blood Area*



(c) *Hair Area*



(d) *Fiber Area*

Figure 5.7: *Classified Areas using two modes,Reflectance on visible spectrum,and Transmission Color mode*

By comparing the results so far we can infer that each mode contributes greatly in different specimens, for example, when we applied classification with only the Reflectance

mode we achieve great accuracy at distinguishing glass and sand that very much look alike, also flawless classification on the fiber area , and great results on fiber and blood. But when it comes to hair, reflectance properties are really low, so we get only 89.3%.

After we applied classification with both Reflectance and Transmission Color modes we achieve 100% accuracy, which is a **10.7% raise** of success rate.Also success rate **slightly increased for Skin, Sand, Blood** areas , but the success rate for Fiber and Glass areas decreased. That leaves us to conclude that if two objects lookalike in one mode , and only two modes are taking part in the classification that may lead to **confuse** our final NN if trained with false or uncertain information.

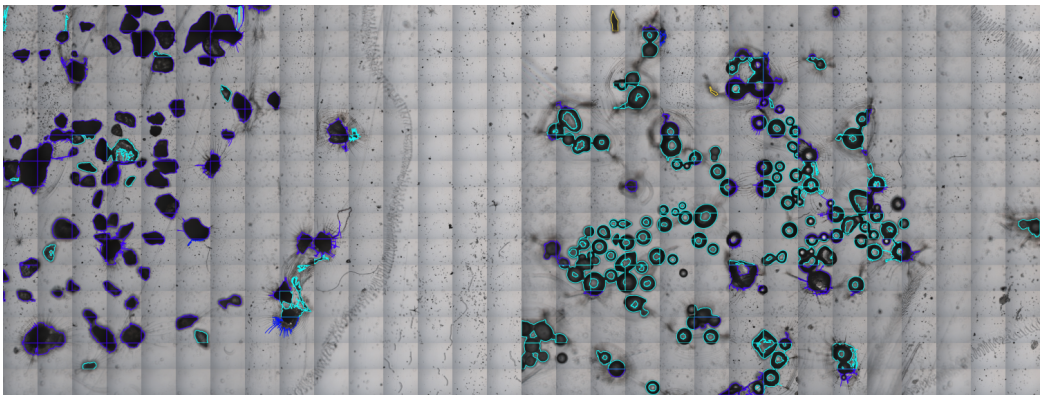
So we have to determine which modes are needed for best results, in terms of success rate as first criterion, secondly we have to understand which modes may lead to misguidance to our FC-NN and last we have to determine which modes do not assist our classification at all and exclude them to reduce the time and computational load of our system.

Another mode we tested was only the 13 Hyperspectral images to determine the power of just the HSI, and here are the results:

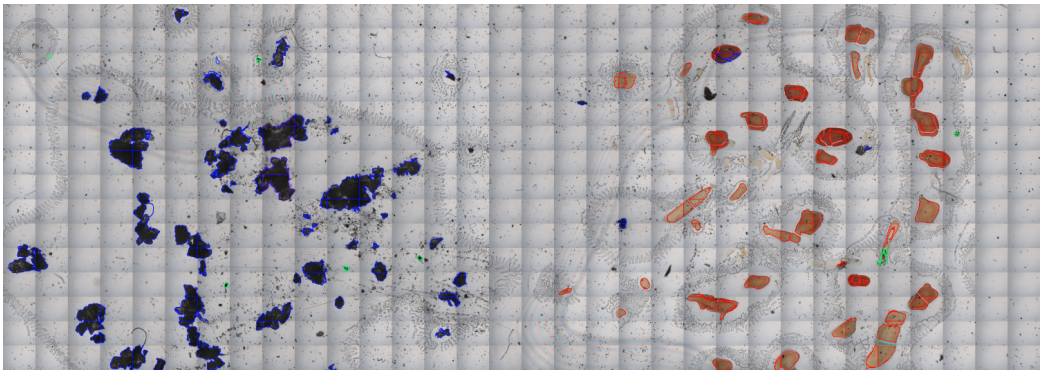
Success Rate table of classification using 13 HSI modes:

Mode	Blood	Glass	Skin	Hair	Fiber	Sand
HSI 13 Bands	97.53%	65.13%	99%	97.87%	100%	90.3%

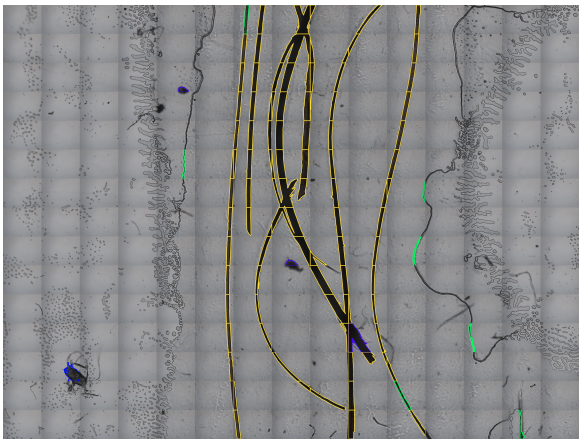
In figure Figure 5.8 , even though the hairs are classified correctly , some bubbles from the tape are misclassified for fibers with Confidence more than 90% because in HSI they look the same.Also the success rate for Glass dropped dramatically to **65.13%** and we had a **7%** decrease in Sand specimens. Even though HSI provides valuable details about specimen composition it is not enough for classification with CNNs.



