ULPGC Universidad de Las Palmas de Gran Canaria

eite escuela de ingeniería de telecomunicación y electrónica

ESCUELA DE INGENIERÍA DE TELECOMUNICACIÓN Y ELECTRÓNICA

BACHELOR THESIS

Obsidian classification using hyperspectral images to study the archaeological heritage of the Canary Islands

Degree:	Engineering in Telecommunications Technologies
Specialization:	Telematic
Author:	Alexandre Moreno Guillén
Supervisors:	Prof. Gustavo Marrero Callicó
	PhD. Himar Fabelo Gómez
	Mrs. Beatriz Martínez Vega
Date:	2022

INDEX

FIGURE INDEX

TABLE INDEX

ACKNOWLEDGEMENTS

Today is the day, I am starting to write my bachelor thesis, I could not be here if it were not for my parents support taking the hard decision to quit my job and start studying again to finish my career.

Certainly, I cannot forget anyone who has made in me a future, hopefully soon, graduate in engineering. I am deeply grateful to Prof. Gustavo Marrero Callicó who bet on me for this project, PhD Himar Fabelo Gómez who has helped me a lot in this challenge and Beatriz Martinez Vega who was my closest teammate.

Being repetitive I would like to express my special thanks of gratitude to my parents, my sister and my girlfriend for their love, support and understanding throughout my life.

Along this path, more people have supported me, my friends, classmates, teachers, and all those who contributed directly or indirectly to the completion of my studies.

RESUMEN

El estudio del patrimonio histórico de un pueblo es un trabajo arduo, complicado y delicado. Durante mucho tiempo el análisis de elementos arqueológicos encontrados en yacimientos cómo cerámicas, piedras, pieles y demás elementos ha requerido de su destrucción parcial o deterioro apreciable. Es por ello por lo que desde el Museo Canario se requiere el estudio sobre técnicas no invasivas para el análisis y la clasificación de las muestras encontradas en distintos yacimientos. Sus actuales métodos de clasificación requieren el envío de sus muestras a laboratorio y la destrucción de pequeñas partes de la muestra. En este documento se presenta una solución basada en imágenes hiperespectrales, que ha demostrado en otros campos su viabilidad y fiabilidad a la hora de realizar este tipo de clasificaciones.

Para poder generar un modelo supervisado de clasificación se dispone de una clasificación de muestras ya estudiadas con ICP-OES, ICP-MS y FRX que se distribuyen en 5 grupos elaborados por su pertenencia a la misma colada de lava. De estos grupos uno pertenece a la isla de Tenerife y el resto a la isla de Gran Canaria.

Las imágenes hiperespectrales (HSI) se tomaron en grupos de muestras localizados en una plantilla. Utilizando métodos de segmentación de imágenes se obtuvieron las máscaras de las muestras con las que se obtiene finalmente la firma espectral de cada muestra. Utilizando la firma espectral de cada muestra y sus etiquetas se generan distintos modelos SVM que han demostrado en otros campos su fiabilidad y su extensivo uso en HSI. Después de una etapa de optimización se obtiene el modelo final con el que poder clasificar posteriormente y evaluar sus métricas. Los resultados muestran una alta precisión (100%) para la clasificación en muestras de distintas islas, pero con algunas dudas

para estudiar en futuros trabajos con respecto a los grupos de coladas de lava de la misma isla.

Palabras clave: imagen hiperespectral; obsidiana; colada de lava; SVM; Islas Canarias; patrimonio arqueológico.

ABSTRACT

The study of historical heritage is an arduous, complicated, and delicate task. For a long time, the analysis of archaeological artefacts found in archaeological sites such as ceramics, stones, skins, and other artefacts requires their partial destruction or deterioration. For this reason, El Museo Canario has required the study of non-invasive techniques for the analysis and classification of the samples found at its various deposits. Its current classification methods require sending samples to the laboratory and the destruction of small parts of the sample. This bachelor thesis presents a solution based on hyperspectral imaging, which has been tested in other fields to be feasible and reliable for classification.

To generate a supervised classification model, a classification of samples already studied with ICP-OES, ICP-MS and XRF is available, which are distributed in 5 groups elaborated by their belonging to the same lava flow. One of these groups belongs to the island of Tenerife and the rest to the island of Gran Canaria.

Hyperspectral image captures were performed on groups of samples located in a template. Using image segmentation methods, the masks of the samples were obtained from which the spectral signature of each sample was finally obtained. Using the spectral signature of each sample and its labels, different SVM models have been generated, which have been tested in other fields to be reliable and widely used in HSI. After an optimisation phase, the final model has been obtained with which to further classify and evaluate its metrics. The results show a high accuracy (100%) for the classification of samples from different islands, but with some doubts to be studied in future works concerning groups of lava flows from the same island.

KEYWORDS: hyperspectral image; obsidian; lava flows; SVM; Canary Islands; archaeological heritage.

1. INTRODUCTION

This bachelor thesis aims to validate the classification of obsidians found in different archaeological deposits in the Canary Islands using a non-destructive technique, namely hyperspectral imaging (HSI). Currently, El Museo Canario, an entity in charge of researching the heritage of the islands, performs chemical studies of various elements found in these deposits to determine their origin. For this identification, the deterioration or partial destruction of the sample is inevitable. Therefore, it is vital to study and validate a non-destructive classification method, such as the HSI technology.

Currently, the use of HSI is wide and varied, from agriculture to medicine, for purposes such as classifying the quality of fruit or vegetables, or the delimitation of a brain tumour. Such wide use is no coincidence, HSI technology has several advantages over others. It allows the analysis of the chemical composition of the materials in the sample to be contact-less, non-ionising and non-destructive. The reason it allows the analysis of the chemical composition is due to it can capture a large number of wavelengths, obtaining information that the human eye is not capable of capturing.

1.1. Motivations

The motivation for this bachelor thesis is the need of museums and archaeological institutions for a non-destructive technique, without the need for laboratory intervention and accurate results of sample classification.

This concern on the part of El Museo Canario was transferred to the supervisors who, with experience in the field of hyperspectral imaging applied to fields such as medicine, decided to explore the application of HSI in this field.

As a result of all this, the classification based on islands, municipalities and sites was studied in a previous work [1]. This study was evaluated by the Canarian Museum, which requested the study based on lava flows instead of deposits, which leads us to this recent bachelor thesis.

1.2. Objectives and Methodology

The main objective of this bachelor thesis is to obtain a valid classification system of the archaeological heritage of the Canary Islands present in El Museo Canario. The elements to be classified are obsidians originating from deposits located in Tenerife and Gran Canaria. This classification will be based on lava flows instead of municipalities per museum request.

To achieve this objective, it is required to develop the following sub-objectives:

1. Knowledge of the technology to be used and the algorithms to be developed.
2. Automatically Segment the samples on the support. This segmentation allows obtaining the HS signature of each sample from a template with several ones in an automated process.
3. Development of the classification algorithms.
4. Design of the developed system. The design of the whole workflow to obtain the sample HS signature, prepare all the classes, the creation of the model classifier and its evaluation.

1.3. Memory Structure

This document is organized as follows:

Chapter 1: Introduction. This chapter explains the motivations for the development of this work and the main objectives. Finally, a brief explanation of the memory structure is provided.

Chapter 2: State-of-the-Art. This chapter covers the state-of-the-art of spectrometry and some methods currently used by El Museo Canario. Finally, hyperspectral imaging and its applications are described.

Chapter 3: Methodology The third chapter talks about the methodology of this project. Firstly, the available database and the new distribution suggested by the museum. Next the acquisition system and available images. Finally, the evaluation metrics used to measure the classifier.

Chapter 4: System Design The fourth chapter explains the system design. From the first phase, the template selection, to the last one, the parameter optimization.

Chapter 5: Results The fifth chapter presents all the results of this work. Presenting the results in groups of two, three and all the classes together. Also, some statistical metrics results and undefined sample predictions are presented in this chapter.

Chapter 6: Conclusion and future lines The last chapter deals with conclusions considering the results and proposes future work that can improve these results.

2. STATE-OF-THE-ART

In this section the necessary concepts for a complete understanding of this project are presented. It is focused on the classification of obsidian using hyperspectral technology. First, the reader will start learning about spectroscopy. Then, the different techniques used to analyse the stones or ceramics found in archaeological sites. After this, the concepts of hyperspectral imaging (HSI) and its different use in several applications will be introduced. Additionally, the relationship between archaeology and HSI which give origin to this final project. This section finish with the image segmentation and classification concepts.

2.1. Spectrometry

Spectrometry is a technique employed to measure the concentration of certain elements using the study of the interaction between matter and radiation. It indicates the components present in the sample and their concentration. The instrument to measure the concentration or amount of a specific element is a spectrometer or spectrograph. There are different spectrometry methods (mass, scattering, atomic emission, ultraviolet and visible light, etc), but only the most common techniques used in the identification of the composition of materials will be explained. In particular, Mass Spectrometry (MS) is the laboratory analytical technique for separating material components by their mass and electrical charge and, is the most common and most important spectrometry used in a laboratory, there are several types [2] i.e., electronic, chemical, or electrospray ionization, atmospheric pressure chemical ionization or photoionization etc.

2.1.1. Inductively Coupled Plasma Optical Emission Spectrometry (ICP-OES)

Inductively Coupled Plasma Optical Emission Spectrometry (ICP-OES) is a well-known chemical analysis technique, this method detects elements in the sample by using plasma and spectrometer [3]. When the sample receives the plasma energy, the elements are excited, and the atoms move to a higher energy position. As soon as the atoms return to the low-energy position, emission rays are released and those corresponding to the photon wavelength are determined by the spectrometer. The element type is measured depending on the position of the photon rays and the component of each element is determined based on the intensity of the rays. This technique can determinate qualitatively and quantitatively more than 60 different elements between them [4].

This technique is employed as the most powerful technique in many fields such as environmental safety, health [5], [6], bioremediation, food quality testing [7]–[9] and pharmaceutical analysis due to its accuracy and sensitivity, simultaneous analysis of multi-element, high throughput, and low costs [10]. But it is also being considered in petrochemical, metallurgical, geological, and nano-technological studies.

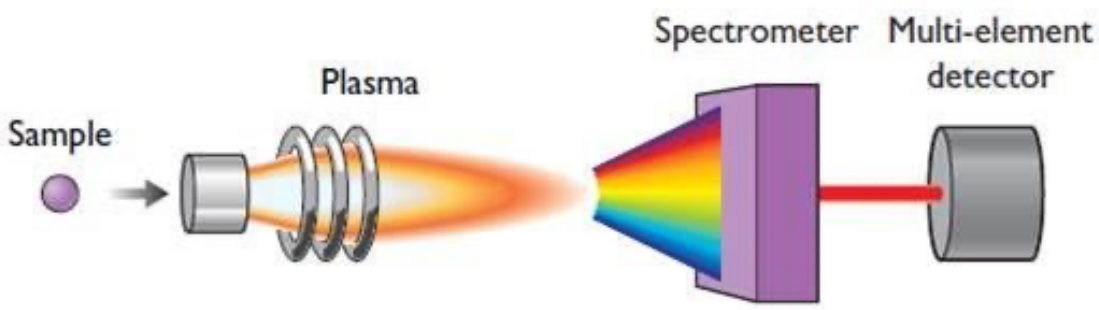

Figure 1: Schematic view of an ICP-OES [11]

2.1.2. Inductively Coupled Plasma-Mass Spectrometry (ICP-MS)

Inductively Coupled Plasma-Mass Spectrometry (ICP-MS) is an analytical technique to quantitatively and semi-qualitatively determine almost all elements of the periodic table which have an ionization potential lower than Argon (Ar) at low concentrations. The sample, in liquid form, is transported to the nebulizer system where it is transformed into an aerosol by the action of Argon gas. This aerosol is conducted to the ionization zone that consists of the plasma generated by Argon gas on magnetic field action induced by a high-frequency current. The plasma dissociates the molecules and removes an electron from the component forming ions which are directed into a mass filtering -mass spectrometer- [12]. Only one mass to charge radiation is allowed to pass through the mass spectrometer from entrance to exit. Each of the masses reaches the detector where its abundance in the sample is evaluated, due to the impact of the ions releases a cascade of electrons which creates an electromagnetic pulse. This pulse is compared with the standardized ones, also called isotopic fingerprint, to determine the concentration of the element. Figure 2 shows the standardized isotropic fingerprint for each element in the periodic table.

ICP-MS is widely used in many different areas like health [6], food quality control [7]–[9], [13]–[15], environmental, biology, agriculture [16], [17], industrial, etc. According to Joint ALSSA-JAIMA-Eurom II Global Laboratory Analytical Instruments Booking Report, over 15% of all new instruments purchased for trace metal analysis are ICP-MS instruments [12].

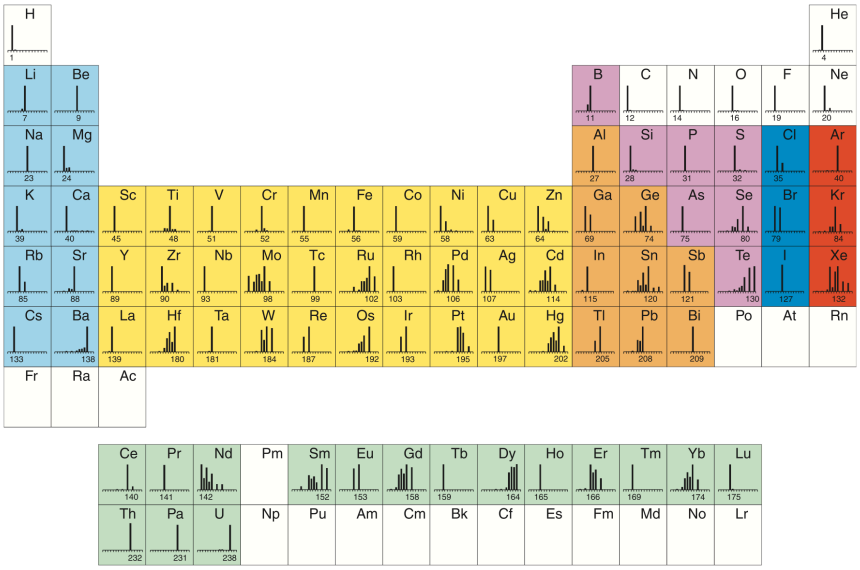

Figure 2: Elements analysed by ICP-MS (elements in colour) [12]

2.1.3. X-Ray Fluorescence (XRF)

X-ray fluorescence spectrometry (XRF) is a method used to perform relatively non-destructive chemical analyses of materials, such as minerals, sediments, rocks, etc. This method relies on fundamental principles that are common to several other instruments involving the interaction between electron beans and X-Ray with samples. When this sample is excited with X-rays, they can ionize it. If this energy is enough to remove an internal electron, the atom becomes unstable, and an external electron will replace it. At this point, energy is radiated, known as fluorescent radiation, which is characteristic of the transition between specific electronic orbitals of a particular element. The resulting X-rays are used to determine the presence of elements in the sample.

Nuclear beans and lasers are becoming increasingly important as analytical tools in art and archaeology for dating and characterization studies [18] due to the portability, quickness, and relatively non-destructive analysis to obtain the first information about the samples. There are plenty of portable or desktop XRF devices with compact design, low weight and even battery-powered.

Figure 3: Spectrace 9000 TN PXRF measuring mural paintings of modern artist Spyros Papalou- kas from the Amfissa Cathedral, near Delphi, Greece

2.2. Hyperspectral imaging

2.2.1. Electromagnetic spectrum

The electromagnetic spectrum [19] is the distribution of all ranges of electromagnetic radiation, i.e., the distribution of this energy emitted in form of waves based on their wavelength. The types of electromagnetic radiation range from radio waves (longest wavelength) to gamma-ray (shortest wavelength). The shorter the wavelength, the higher energy it emits.

The electromagnetic spectrum is important in this study due to the interaction of the electromagnetic radiation with materials is different along the spectrum, which allows the hyperspectral signature of materials to be determined.

2.2.2. HSI fundamentals

HSI is a technology that has evolved from spectroscopy, combining spectroscopy with digital imaging. The result of both technologies, spatial and spectral information provides a huge amount of data in the form of an HSI cube, as can be observed in Figure 4. Each pixel provides spectral information, which is the radiance of the materials within the area that is covered by the pixel. The fact is that all the materials reflect, absorb, or emit electromagnetic energy in different wavelengths. For this reason, hyperspectral (HS) cameras use different sensors to measure them in hundreds of spectral wavelength bands and these values normally behave like a continuous spectrum. This continuous spectrum is the so-called spectral signature, and it is the equivalent of a fingerprint. This means that each material has a unique spectral signature, as shown in Figure 4.

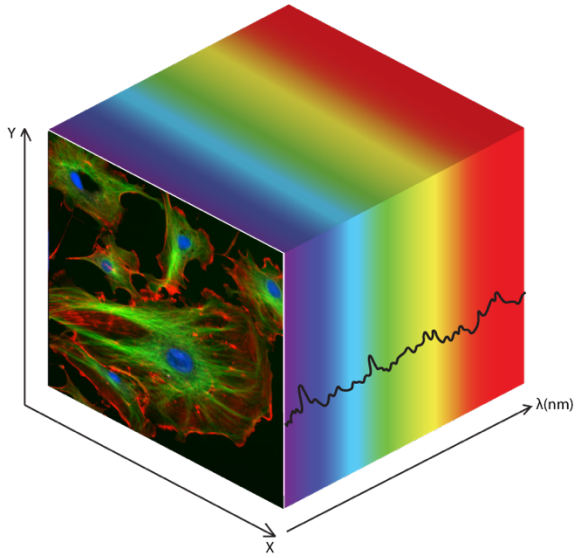

Figure 4: HSI cube

HSI is a growing technology originally developed for remote sensing, but nowadays it is used in many research fields, such as medicine [20]–[26], food quality [27]–[30], defence [31], [32], drug identification [33]–[35], art [36]–[39], and archaeology [1], [37], [40], [41].

2.2.2.1. Remote Sensing

Remote sensing refers to the collection of information about an object without being in physical contact with it [42], typically using satellites, aeroplanes or drones equipped with multispectral or HS cameras. HS data processing is widely used for the detection and identification of surface, topographic and geological features on The Earth [18] [19].

The National Aeronautics and Space Administration (NASA) has employed remote sensing for the detection of ice, water, and snow on The Earth processing HS data with images captured from NASA's Earth Observing-1 satellite in 2012 [44].

Remote sensing is also used in agriculture, e.g., to measure the content of leaf water on agricultural productions [45], [46] or to estimate some performance keys on cereal lands [47].