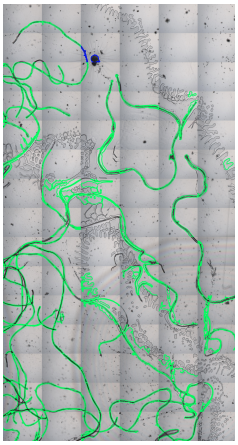
(a) *Glass and Sand Area*



(b) *Skin and Blood Area*



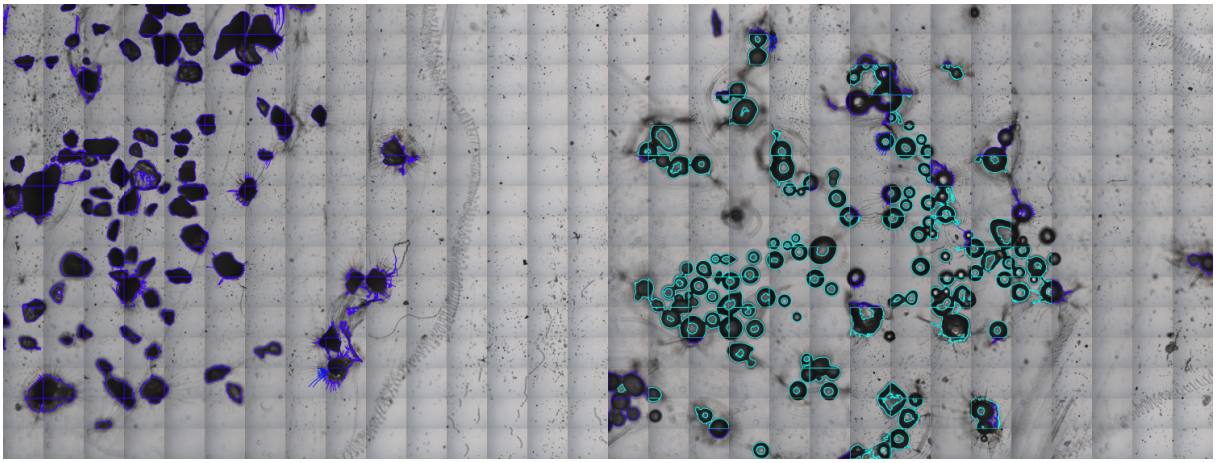
(c) *Hair Area*



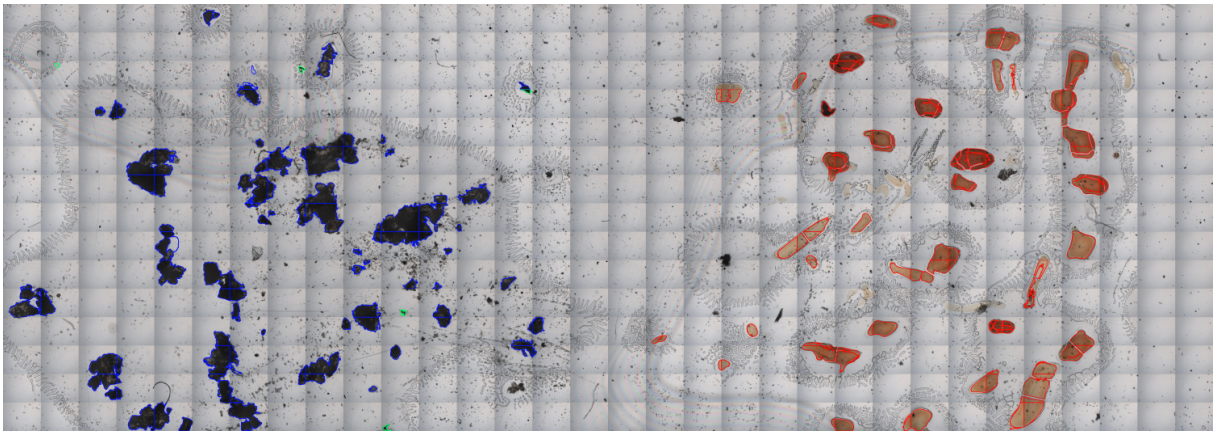
(d) *Fiber Area*

Figure 5.8: *Classified Areas using only the HSI at 13 different bands.*

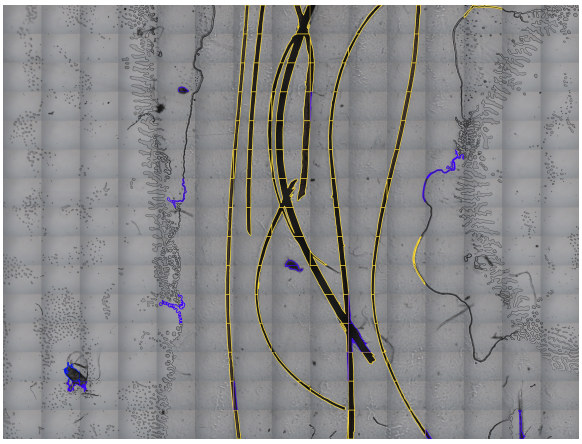
On our next combination of modes includes Transmission Color,Reflectance on the visible and the ultraviolet spectrum, Polarized mode, and one Fluorescence mode with light source at 450nm.Below on figure Figure 5.9 are the results:



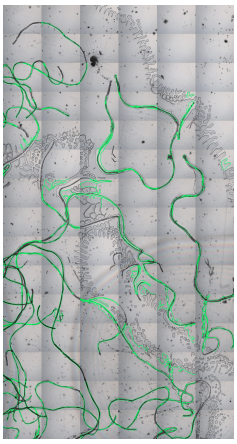
(a) *Glass and Sand Area*



(b) *Skin and Blood Area*



(c) *Hair Area*



(d) *Fiber Area*

Figure 5.9: *Classified Areas using Polarized mode, Fluorescence mode at 450nm, Transmission Color, Reflectance in visible and ultraviolet spectrum.*

Success Rate table for classification using Transmission Color, Polarized, both Reflectance modes and Fluorescence mode with light source at 450nm.

Blood	Glass	Skin	Hair	Fiber	Sand
98.8%	84.4%	100%	100%	100%	100%

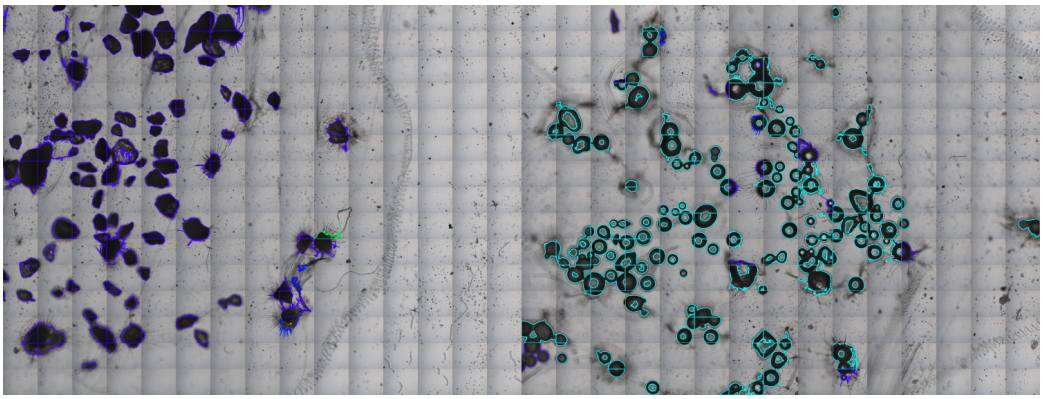
By comparing the combinations so far we see that in the last combination we have **100%** success rate for Sand , but success rate for Glass dropped by 9.2% from our best Success Rate so far when we combined Transmission Color and Reflectance.