2.2.2.2. Medical

Researchers have found many applications of the HSI in the medical field due to the interaction between electromagnetic radiation and tissues, which provides useful information for diagnostic and non-invasive techniques. From atrial ablation lesions [48], [49] or cervical neoplasia [50], [51] to cancer detection [20]–[24], [52].

Several studies are related to the identification of brain cancer during surgical resection [20], [23], [24], [52]. This procedure is critical and challenging for neurosurgeons as there is no perfect tool to delineate the tumour to be resected. The accuracy with which it is delimited will be the key to the treatment success, due to less malignant tissue resected will make the patient relapse and good tissue resected could limit vital functions.

2.2.2.3. Cultural Heritage Conservation

The conservation of cultural heritage, such as paintings, objects, buildings, etc., is highly important for human beings and requires high technology and hard work to carry out. This field has used many different techniques for identification, conservation, and preservation throughout its history. HSI has come into this field [41], [53], [54] to replace other invasive techniques as mentioned above.

Polychronis Kolokoussis et al. [41] studied the degradation of materials of four buildings about 2,400 years old that have been affected by climate change in recent decades using HSI technology. It aims to determine whether the combination of detailed 3D texture models and HSI can provide a tool for the creation of degradation maps as shown in Figure 5.

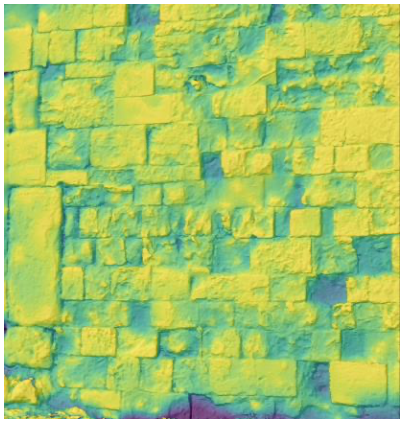

Figure 5: Material loss estimation using the valley depth algorithm (hillshade has been added as background for easier interpretation of the results) [41]

In [53], a panel painting from the collection of the Ethnographic Museum of Slovenia was employed to evaluate the potential of HSI for the evaluation of heritage objects. Four different HS cameras were used to scan the document and obtain the HS data and merge it with other reference methods databases to allow the identification of the materials used by the artist on the panel. This allows the identification and characterization of the colourants, binders, and coatings

originally used or later additions, i.e. In Figure 6 it is possible to identify hidden underdrawings detected by HSI.

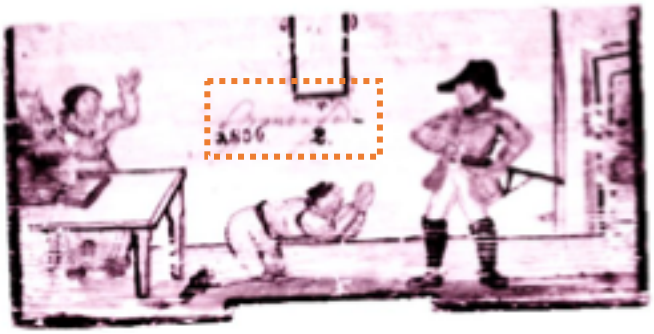

Figure 6: The hidden underdrawings in the inner layer of the studied beehive panel painting were detected using the FX17 and SWIR hyperspectral cameras (NIR wavelength: 1500 nm) [53]

2.3. Image segmentation

2.3.1. Morphological segmentation

Image segmentation is a process to divide a digital image into regions or parts based on some particular feature such as discontinuities in pixels values, shape or colour differences [55]. Morphological segmentation is an image segmentation based on the shapes of interest to split the sample.

Morphological toolboxes can be used in many fields, but one of the most incredible uses is in medical procedures. Image segmentation is used for distinguishing between tissue types during a neurosurgical operation [56]. During the procedure, pathologists stain body tissue with haematoxylin and eosin and take a digital image of the brain, then using deep learning they segment the image into tumour and background. This segmentation helps the surgeon to resect the tumour effectively.

2.3.2. Threshold segmentation

This is the simplest method for image segmentation. In this method, pixels are separated according to grayscale intensity value. A threshold value is

employed at which the grayscale image is converted into a binary image, with a value of 1 for those equal to or above the threshold and 0 for those below. In threshold segmentation, methods can be used that adapt the threshold value at each pixel based on image characteristics, this is called automatic thresholding.

There are many automatic thresholding methods based on different algorithms and they can be classified as histogram shape, clustering, entropy, object attribute and spatial methods. The most famous method is Otsu [57], that it determines the threshold by minimizing intra-classes intensity variance.

The reason behind the automatic thresholding could be explain with a landscape photo. Using the same threshold (manual thresholding) for the land and the sky would provide a low-quality segmentation, in the other hand if the threshold used in the sky is different from the one used in the land it will provide better segmentation.

2.4. Image classification

For imaging classification there are different methods based on Machine Learning (ML). There are three ML methods for classification, supervised, unsupervised and semi-supervised. The main difference between supervised and unsupervised methods is mostly in the use of labelled data: supervised algorithms use labelled inputs to generate the models, while unsupervised algorithms try to find some patterns in the input data (unlabelled data). The latest one, semi-supervised uses a mix of both approaches, patterns, and labelled data.

In this work, one of the most used supervised method, Support Vector Machine will be used due to the number of researches done with HS images classification.

2.4.1. Support Vector Machine (SVM) Classification

Support Vector Machine (SVM) Classification is a supervised learning technique that allows the separation of the different groups that make up the data by means of hyperplanes, see Figure 7. The main goal of SVM is to find a maximum marginal hyperplane which best divides the dataset into different classes. This means that the hyperplane will be the best division that can be made with the different classes using such values.

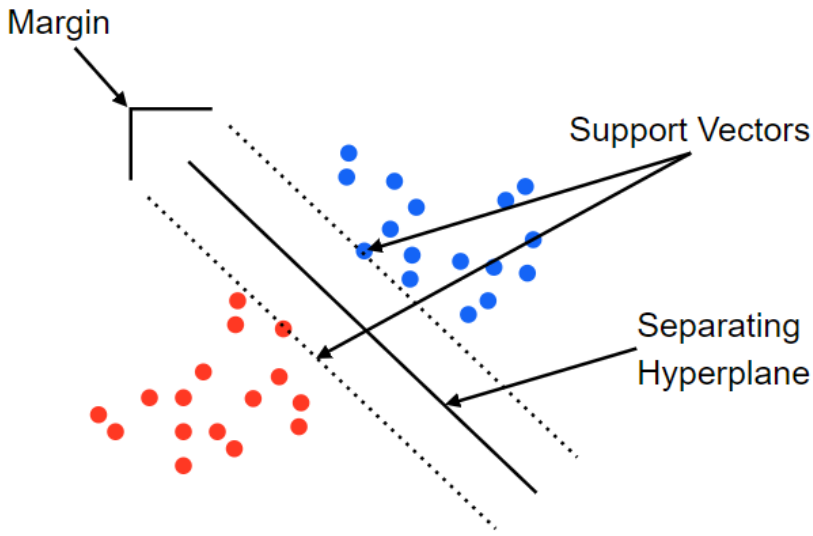

Figure 7: Representation of two classes of SVM classification with lineal hyperplane [58]

In Figure 7 is possible to see the support vectors, those data points closest to the hyperplane and the margin that could be defined as the gap between the support vectors of each class.

The calculation of data point division depends on a kernel. Exists different kinds of kernel functions: Linear, Polynomial, Gaussian, and Radial Basis Function (RBF). This means that the determination of the hyperplane is based on the kernel function. Table 1 shows the different kernel functions for support vector machines and their different formulas.

Table 1: Kernel functions formulas

Kernel Function Name	Formula
Linear	$G(x_j, x_k) = x_j' x_k$
Polynomial	$G(x_j, x_k) = (1 + x_j' x_k)^q$
Gaussian or RBF	$G(x_j, x_k) = \exp(\|x_j - x_k\|)$

Support vector machine classification is one of the most used models in all kind of fields [59]–[62].

2.5. Summary

In this section all the theorical concepts necessary for the development of this project are presented. Spectrometry and the methods used in El Museo Canario to analyse its samples (ICP-MS, ICP-OES, FRX) are explained. Then, the concepts related to Hyperspectral Image and its usages are covered. Finally, Image Segmentation and Classification with their different methods used in this work are explained like threshold segmentation and SVM classification.

3. METHODOLOGY

This section covers the new definition of classes given by El Museo Canario based on lava flows instead of municipalities that prompted this work. The HS acquisition system and the available HS images are explained too. Evaluation metrics used to measure the classifier are defined at the end of the section.

3.1. Samples database and description

3.1.1. Distribution from a previous master's final work

In the previous work [1] 69 obsidians provided by El Museo Canario and the Department of Historical Sciences of the University of Las Palmas de Gran Canaria were studied. These samples were divided based on deposit (orange), municipality (green) and island (blue), as shown in Figure 8.

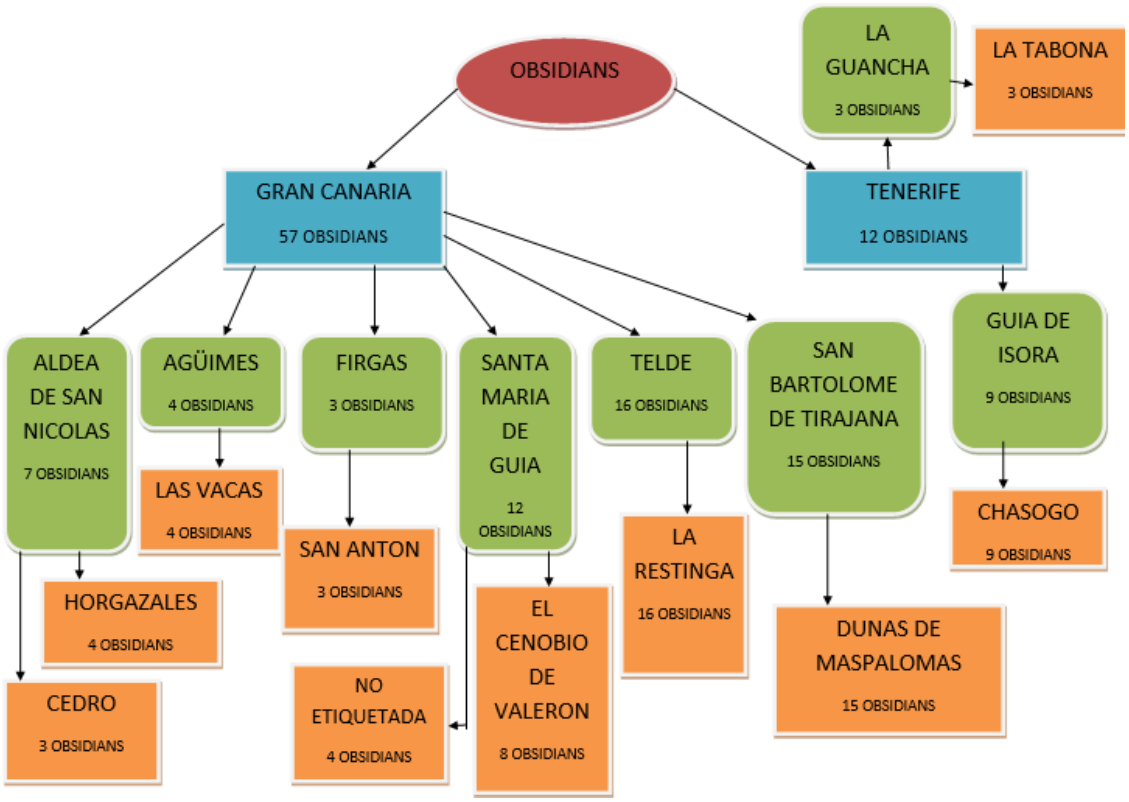

Figure 8: Previous work organisation scheme of the obsidian sample database [1]

Table 2 displays the labelling created to facilitate the processing of the images based on the three levels (island, municipality, and deposit) to match the distribution provided by the archaeologist. The class identifications are represented by 3 digits, the first represents the island, the second the municipality and the last the deposit where the sample was found, for example, 122 is from Hogarzales in Aldea de San Nicolás (Gran Canaria). This table also indicates the number of samples found in each deposit, in total 57 samples for Gran Canaria and 12 for Tenerife.

Table 2: Obsidians labelled by deposit, municipality, and island on previous work [1]

Class ID			**Class Name**			**Number of Obsidians**		
LEVEL 1	**LEVEL 2**	**LEVEL 3**				**LEVEL 1**	**LEVEL 2**	**LEVEL 3**
100	110	111	Gran Canaria	Telde	La Restinga	57	16	16
	120	121		Aldea de San Nicolas	Cedro		7	3
		122			Hogarzales			4
	130	131		Firgas	San Antón		3	3
	140	141		Agüimes	Las Vacas		4	4
	150	151		Santa María de Guía	El Cenobio		12	8
		152			No label			4
	160	161		San Bartolomé de Tirajana	Dunas de Maspalomas		15	15
200	210	211	Tenerife	Guía de Isora	Chasobo	12	9	9
	220	221		La Guancha	La Tabona		3	3

Not all the samples were captured in previous work and one of the objectives of the current work was to capture most of the rest. But due to the access difficulty to the obsidians and the acquisition system, since this was used by other projects in the medical field made it impossible.

3.1.2. New distribution

El Museo Canario has suggested a new distribution based on the analysis of samples by ICP-OES, ICP-MS and FRX techniques. The reason for the change is that several tanks can share lava flows and thus be very similar chemically, which would result in similar spectral signatures. This may be the reason why the results of the previous work were not very encouraging.

This new distribution consists of five groups: TAB, N1, N2, N3, HOG and the rest of the samples are undefined, ANNEX shows the samples that belong to each class. The first one, TAB is mainly composed of the samples found in Tenerife, and the remaining four are those corresponding to the samples found in Gran Canaria. One with the samples found in Hogarzales, Vacas and Cedro and three new groups for future study. The undefined samples are cases to be studied, some of them could be analytical issues or samples not discarded previously by experts (deteriorated surfaces, presence of titanium white in the acronym, etc).

Table 3, Table 4, Table 5, Table 6 and Table 7 contain the list of samples that are captured for each defined group.

Table 3: TAB samples captured

TAB		
CHA-27	CHA-28	CHA-30
CHA-31	CHA-33	CHA-35*
CHA-36	CHA-39	TAB-1
TAB-2	TAB-3	

* Two captured samples are called CHA-35.

Table 4: HOG samples captured

HOG		
ANT-158	CED-18-114	CED-C-155
CED-T-113	CNB-147-A	CNB-147-B
CNB-149-A	CNB-151	CNB-152
CNB-153	CNB-155	DUM-78
DUM-80	DUM-82	DUM-83
DUM-85	DUM-88-1	DUM-90
DUM-91	HOG-38-1368-73	HOG-38-816-65
HOG-38-818-65	RES-10-180	RES-10-181
RES-10-183	RES-10-184	RES-10-185
RES-10-187	RES-10-190-I	RES-10-192
RES-10-193	RES-7-174	RES-7-175
VAC-1-119	VAC-2-120-A	VAC-2-120-B
VAC-2-120-C		

Table 5: N1 samples captured

N1		
DUM-77	DUM-89	RES-7-173
RES-7-178		

Table 6: N3 samples captured

N3		
CNB-149-B	DUM-81	DUM-92

Table 7: Undefined samples captured

Undefined		
ANT-157	DUM-88-2	DUM-93
RES-7-179	RES-10-194	

3.2. Acquisition system

The acquisition system consists of two cameras, one SWIR and one VNIR camera, both were placed on a scanning platform with a light system to illuminate the stones. Figure 9 illustrates the acquisition system, which is composed of the illumination source (1), the SWIR camera (2), the VNIR camera (3), and one linear displacement (4). Table 8 contains the specifications of the HS cameras.

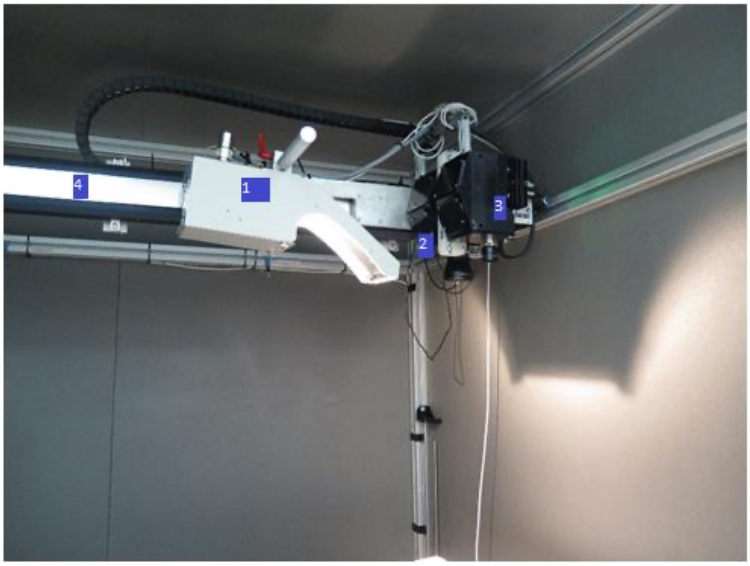

Figure 9: Acquisition system [1]

Table 8: Specifications of the HS cameras

Characteristic	Headwall Hyperspec® SWIR	Headwall Hyperspec® VNIR E-Series
Spectral range (nm)	900-2500	380-1000
Spectral resolution (nm)	12	3
Spectral bands	267	923
Spatial bands	384	1600
Dispersion/Pixel (nm/pixel)	6	0.65
Capture type	Push broom	Push broom
Pixel Pitch (μm)	24	6.5

These cameras are push-broom type, as we have explained before they take the HS cube moving from one side to the other to catch the whole one. This movement is controlled by the camera using a linear actuator driven by a stepper motor, and it is synchronous with the shooter.

The illumination system is composed of halogen light and a power supply. The intensity of the light is controlled by a regulator to allow the user to adjust the parameters dependent on the light.

3.3. Available images

The database is composed by 19 HSI from 5 compositions with a template where the sample is situated inside a square, these compositions are called 'Toma' from 1 to 5. Each composition has different sample positions, formal is the initial one, the reverse side of the sample and the reverse and rotate is with the sample in reverse position and rotated. Some compositions have two shots, for example, Reverse Composition 1.

One example of this template is shown in Figure 10. In this case, it is for Composition 5, squares have different sizes as the samples captured are higher than the regular square for all the templates, so it was customized.