Next combination of modes is the same as previous combination but we also added two Fluorescence modes with the light source at 365 and 405nm.

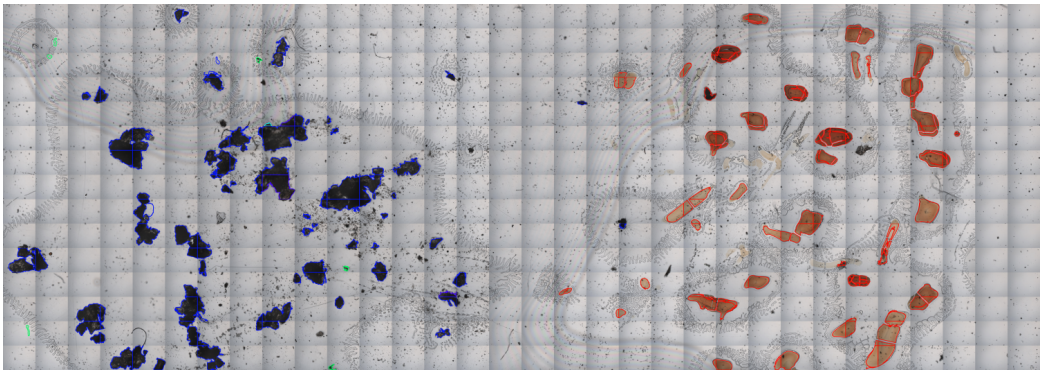
Success Rate table of classification by using every mode except HSI.

Blood	Glass	Skin	Hair	Fiber	Sand
98.8%	89.9%	100%	100%	100%	98.3%

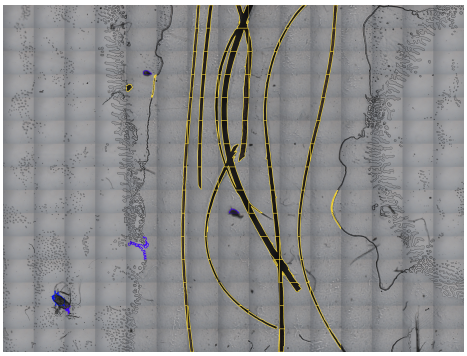
By adding 2 more Fluorescence mode capturing the emmited light at lower wavelengths we see **7.4%** increase in the success rate for the Glass Area but also one falsely classified Sand contour.



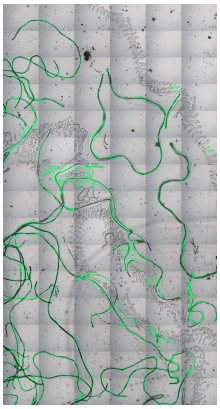
(a) *Glass and Sand Area*



(b) *Skin and Blood Area*



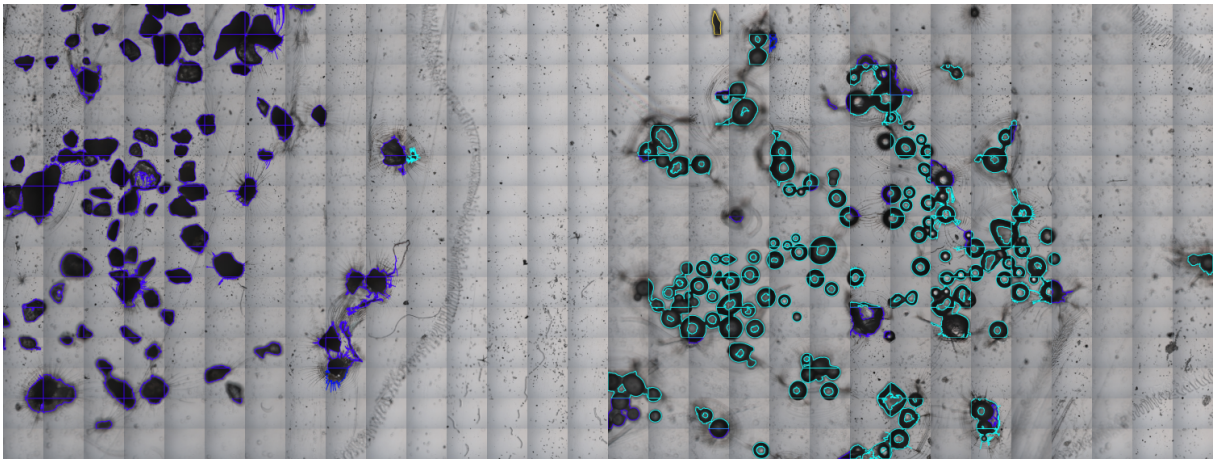
(c) *Hair Area*



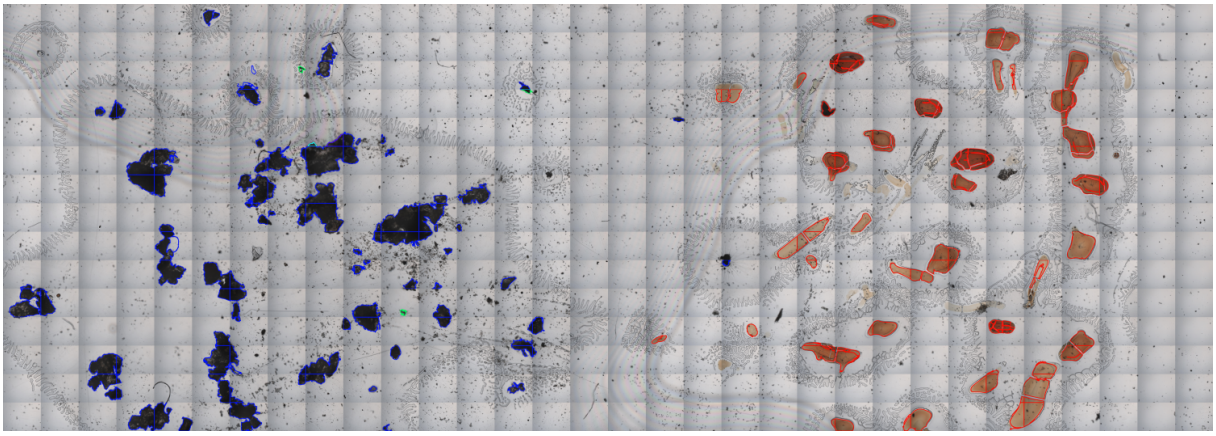
(d) *Fiber Area*

Figure 5.10: *Classified Areas using Polarized mode, Fluorescence mode at 365,405 and 450nm, Transmission Color, Reflectance in visible and ultraviolet spectrum.*