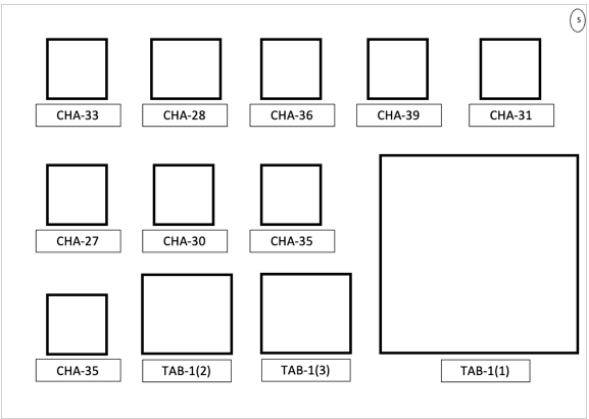

Figure 10: Example of template

Table 9: Different compositions captured in previous work

	Formal	Reverse	Reverse & Rotation	Total
Composition 1	1	2	2	5
Composition 2	1	1	1	3
Composition 3	1	2	1	4
Composition 4	1	1	1	3
Composition 5	1	2	1	4

3.4. Evaluation metrics

The following is an explanation of the accuracy, precision, sensitivity, specificity and f1 score metrics, which have been used in this study. In general, these metrics evaluate the performance of the generated model.

3.4.1. Accuracy

The accuracy of a classifier refers to how close an estimation comes to the known value. It can be defined as the number of correct estimations divided by the total of estimations. Using TF as True Positive, TN as True Negative, FP as False Positive and FN as False Negative, accuracy for binary classification is:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

3.4.2. Precision

Precision is the true positive rate to predicted positives, in binary classification is how truly the positive cases are. It is defined by the equation below.

$$Precision = \frac{TP}{TP + FP}$$

3.4.3. Sensitivity

Also called recall, sensitivity is the true positive rate to total, in binary classification is how well the classifier detects the positive cases. It is defined by the equation below.

$$Sensitivity = \frac{TP}{TP + FN}$$

3.4.4. Specificity

Specificity is the true negative rate, in binary classification is how well the classifier detects the negative cases. It is defined by the equation below.

$$Specificity = \frac{TN}{TN + FP}$$

3.4.5. F1 Score

F1 Score is the harmonic mean of Precision and Sensitivity, in binary classification is one of the most common measures [63] to rate how well a classifier is when there are unbalanced classes. It is defined by the equation below.

$$F1\ Score = \frac{2 \times Precision \times Sensitivity}{Precision + Sensitivity}$$

3.4.6. Jeffries-Matusita Distance

Jeffries-Matusita [64]–[67] distance calculates the separability of a pair of probability distributions. Provides a reliable criterion of classes separability because it behaves like the probability of a correct classification. Its values range from 0 to 2, where 0 means that the classes are not separable and 2 that they are completely separable. It is defined by the equation below where B is the Bhattacharyya distance.

$$JM_{ij} = \sqrt{2(1 - e^{-B})}$$

$$B = \frac{1}{8}(\mu_i - \mu_j)^T \left(\frac{C_i + C_j}{2}\right)^{-1} (\mu_i - \mu_j) + \frac{1}{2} ln \left(\frac{\left|\frac{C_i + C_j}{2}\right|}{\sqrt{|C_i||C_j|}} \right)$$

Where *i,j* are the compared classes, C_i is the covariance matrix of class *i*, μ_i is the average vector of the same class and $|C_i|$ is determinant of the covariance matrix.

3.4.7. Rand Index for clustering

The rand index is a measure of the similarity between two data clustering, like accuracy in supervised classification, the rand index measures the percentage of correct decisions. Using TF as True Positive, TN as True Negative, FP as False Positive and FN as False Negative is:

$$Rand\ Index = \frac{TP + TN}{TP + TN + FP + FN}$$

3.5. Summary

This section presents all the methodology elements. First, the sample database, with the differences between the new classification provided by El Museo Canario and the classification of the municipality. Then, the acquisition system and the available HS images are explained. Finally, the evaluation metrics used to measure the classifier are defined.

4. SYSTEM DESIGN

In this section all the different phases of the system are explained, starting with the template selection to the creation of the SVM model with parameter optimization.

The system is designed around the available database, using the captures already obtained a few steps are needed before classification to create a dataset ready for it. Figure 11 shows the complete system workflow, from the template section to the classifier model creation. In this workflow there are steps in light orange that represent those steps done only with the binary image and the blue steps that represent those done with the entire HS cube. Finally, the green step, evaluation metrics, is done only with the best features, not the entire cube. This is done to improve the system performance since working with the whole cube is quite demanding.

Figure 11: System workflow diagram

In the first phase, each template box containing the sample is automatically separated from the others. Then, in the second phase, once the box is separated from the template, the sample inside the box is obtained. For this, a binary mask is created to know which pixels represent the sample, and the spectral signatures of each pixel which compose the sample are obtained. This process is carried out for different samples, using prior knowledge of which obsidian corresponds to each class. In the third phase, two subclasses are generated, one for training and one for testing. The fourth phase is the generation of the model with the training classes and using the test ones for hyperparameter optimization. In the last phase,

the generated models are evaluated. MATLAB software is employed throughout each of the phases.

4.1. Template selection phase

The first phase of this system is the selection of the template, i.e., the cropping of the squares in the template. This process is fully automated in different scripts. The first function of the process is employed to detect the square shapes in one binary image created from a snapshot cube.

Morphological analysis is performed to find the rectangular shapes presented in the image. This morphological analysis needs some preparation, e.g., it is necessary to fill in all possible gaps found in the image. The reasons for this filling are due to irregular samples, reflections in the samples and the names of the samples.

After finding all the possible rectangular shapes, the following steps are necessary: first, the removal of those pixels which are on the edges, as they may come from the page, from shadows, or from anomalies in the template. This deletion is performed using the position of the vertices. If any of the vertices are on the margin of the template, they are removed. Based on the templates used, the deletion of 5 pixels is enough to remove the undesired rectangles. However, this parameter is not fixed, and the code is prepared to easily change this value if it is necessary to employ other types of templates.

After removing the anomalous shapes on the image margin, the next step is that the system increases the size of each shape to include the surrounding pixels for each shape, taking care of some anomalies in some snapshots like Compositions 2, 3 and 5. Increasing the size of the largest frame too much will overlap it with the surrounding frames, so this needs to be customised for some snapshots. This scaling is necessary to blend all the shapes that are in each box

since the sample could be one shape, the surrounding box another one, or some anomalous like each letter of the sample name could be more shapes.

After scaling all the shapes, the third step is to overlap and merge the detected shapes since they are different parts of the same portion of the template. That is, the sample could be one shape, the box around it another and the text for the name of the sample could be many of the shapes, so they all need to be blended into one. Using MATLAB function for shape overlaps, a matrix with all overlaps is retrieved. The size of the matrix is *NxN* where *N* is the number of shapes, and the content of the matrix is 1 when the corresponding pairs of elements of the polyform objects overlap, 0 otherwise. Using these overlaps, the final shape containing them is created for next steps.

Finally, the detected shapes are employed to cut the image and work in the following steps only with this part of the template.

4.2. Segmentation phase

The segmentation procedure is first performed by creating a binary mask of the image cropped in the previous phase, which allows detection of the sample. Again, MATLAB morphological analysis is used to detect the components connected by the 8-pixel connectivity. This means that the pixels are connected in all directions, so they cannot be lines or dots. In addition, the sample should be solid and with some circularity due to the stones are not perfect squares. Finally, with all the above restrictions, the sample will be within the largest component found in the analysis.

Based on these assumptions, the hyperspectral cube is cropped with the coordinates of the coincident shape.

The result of this phase is an object with all the smallest cubes that can contain the samples. So, all the harder processing will be done only with smaller cubes improving the performance of the whole system.

4.3. Pre-processing

4.3.1. Binary Mask phase

After segmentation, it is necessary to select only the pixels that represent the sample. Using a binary mask of the cube, only the pixels representing the samples are obtained in a vector. This binary mask is created with the mean of the Z coordinate of the cube to avoid small differences in the HS signature of each pixel, removing anomalous pixels. Figure 12 shows all the binary masks from Composition 1 with the indexes used to select the samples for the creation of the data vector.

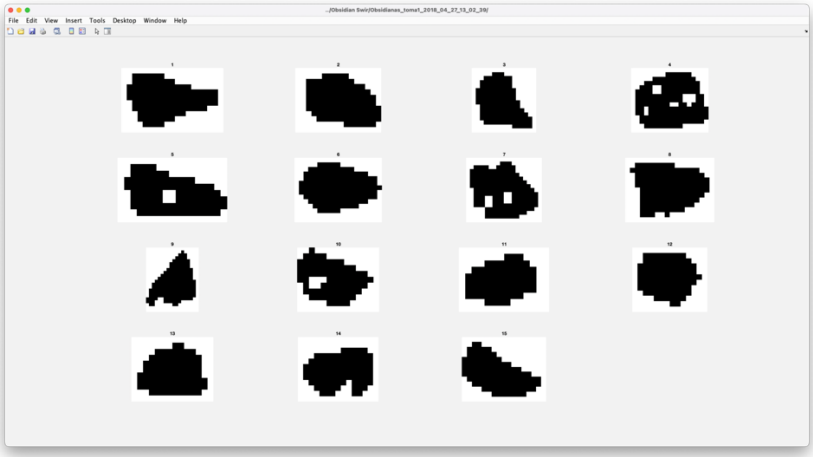

Figure 12: Binary masks from Composition 1

This binary mask allows the system not only to obtain the pixels corresponding to the sample. Also, it allows for eliminating light reflections on the obsidian surface that would have an anomalous HS signature and could prejudice the generation of the model. In samples 4, 5, 7, 9, 10 and 14 in Figure

12 there are holes or indentations in the sample masks that represent these types of anomalous pixels.

For this removal, a threshold of 0.4 has been based on different tests on several compositions. Some of them have customized thresholds in order to clean up with the highest accuracy. Figure 13 illustrates the differences between binary images with the 0.6 and 0.4 threshold values. Samples with imperfections such as 4, 5, 7, 10 and 14 have different shapes in their imperfections due to this threshold, and the classification accuracy can be affected by these edge pixels.

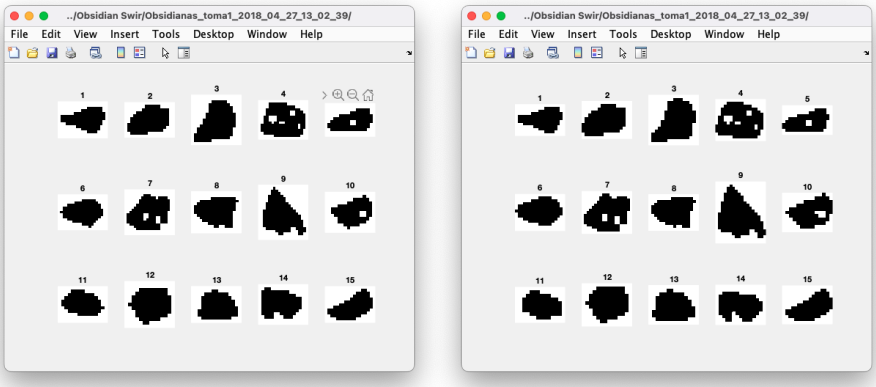

Figure 13: Binary image comparison between threshold 0,6 (left) and 0,4 (right)

4.3.2. Class generation phase

All the previous steps correspond to a single sample. However, it is necessary to create sets of samples of each type defined by the archaeologists. Each set of samples would correspond to a class, and once the classes are defined, the model for the classification of future samples could be generated.

For classification purposes, training and testing subclasses are needed. So, for each class, one for creating the classification model, and the other for estimating metrics such as accuracy, sensitivity, and specificity of its model. An important constraint for the creation of the model is that any sample employed in the training subclass cannot be used for testing due to it could distort the results. So, in this work, there are two different approaches: the first is to do the splitting

automatically by providing a script with an array of cells and the second one is manually done. In both approaches, each cell will correspond to a sample, and the automated script will split the samples into approximately 80% for the training purpose and the rest of the samples for testing.

For the automatic approach, a dataset has been generated with the four classes available for HS captures TAB, HOG, N1 and N3. Each class is defined by a set of cells containing a matrix of pixels of each sample obtained by applying the binary mask as a selection of the pixels of the cube. These sets of cells are obtained from the selection of the samples in each of the compositions using the information provided by the archaeologists from the chemical analysis and saving them in the matrix of cells. To improve the results, the three versions of each composition, original, inverted and inverted and rotated, are used, obtaining three times more samples and therefore more spectral signatures for the generation of the model.

In the manual approach, the dataset has been created based on the classification shown in Table 10 (all the corresponding samples for each class are in Table 44, Table 45, Table 46, Table 47 and Table 48 at ANNEX I). For each class, the dataset has two different subclasses, training, and test, both generated by all the samples in the three compositions, original, inverted and inverted and rotated. As can be seen, the classes are very unbalanced, with the HOG class with 24 samples for training and 11 samples for testing, while N3 has only two samples for training and only one sample for testing.

Table 10: Number of samples in the new distribution per composition, deposit, and type

Class	Type	Deposit	Composition	Samples
HOG	TRAIN	RES	TOMA1	7
		CED	TOMA2	2
		HOG	TOMA2	2
		ANT	TOMA2	1
		VAC	TOMA2	1
			TOMA3	1
		CNB	TOMA3	5
		DUM	TOMA4	5
	TEST	RES	TOMA1	3
		CED	TOMA2	1
		HOG	TOMA2	1
		VAC	TOMA2	1
			TOMA3	1
		CNB	TOMA3	2
		DUM	TOMA4	2
TAB	TRAIN	CHA	TOMA5	6
		TAB	TOMA5	2
	TEST	CHA	TOMA5	2
		TAB	TOMA5	1
N1	TRAIN	RES	TOMA1	1
		DUM	TOMA4	2
	TEST	RES	TOMA1	1
N3	TRAIN	CNB	TOMA3	1
		DUM	TOMA4	1
	TEST	DUM	TOMA4	1

In this work, the process of obtaining the HS signature of each sample for the creation of the class is created in a way that eliminates the need for the user to select the region in which the obsidian is located as described in previous steps, this is done with a script that works as an assistant. In the first step, it reads the different Compositions that exist in the folder dedicated to it and displays a window with the list (as it can be seen in Figure 14) to select the composition that will be used to obtain the corresponding samples to the class that the user wants to generate. Once the capture is analysed to perform the steps described above (template and sample segmentation), the template is displayed to the user (Figure 15) with the automatically detected obsidians highlighted in different colours, and also in another window, the binary mask of all the samples with a numerical index in each one of them. In addition to the two windows, a dialogue box appears, shown in Figure 16, where it requests which indexes the user wants to include in the class, identifying the index in the binary mask with the corresponding one in the template, the user can select the ones that belong to the class to be obtained. Once the array with the indexes has been provided, an array with all the pixels of each of the chosen samples will be returned.

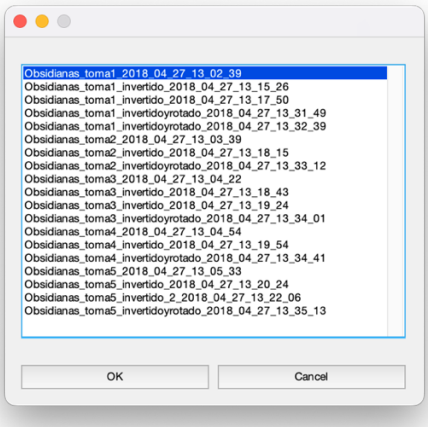

Figure 14: Composition selection

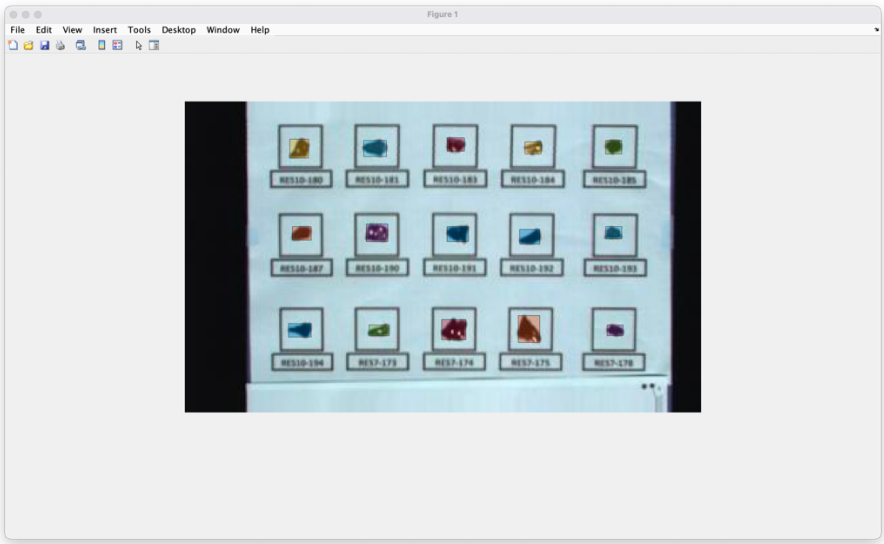

Figure 15: Composition with samples automatically highlighted

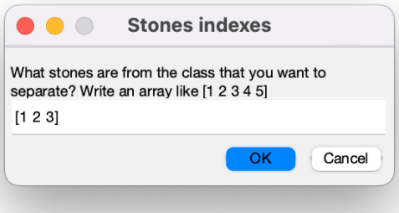

Figure 16: Dialog box asking for samples indexes

For the whole class, a few executions are needed. First, it is needed one for each composition version and some of the classes are in different captures, so the wizard needs to be executed a few times for each. This wizard differs from the manual to automatic approach, as the automatic workflow needs to keep the samples separated to be able to select automatically the stones used for testing and training, and in the manual one, the subclasses are already defined and need to be generated to selecting the samples for testing and training in different executions.

The dataset created in manual workflow contains each subclass for each kind of shot, for example for the HOG class the data set is composed of each shot, formal, reverse and reserve and rotate, with subclasses for test and train

separately at Composition 1, 2, 3 and 4. This makes four compositions times three shots times two subclasses (test and train) so 24 different sub-subclasses in HOG group. This is done to avoid composition or shot issues, as test and train contain samples for each composition and each shot. In N1 and N3 group cannot be done due to low number of samples in each composition, only one.

There is a special situation in shot 2 inverted there are two subclasses, this is because the binary threshold had to be adjusted for different samples as the obsidian catalogued as VAC-2-120-B cannot be obtained with the same threshold as the rest of the samples in the template due to the poor quality of the shot.