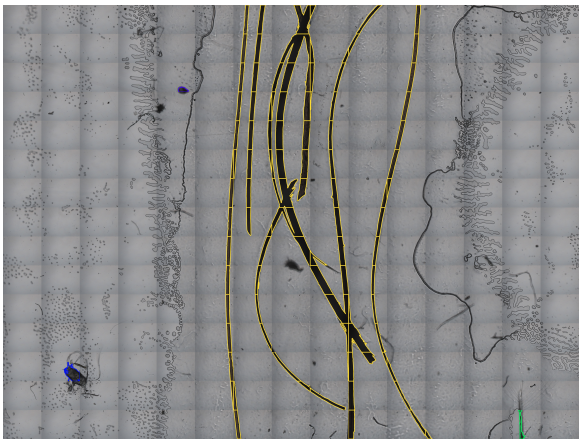
Afterwards we **added 5 HSI modes** covering the electromagnetic spectrum from **325nm to 450nm**.On Figure Figure 5.11 the results of the classification and the success rate table:



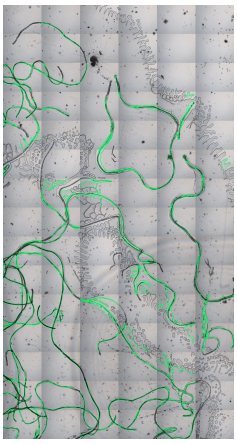
(a) *Glass and Sand Area*



(b) *Skin and Blood Area*



(c) *Hair Area*



(d) *Fiber Area*

Figure 5.11: *Classified Areas using Polarized mode, Fluorescence mode at 365,405 and 450nm, Transmission Color, Reflectance in visible and ultraviolet spectrum and our 5 first Hyperspectral modes.*

Success Rate table for classification with every mode except the last 8 Hyperspectral modes:

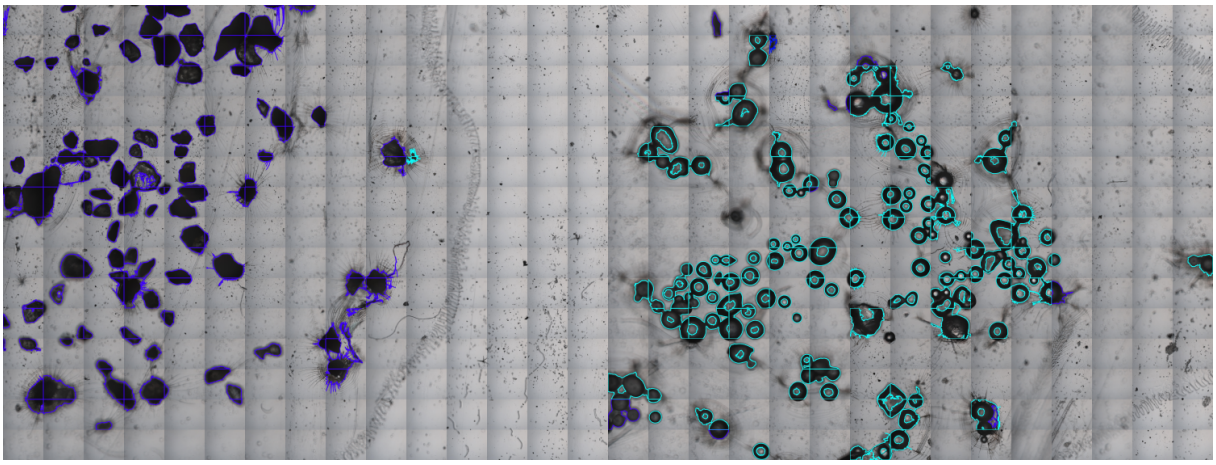
Blood	Glass	Skin	Hair	Fiber	Sand
98.8%	88%	100%	100%	100%	100%

Even though we added 5 Hyperspectral modes, it affected our success rate only by little, increasing the Sand accuracy by 1.7% but decreasing the Glass accuracy by 1.9%

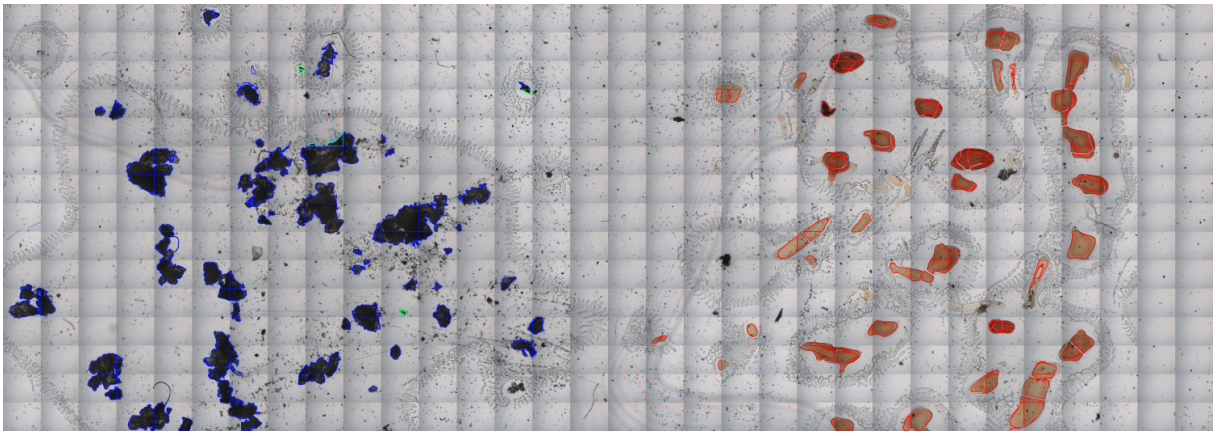
Our next step is to **add 5 more HSI modes** covering the electromagnetic spectrum from **325nm to 750nm**.Below are the results of the classification and the success rate table:

Success Rate table for classification with every mode except the last 3 Hyperspectral modes:

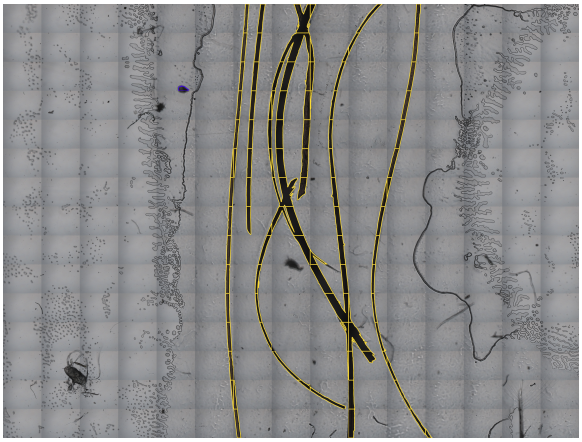
Blood	Glass	Skin	Hair	Fiber	Sand
98.8%	93.5%	100%	100%	100%	99.2%