Finally, an array of HS signatures is created for each subclass, testing, and training, with all the shots mixed, original, inverted and inverted and rotated for being used in the next step. Table 11 contains all the number of samples and pixels for each class used for the creation of the models in this work and its corresponding results.

Table 11: Characteristics of the used classes

Class	Subclass	Number of Samples	Number of Pixels
HOG	TRAIN	24	12540
HOG	TEST	11	5585
TAB	TRAIN	8	8818
TAB	TEST	4	2288
N1	TRAIN	3	874
N1	TEST	1	264
N3	TRAIN	2	567
N3	TEST	1	216

4.4. Model generation phase

The classification method employed is the SVM algorithm. For this, it is necessary to install the MATLAB Statistics and Machine Learning Toolbox, which contains well-known and fully implemented functions for SVM. Because there are several classes, all possible combinations are evaluated, first biclasses evaluation (HOG vs TAB, HOG vs N1, etc) then multiclass in combinations of three (HOG vs TAB vs N1, HOG vs N1 vs N3, etc) and finally the four classes, so for the model generation, this work provides an automatic selection of SVM classifier functions providing the kind of kernel (linear, Gaussian or RBF) and the number of classes.

For the execution of the classifier, this work provides a dialog box, shown in Figure 17, where the user can select which classes want to use to create the classifier, the SVM kernel and if the user wants to balance the classes, as there are two classes (N1 and N3) with a big difference with the other two (HOG and TAB) and this could affect to the results. This dialog box is prepared for two and three classes to help in future works to create new models with new classes or the same used in this work but with more samples in the dataset.

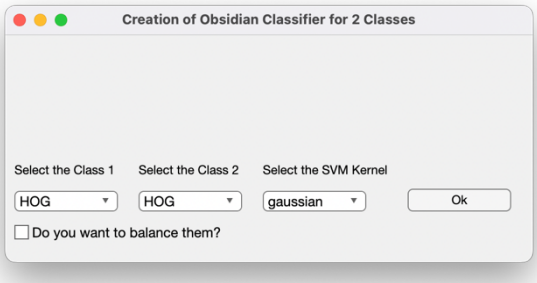

Figure 17: Classifier menu for two classes

When selecting the classes, the script will load the test and training classes for each class from their files and will execute the classifier. If the user has selected to balance the classes, the classes will be limited to the number of pixels of the

smallest class. Before balancing, to avoid the model being generated with only a few samples from the bigger classes the pixels are randomly reordered. In this way, the model will be created with pixels corresponding to all samples. As the test classes are also quite unbalanced, when activating the option, test classes are also balanced, so the metrics are more accurate.

Before creating the SVM template, a random cross-validation partition with five folds is created to execute the model creation in five iterations for feature selection. This means that the script is creating a random partition of the dataset, then developed an SVM model for each partition and evaluates them to check the best hyperparameter.

4.5. Parameter optimization phase

After the feature selection, the final SVM model is created only with the selected features in the previous steps. This step is called parameter optimization as the new classes to be predicted are only predicted by the selected features instead of the whole HS signature which is faster.

Finally, with this latest model, the metrics are created for the test dataset, the model is used to predict all this testing data and they are compared with the real data label to measure accuracy, sensitivity, and specificity.

4.6. Summary

This section explains each of the phases of the system, starting with the template selection, then sample segmentation, pre-processing chain, the model creation and finally parameter optimization.

5. RESULTS

This chapter present the results obtained after analysing the different classifications based on lava flows after mixing the different shots taken from the same sample.

All classes are evaluated by performing all possible combinations. First, the classes are evaluated by groups of two i.e., N1 vs N3, HOG vs TAB, etc. Then, by groups of three i.e., N1 vs N3 vs HOG, etc; and finally, all four classes at the same time.

However, before performing the classifications, a previous study is performed to identify the kernel with the best performance. In this case, all the classification results are performed with the Gaussian kernel due to the fact that this, on average, provides a slightly higher accuracy than the lineal and RBF kernels. See Table 12 for this comparison, the N1 and N3 classes are employed because the use of TAB or HOG classes could distort the results and make it difficult to compare the kernels.

Table 12: Comparison between kernels with N1 and N3 classes

Kernel	Accuracy	Class	Precision	Sensitivity	Specificity	F1 Score
Gaussian	41.87%	N1	0.4744	0.5265	0.2870	0.6435
		N3	0.2116	0.2870	0.5265	0.4980
Gaussian balanced	53.01%	N1	0.4082	0.4630	0.5972	0.6969
		N3	0.5265	0.5972	0.4630	0.6898
Linear	55%	N1	0.55	1	0	0.7097
		N3	0	0	1	NaN
Linear balanced	46.53%	N1	0.3797	0.4167	0.5139	0.6316
		N3	0.4684	0.5139	0.4167	0.6379
RBF	45.62%	N1	0.5052	0.5530	0.3380	0.6713
		N3	0.2526	0.3380	0.5530	0.5530

RBF balanced	48.61%	N1 N3	0.2148 0.4899	0.2963 0.6759	0.6759 0.2963	0.6465 0.6577

Table 13 shows the confusion matrix for the predicted classes with unbalanced N1 and N3 using the Gaussian kernel, confusion matrix for balanced classes can be seen in Table 14. As it can be seen, with unbalanced classes N1 has more true positives than N3 due to it has more pixels to create the model than N3, but when the classes are balanced, N3 has larger number of true positives than N1. Table 15 and Table 16, with unbalanced classes, the prediction for N3 is zero which means that Linear kernel is not appropriate for the unbalanced classification. Table 17 and Table 18 show the confusion matrix for RBF kernel, with poorest results than Gaussian but not anomalous results such as Linear.

Table 13: Confusion Matrix from Gaussian Kernel with N1 and N3 classes

CLASS	**Predicted N1**	**Predicted N3**
N1	139	125
N3	154	62

Table 14: Confusion Matrix from Gaussian Kernel with N1 and N3 balanced classes

CLASS	**Predicted N1**	**Predicted N3**
N1	100	116
N3	87	129

Table 15: Confusion Matrix from Linear Kernel with N1 and N3

CLASS	**Predicted N1**	**Predicted N3**
N1	264	0
N3	216	0

Table 16: Confusion Matrix from Lineal Kernel with N1 and N3 balanced classes

CLASS	Predicted N1	Predicted N3
N1	90	126
N3	105	111

Table 17: Confusion Matrix from RBF Kernel with N1 and N3

CLASS	Predicted N1	Predicted N3
N1	146	118
N3	143	73

Table 18: Confusion Matrix from RBF Kernel with N1 and N3 balanced classes

CLASS	Predicted N1	Predicted N3
N1	64	152
N3	70	146

5.1. Two classes

Table 19 shows the classification results of all possible pairwise combinations (biclasses), both with unbalanced and balanced samples. As it can be seen, the TAB group provides the highest accuracy to the classifier since the samples come from another island. But with HOG, N1 and N3 groups may be inaccurate due to the lack of samples in the last two to generate the model. As can be seen in Table 11, these two classes have only two and three samples for model creation. Furthermore, these samples are small, providing only 874 or 576 pixels for model generation compared to HOG and TAB classes with 12540 or 8818 pixels respectively.

Table 19: Results in groups of two classes

Group	Accuracy	Class	Precision	Sensitivity	Specificity	F1 Score
HOG-TAB	100%	HOG TAB	1 1	1 1	1 1	1 1
HOG-TAB balanced	100%	HOG TAB	1 1	1 1	1 1	1 1
HOG-N1	95.49%	HOG N1	0.9549 0	1 0	0 1	0.9769 NaN
HOG-N1 balanced	52.27%	HOG N1	0.5146 0.1585	0.7992 0.2462	0.2462 0.7992	0.6795 0.7104
HOG-N3	96.28%	HOG N3	0.9628 0	1 0	0 1	0.9810 NaN
HOG-N3 balanced	72.92%	HOG N3	0.6645 0.3821	0.9259 0.5324	0.5324 0.9259	0.7984 0.9350
TAB-N1	100%	TAB N1	1 1	1 1	1 1	1 1
TAB-N1 balanced	100%	TAB N1	1 1	1 1	1 1	1 1
TAB-N3	100%	TAB N3	1 1	1 1	1 1	1 1
TAB-N3 balanced	100%	TAB N3	1 1	1 1	1 1	1 1
N1-N3	41.87%	N1 N3	0.4744 0.2116	0.5265 0.2870	0.2870 0.5265	0.6435 0.4980
N1-N3 balanced	53.01%	N1 N3	0.4082 0.5265	0.4630 0.5972	0.5972 0.4630	0.6969 0.6898

Table 20, Table 21, Table 22 and Table 23 show the confusion matrixes of HOG-N1 and HOG-N3 groups unbalanced and balanced respectively. Confusion matrixes for N1-N3 groups are in the previous section and the HOG-TAB, TAB-N1 and TAB-N3 confusion matrixes are shown in ANNEX I in Table 49 ,Table 50,

Table 51, Table 52, Table 53 and Table 54. Where it can be observed that each class is correctly identified.

Table 20: Confusion Matrix with HOG and N1

CLASS	Predicted HOG	Predicted N1
HOG	5585	0
N1	264	0

Table 21: Confusion Matrix with HOG and N1 balanced classes

CLASS	Predicted HOG	Predicted N1
HOG	211	53
N1	199	65

Table 22: Confusion Matrix with HOG and N3

CLASS	Predicted HOG	Predicted N3
HOG	5585	0
N3	216	0

Table 23: Confusion Matrix with HOG and N3 balanced classes

CLASS	Predicted HOG	Predicted N3
HOG	200	16
N3	101	115

5.2. Three classes

The set of trios that can be generated with the four available classes can be seen in Table 24 with the corresponding metrics. There are groups of three unbalanced and balanced classes. Like in the previous section, there are a lot of differences when classifying inter or intra island and the number of samples available for the creation of the model.