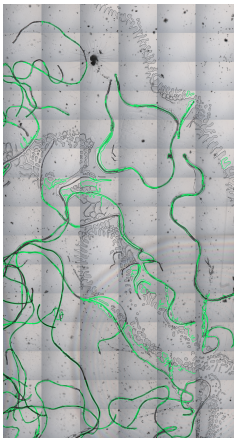
(a) *Glass and Sand Area*



(b) *Skin and Blood Area*



(c) *Hair Area*



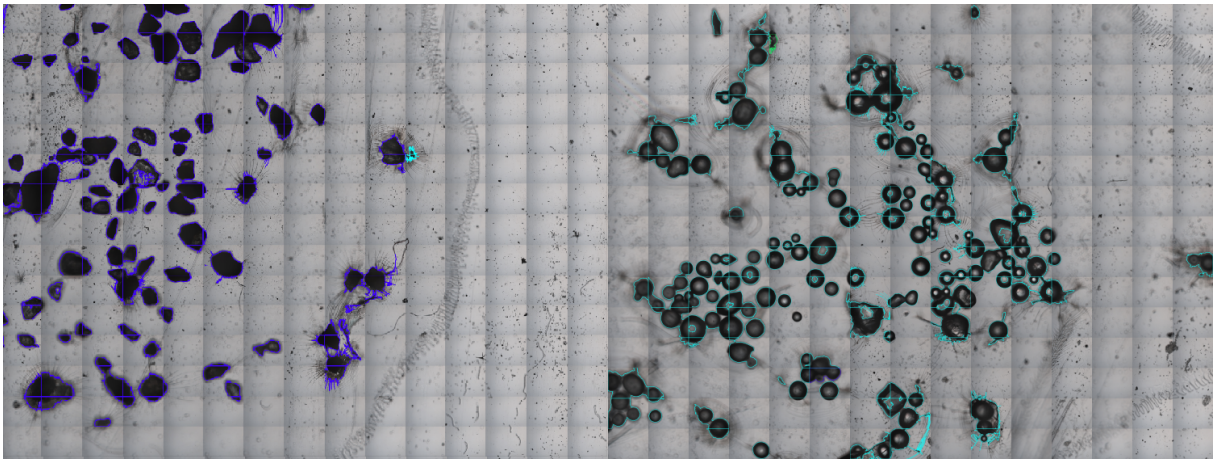
(d) *Fiber Area*

Figure 5.12: *Classified Areas using Polarized mode, Fluorescence mode at 365,405 and 450nm, Transmission Color, Reflectance in visible and ultraviolet spectrum and our 10 first Hyperspectral modes.*

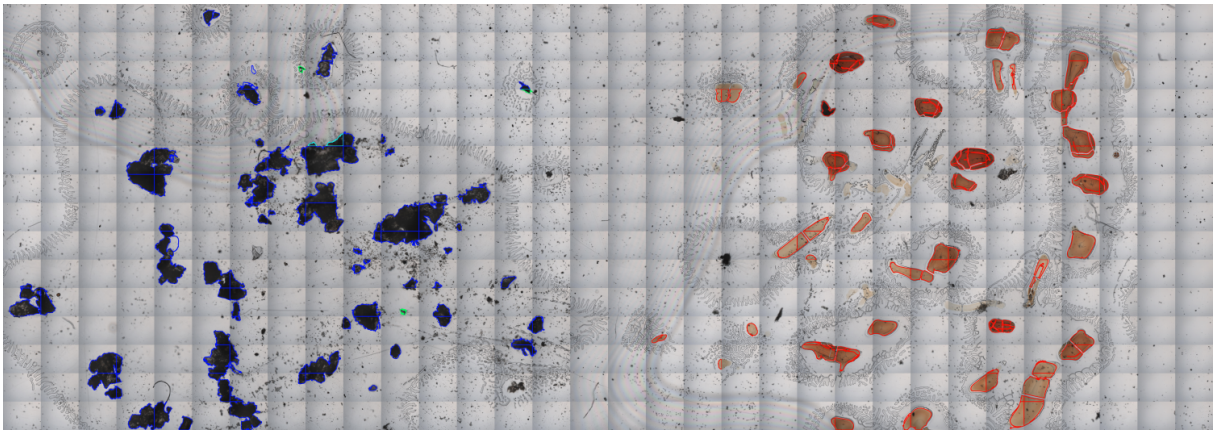
For our last combination we **added every HSI mode** covering the electromagnetic spectrum from **325nm to 950nm**.Below on Figure Figure 6.1are the results of the classification and the success rate table:

Success Rate table for classification with all 20 modes:

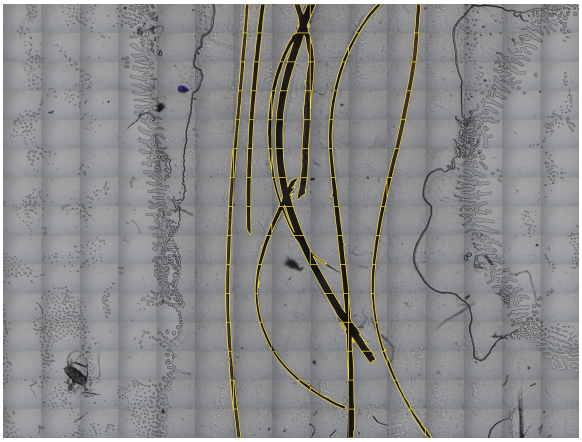
Blood	Glass	Skin	Hair	Fiber	Sand
98.8%	100%	100%	100%	100%	99.2%



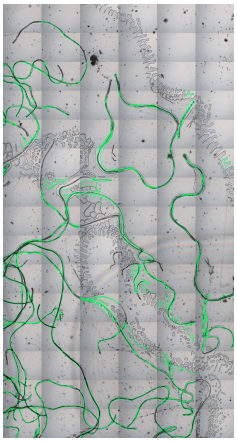
(a) *Glass and Sand Area*



(b) *Skin and Blood Area*



(c) *Hair Area*



(d) *Fiber Area*

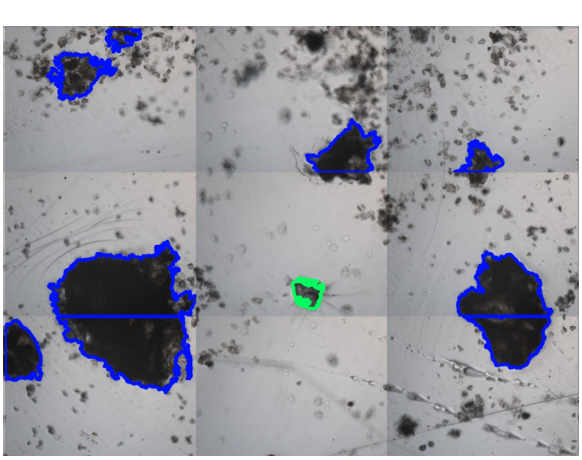
Figure 5.13: *Classified Areas using 20 modes*

Chapter 6

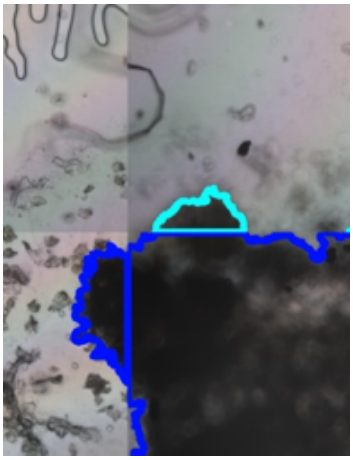
Conclusions and Future Work

We specific chose this kind of Glass specimen that it is not transparent to make it more difficult to distinguish from Sand, and as we can see those 2 areas are the only ones who confuse our CNNs.

Also **very small objects or parts of a bigger specimen** that have been disjoined by a adjacent tile may be **falsely classified** due to the fact that our **dataset do not contain very small objects** , and our CNNs do not have a reference to match them with.



(a) *Very small object that our CNNs were not trained to classify,*



(b) *Part of a bigger specimen that was disjoined by the tile below*

Figure 6.1: *Examples of very small objects*

Even though our dataset should be larger or we should find a different way for object detection, as we can see , when we use all of our modes we achieve a perfect classification and that demonstrates the power of CNNs.

While testing out different combinations of modes to determine which modes are

critical for the classification, we noticed that some modes can distinguish some objects better than other.

First combination was only **Reflectance Mode** with our light source on the visible spectrum which is known for its capabilities to display objects with unique appearances and even though we get great results with only one mode some specimen do not reflect much light and they can not be classified, like some parts of hairs so we continue to add different modes that can detect those objects.

When we combined **Transmission Color and Reflectance on the visible spectrum** we achieve great Success Rate, but for Glass Area we got **92%** and we want to increase that rate even more.

ur fourth combination was only the 13 bands of wavelengths that we captured with HSI. Even though Blood, Skin, Hair and Fiber which differ a lot as matter components, and that is what HSI tries to display, Glass that is processed Sand can not be distinguished by the CNNs , and even though we get **90.3%** Success Rate on the Sand area , we get only **65.13%** Success Rate on the Glass area.

Afterwards we added **Reflectance** with our light source at the **ultraviolet spectrum**, **Polarized** Mode and **Fluorescence** mode that captures the emmited light from the specimen at **450nm** and we achieve only **84.4%** at Glass area but 100% to every other area.Our CNNs predictions were worse on the Glass area but better on every other area.

Next we **added two Fluorescence** modes capturing light at different bands at 365nm and 405nm we see an increase of **5.5%** at the Glass Area that it is the only area we want to improve, as every other area is almost perfectly classified.

On table Table 6.1 are all the Success Rate tables so far:

	Modes	**Blood**	**Glass**	**Skin**	**Hair**	**Fiber**	**Sand**
1	Reflectance on the visible spectrum	98%	93.57%	98%	89.36%	100%	95.16%
2	Transmission Color	100%	68.8%	99%	94.6%	98.7%	95.1%
3	Transmission Color and Reflectance	100%	91.74%	98%	100%	98.7%	97.58%
4	Transmission Color + Reflectance + Polarization	100%	88%	98%	96.8%	100%	99%
5	13 HSI Modes from 325nm to 980nm	95%	65.13%	99%	97.87%	100%	90.3%
6	4 + Reflectance on the ultraviolet spectrum + Fluorescence capturing the emmited light at 450nm	100%	84.4%	100%	100%	100%	100%
7	6 + Fluorescence Modes capturing light at 365nm and 405nm	100%	89.9%	100%	100%	100%	98.3%
8	7 + 5 HSI Modes from 325nm to 450nm	100%	88%	100%	100%	100%	100%
9	8 + 5 HSI Modes from 450nm to 750nm	100%	93.5%	100%	100%	100%	99.2%
10	7 + 4 HSI Modes at 325,490,630,980	100%	93.5%	100%	100%	100%	99.2%
11	9 + 3 HSI Modes from 750nm to 980nm	100%	100%	100%	100%	100%	99.2%

Table 6.1: *Success Rate table for every combination.*

By adding only 5 HSI modes capturing the specimnes between 325-450nm we see a tiny decrease of **1%** on our Glass Area and a tiny increase of **1.7%** at the Sand Area that gave us 100% Success Rate.So we continue and we add the next **5 HSI modes** that capture

the specimen from 450nm-750nm and we see a **5.5% increase** on the Success Rate of the Glass area but we find a **single mistake** on the Sand Area. That mistake is caused probably due to the fact that the tape placed on the sand specimen is pushed too hard and the image is distorted and looks a lot like **transparent Glass**.Our dataset do not contained distorted specimens like that and so it is logical for our CNN to falsely predict it as glass. Because the results is ambiguous we place that predict as error and we continue with our combinations.

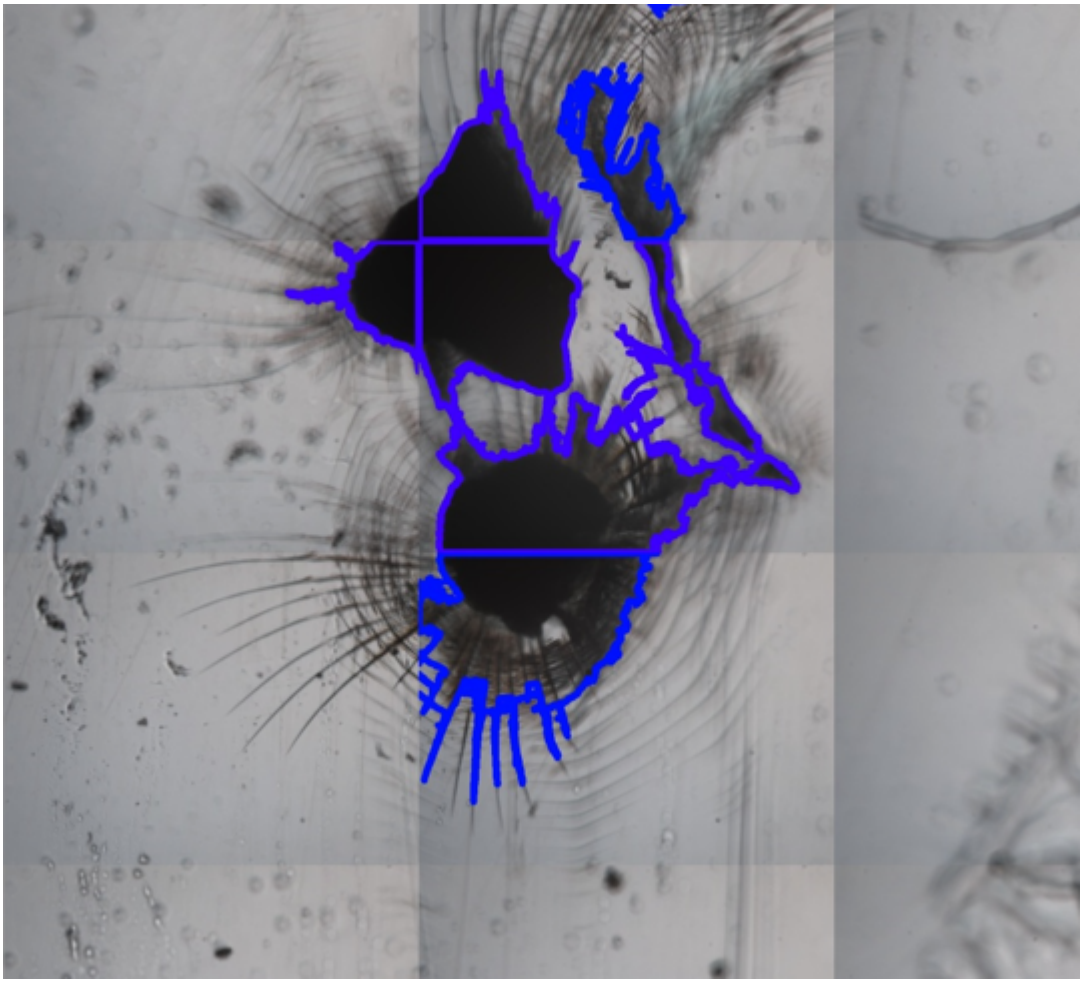


Figure 6.2: *Mistake in the Sand area caused by the tape.*

We continue by **adding 5 more HSI** Modes that capture the specimen from 450nm-750nm and the Success Rate **increased** by **5.5%** at Glass Area and no changes at the rest of the areas.

By including all 13 modes as we can see in Figure Figure 6.1 we increase our

Success Rate by **6.5%** and the only contour that is classified as Fiber is actually tape which has taken the color of the Glass, and our dataset does not contain any images as reference.

Every mode offers its own unique characteristics that help the final FC-NN identify the objects.As the modes increase so does the complexity of the encoding.Each mode adds as many elements as our different classes to the training array for the FC-NN, and this is an advantage for a powerful FC-NN like ours. When we combine 20 modes the final NN is trained with an array of 161 elements, the first 160 elements are the 20 different predictions from our CNNs and the 161th element is the object's actual label.

While classification with only one mode took 237s, the time is **not linear** because the time for prediction do not change, but loading the images is, if we load all images together will take less amount of time. So when combine 5 modes we need approximately 600 seconds , and when we use all 20 modes we need **2100 seconds**. If time was linear we would need at least **4000 seconds**.

6.1 Future Work

Improves can be made in a couple of sections in this thesis. Our first job is to further increase our dataset to include even the smaller objects and even specimens with diffused display from the tape. Increase the tape class so they are never misclassified us fibers or hairs,

Another potentially great change would be at the way we detect the objects not from tile to tile but from the whole overview image so there is no need to classify small objects but the whole specimen all together.

Also the process of detecting the objects with threshold according to the intensity of each pixel will change and replaced by **Convolutional Neural Networks that work like autoencoders and autodecoders** of an image.Unlike traditional methods of denoising, autoencoders do not search for noise, they extract the image from the noisy data that has been fed to them via learning a representation of it. The representation is then decompressed to

form a noise-free image[35].

Even more combinations will also determine if a mode can be completely extracted from our system, because it may decrease our Success Rate or do not provide anything, and by doing so we will decrease the overall time and maybe make our system more stable.

References

[1] V. Aginsky, "Forensic examination of "slightly soluble" ink pigments using thin-layer chromatography," *Journal of Forensic Science*, vol. 38, no. 5, pp. 1131–1133, 1993.

[2] R. H. Bremmer, S. C. Kanick, N. Laan, A. Amelink, T. G. van Leeuwen, and M. C. Aalders, "Non-contact spectroscopic determination of large blood volume fractions in turbid media," *Biomedical optics express*, vol. 2, no. 2, pp. 396–407, 2011.

[3] M. Govender, K. Chetty, and H. Bulcock, "A review of hyperspectral remote sensing and its application in vegetation and water resource studies," *Water Sa*, vol. 33, no. 2, pp. 145–151, 2007.

[4] A. A. Gowen, C. P. O'Donnell, P. J. Cullen, G. Downey, and J. M. Frias, "Hyperspectral imaging–an emerging process analytical tool for food quality and safety control," *Trends in food science & technology*, vol. 18, no. 12, pp. 590–598, 2007.

[5] G. P. Sabin, V. A. Lozano, W. F. Rocha, W. Romão, R. S. Ortiz, and R. J. Poppi, "Characterization of sildenafil citrate tablets of different sources by near infrared chemical imaging and chemometric tools," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 85, pp. 207–212, 2013.

[6] R. M. Levenson, E. S. Wachman, W. Niu, and D. L. Farkas, "Spectral imaging in biomedicine: A selective overview," *Imaging Spectrometry IV*, vol. 3438, pp. 300–312, 1998.

[7] G. Edelman, E. Gaston, T. van Leeuwen, P. Cullen, and M. Aalders, "Hyperspectral imaging for non-contact analysis of forensic traces," *Forensic Science International*, 2012. DOI: https://doi.org/10.1016/j.forsciint.2012.09.012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0379073812004495.

[8] D. L. Exline, C. Wallace, C. Roux, C. Lennard, M. P. Nelson, and P. J. Treado, "Forensic applications of chemical imaging: Latent fingerprint detection using visible absorption and luminescence," *Journal of forensic sciences*, vol. 48, no. 5, pp. 1047–1053, 2003.

[9] G. Payne, B. Reedy, C. Lennard, B. Comber, D. Exline, and C. Roux, "A further study to investigate the detection and enhancement of latent fingerprints using visible absorption and luminescence chemical imaging," *Forensic science international*, vol. 150, no. 1, pp. 33–51, 2005.

[10] N. J. Crane, E. G. Bartick, R. S. Perlman, and S. Huffman, "Infrared spectroscopic imaging for noninvasive detection of latent fingerprints," *Journal of forensic sciences*, vol. 52, no. 1, pp. 48–53, 2007.

[11] M. Tahtouh, J. R. Kalman, C. Roux, C. Lennard, and B. J. Reedy, "The detection and enhancement of latent fingermarks using infrared chemical imaging," *Journal of Forensic Science*, vol. 50, no. 1, JFS2004213–9, 2005.

[12] M. Tahtouh, P. Despland, R. Shimmon, J. R. Kalman, and B. J. Reedy, "The application of infrared chemical imaging to the detection and enhancement of latent fingerprints: Method optimization and further findings," *Journal of forensic sciences*, vol. 52, no. 5, pp. 1089–1096, 2007.

[13] R. Bhargava, R. Schwartz Perlman, D. C. Fernandez, I. W. Levin, and E. G. Bartick, "Non-invasive detection of superimposed latent fingerprints and inter-ridge trace evidence by infrared spectroscopic imaging," *Analytical and bioanalytical chemistry*, vol. 394, no. 8, pp. 2069–2075, 2009.

[14] K. S. Kalasinsky, J. Magluilo Jr, and T. Schaefer, "Hair analysis by infrared microscopy for drugs of abuse," *Forensic science international*, vol. 63, no. 1-3, pp. 253–260, 1993.

[15] K. S. Kalasinsky, "Drug distribution in human hair by infrared microscopy.," *Cellular and molecular biology (Noisy-le-Grand, France)*, vol. 44, no. 1, pp. 81–87, 1998.

[16] L. L. Randeberg, B. Skallerud, N. E. Langlois, O. A. Haugen, and L. O. Svaasand, "The optics of bruising," in *Optical-thermal response of laser-irradiated tissue*, Springer, 2010, pp. 825–858.

[17] G. Payne, N. Langlois, C. Lennard, and C. Roux, "Applying visible hyperspectral (chemical) imaging to estimate the age of bruises," *Medicine, science and the law*, vol. 47, no. 3, pp. 225–232, 2007.

[18] L. L. Randeberg, E. L. P. Larsen, and L. O. Svaasand, "Characterization of vascular structures and skin bruises using hyperspectral imaging, image analysis and diffusion theory," *Journal of biophotonics*, vol. 3, no. 1-2, pp. 53–65, 2010.

[19] B. Stam, M. J. van Gemert, T. G. van Leeuwen, A. H. Teeuw, A. C. van der Wal, and M. C. Aalders, "Can color inhomogeneity of bruises be used to establish their age?" *Journal of biophotonics*, vol. 4, no. 10, pp. 759–767, 2011.

[20] F. Li, S. Liu, R. Qi, H. Li, and T. Cui, "Effective visualization of latent fingerprints with red fluorescent la2(moo4)3:eu3+ microcrystals," *Journal of Alloys and Com-*

pounds, 2017. DOI: https://doi.org/10.1016/j.jallcom.2017.08.182. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925838817329110.

[21] A. M. Boddis and D. A. Russell, "Development of aged fingermarks using antibody-magnetic particle conjugates," *Anal. Methods*, vol. 4, no. 18, pp. 637–641, 2012.

[22] D. K. BIRD, K. M. AGG, N. W. BARNETT, and T. A. SMITH, "Time-resolved fluorescence microscopy of gunshot residue: An application to forensic science," *Journal of Microscopy*, 2007. DOI: https://doi.org/10.1111/j.1365-2818.2007.01752.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2818.2007.01752.x.

[23] S. R. Khandasammy, A. Rzhevskii, and I. K. Lednev, "A novel two-step method for the detection of organic gunshot residue for forensic purposes: Fast fluorescence imaging followed by raman microspectroscopic identification," *Analytical chemistry*, vol. 91, no. 18, pp. 11 731–11 737, 2019.

[24] M. D. Forrest, "Simulation of alcohol action upon a detailed purkinje neuron model and a simpler surrogate model that runs¿ 400 times faster," *BMC neuroscience*, vol. 16, no. 1, pp. 1–23, 2015.

[25] P. Yang, R. Ni, and Y. Zhao, "Recapture image forensics based on laplacian convolutional neural networks," in *International Workshop on Digital Watermarking*, Springer, 2016, pp. 119–128.

[26] O. Sporns, "Brain connectivity," *Scholarpedia*, vol. 2, no. 10, p. 4695, 2007, revision #91084. DOI: 10.4249/scholarpedia.4695.

[27] M. Ozkan, B. Dawant, and R. Maciunas, "Neural-network-based segmentation of multi-modal medical images: A comparative and prospective study," *IEEE Transactions on Medical Imaging*, vol. 12, no. 3, pp. 534–544, 1993. DOI: 10.1109/42.241881.

[28] G. Riegler, D. Ferstl, M. Rüther, and H. Bischof, "A deep primal-dual network for guided depth super-resolution," *arXiv preprint arXiv:1607.08569*, 2016.

[29] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE transactions on image processing*, vol. 19, no. 9, pp. 2241–2253, 2010.

[30] Y. Aksoy *et al.*, "A dataset of flash and ambient illumination pairs from the crowd," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 634–649.

[31] X. Deng and P. L. Dragotti, "Deep convolutional neural network for multi-modal image restoration and fusion," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 10, pp. 3333–3348, 2020.

[32] H.-Y. Choi, H.-U. Jang, J. Son, D. Kim, and H.-K. Lee, "Content recapture detection based on convolutional neural networks," in *International Conference on Information Science and Applications*, Springer, 2017, pp. 339–346.

[33] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *International conference on artificial neural networks*, Springer, 2010, pp. 92–101.

[34] A. Ismail, S. A. Ahmad, A. C. Soh, K. Hassan, and H. H. Harith, "Improving convolutional neural network (cnn) architecture (minivggnet) with batch normalization and learning rate decay factor for image classification," *International Journal of Integrated Engineering*, vol. 11, no. 4, 2019.

[35] M. Sundermeyer, Z.-C. Marton, M. Durner, and R. Triebel, "Augmented autoencoders: Implicit 3d orientation learning for 6d object detection," *International Journal of Computer Vision*, vol. 128, no. 3, pp. 714–729, 2020.