Table 24: Results in groups of three classes

Group	Class	Accuracy	Precision	Sensitivity	Specificity	F1 Score
HOG-TAB-N1	HOG TAB N1	97.81%	0.9676 0.3912 0	1 1 0	0.8966 1 1	0.9769 1 NaN
HOG-TAB-N1 balanced	HOG TAB N1	86.88%	0.7032 1 0.0214	0.7032 1 0.1856	0.9158 1 0.8516	0.9374 1 0.1261
HOG-TAB-N3	HOG TAB N3	98.20%	0.9628 0.3944 0	1 1 0	0.9137 1 1	0.9865 1 NaN
HOG-TAB-N3 balanced	HOG TAB N3	86.41%	0.6922 1 0.0172	0.6862 0.9913 0.1806	0.9276 1 0.8393	0.9458 1 0.0959
HOG-N1-N3	HOG N1 N3	94.58%	0.9209 0 0	1 0 0	0 1 1	0.9588 NaN NaN
HOG-N1-N3 balanced	HOG N1 N3	30.43%	0.2258 0.0393 0.0537	0.4659 0.0720 0.1204	0.1108 0.6963 0.4863	0.4053 0.3689 0.2574
TAB-N1-N3	TAB N1 N3	94.07%	0.9996 0.0826 0.0227	1 0.7159 0.2407	0.9959 0.9349 0.9706	0.9998 0.6987 0.5810
TAB-N1-N3 balanced	TAB N1 N3	74.77%	0.7021 0.1569 0.4548	1 0.2235 0.7917	1 0.9062 0.6117	1 0.7239 0.6252

Table 25, Table 26, Table 27 and Table 28 show the confusion matrixes of unbalanced classes for all groups since they are the ones that show the most anomalous results, the classes N1 and N3 are completely misclassified due to the difference of amount of labelled data per class. Confusion matrixes for the

balanced groups can be seen at ANNEX I (Table 55, Table 56, Table 57 and Table 58).

Table 25: Confusion Matrix with HOG, TAB and N1

CLASS	Predicted HOG	Predicted TAB	Predicted N1
HOG	5585	0	0
TAB	0	2288	0
N1	264	0	0

Table 26: Confusion Matrix with HOG, TAB and N3

CLASS	Predicted HOG	Predicted TAB	Predicted N3
HOG	5585	0	0
TAB	0	2288	0
N3	216	0	0

Table 27: Confusion Matrix with HOG, N1 and N3

CLASS	Predicted HOG	Predicted N1	Predicted N3
HOG	5585	0	0
N1	264	0	0
N3	216	0	0

Table 28: Confusion Matrix with TAB, N1 and N3

CLASS	Predicted TAB	Predicted N1	Predicted N3
TAB	2288	0	0
N1	0	189	75
N3	1	163	52

5.3. Four classes

The results of the classifier for the four classes (HOG, TAB, N1 and N3) are shown in Table 29, as it can be seen the results are quite different from balanced to unbalanced classes, since the classifier works perfectly with islands

classification when there are a enough samples for the creation of the model, but when the classes are balanced, the accurate decreases. Also, in classes with not enough samples for the generation of the sample, if these are unbalanced, the models for them are not accurate and all the pixels are labelled as the closest major class, this situation is shown at Table 30.

Table 29: Result of the four-class classifier

Balanced	Accuracy	Class	Precision	Sensitivity	Specificity	F1 Score
No	97,04%	HOG	0.9207	1	0.8262	0.9587
		TAB	0.3770	0.9996	1	1
		N1	0	0	1	NaN
		N3	0	0	1	NaN
Yes	66,67%	HOG	0.3176	0.6250	0.5060	0.4821
		TAB	0.5082	1	1	1
		N1	0.0871	0.1713	0.9144	0.6667
		N3	0.1035	0.2037	0.7870	0.4560

Table 30: Confusion Matrix of the four-class classifier with unbalanced classes

CLASS	Predicted HOG	Predicted TAB	Predicted N1	Predicted N3
HOG	5585	0	0	0
TAB	1	2287	0	0
N1	264	0	0	0
N3	216	0	0	0

5.4. Jeffries-Matusita Distance

The results with the JM distance do not give off any clear idea, being the separable classes but not agreeing with previous results, that the separability between islands using the SVM was maximal. Table 31 shows the results of each pair of classes.

Table 31: Jeffries-Matusita distance of each pair of classes

Class 1	Class 2	Score
HOG	TAB	1.0023
HOG	N1	1.0035
HOG	N3	1.0041
TAB	N1	1.0106
TAB	N3	1.0116
N1	N3	1.0006

5.5. K-means

In the case of unsupervised K-means clustering, poor success results are obtained with respect to the creation of clusters based on pairs of classes. Table 32 shows the results of the rand index of the clustering of each pair of classes.

Table 32: Accuracy of K-means clustering of each pair of classes

Class 1	Class 2	Rand Index
HOG	TAB	28.46%
HOG	N1	3.56%
HOG	N3	2.41%
TAB	N1	5.53%
TAB	N3	3.79%
N1	N3	26.25%

5.6. Undefined samples prediction

For the prediction of undefined samples, the balanced four-class model is used, since using the unbalanced model could cause samples belonging to N1

or N3 to be considered as HOG. Table 33 shows the prediction of each sample, as it can be seen, there are samples with high percentage of prediction in HOG group such as ANT-156, DUM-79, RES-10-191, HOG-38-1575-69 and NE-820-1.

Table 33: Prediction of undefined samples with four classes model

Sample	HOG	TAB	N1	N3
ANT-156	82.38%	0%	9.06%	8.56%
ANT-157	21.61%	0%	33.24%	45.15%
DUM-93	22.42%	0%	33.41%	44.17%
DUM-88-2	10.60%	0%	47.35%	42.05%
DUM-79	83.94%	0%	6.06%	10.00%
RES-7-179	9.16%	0%	43.63%	47.21%
RES-10-190	56.55%	0%	17.27%	26.18%
RES-10-191	86.02%	0%	4.03%	9.95%
RES-10-194	27.62%	0%	22.10%	50.28%
HOG-38-1575-69	74.90%	0%	9.27%	15.83%
NE-820-1	76.99%	0.44%	7.52%	15.04%
NE-820-1-2	49.46%	0%	9.68%	40.86%
NE-811-2	74.75%	0%	6.57%	18.69%
NE-811-4	45.40%	0%	23.93%	30.67%

To avoid the HOG confusion due to the number of samples used, a three-class model has been used with TAB, N1 and N3 classes. Table 34 shows the results of the prediction with this three-class model. In this case, some similarity between N3 and HOG is appreciable, as the samples predicted as HOG using the four-class model are predicted almost N3 with three-class model.

Table 34: Prediction of undefined samples with TAB-N1-N3 balanced classes model

Sample	TAB	N1	N3
ANT-156	0%	20.13%	79.87%
ANT-157	0%	39.61%	60.39%
DUM-93	0%	43.05%	56.95%
DUM-88-2	0%	53.71%	46.29%
DUM-79	0%	25.15%	74.85%
RES-7-179	0%	65.14%	34.86%
RES-10-190	0%	32.87%	67.13%
RES-10-191	0%	24.19%	75.81%
RES-10-194	0%	57.18%	42.82%
HOG-38-1575-69	0%	19.50%	80.50%
NE-820-1	0%	28.32%	71.67%
NE-820-1-2	0%	30.11%	69.89%
NE-811-2	0%	54.60%	45.40%
NE-811-4	1.01%	30.30%	68.69%

6. CONCLUSION AND FUTURE LINES

The objective of this bachelor thesis was to design a supervised classification system with hyperspectral images using the classification by lava flows proposed by the Canarian Museum using chemical analysis systems. For this purpose, the provided database was employed, an automatic segmentation of the template was performed to facilitate the extraction of the spectral signature of the obsidian and finally the SVM classifier.

6.1. Conclusions

After the results obtained in chapter 5, it is shown that good results are obtained when identifying the different samples using an SVM classifier. In Table 12 the comparison of different kernels is performed, where the Gaussian kernel obtains better results, especially in specificity, also using balanced classes has better results in all metrics than other kernels. This points the way to which kernel should be employed for obsidian classification by HSI.

Regarding the classification in pairs of classes there is a clear conclusion, the chemical differences that present the obsidians with origin from different islands makes their classification much easier. As can be seen in Table 19, all the groups with the TAB class, coming from Tenerife, present maximum values of detection, so an error-free classification of obsidian by islands could be guaranteed, even with few samples as can be seen in the balanced groups TAB-N1 and TAB-N3 groups.

However, with respect to the classification within the same island, the HOG, N1 and N3 classes present a difficulty when it comes to be accurately classified, and only the balanced HOG-N3 pair presents positive results, achieving 72 % of

accuracy, which may suggest that they are pairs with more chemical differences than between HOG and N1.

With respect to classifications in groups of three, the conclusions are similar, the differences presented in samples of different islands give very positive results, as can be seen in Table 24. The balanced HOG-TAB-N1, HOG-TAB-N3 and TAB-N1-N3 groups present results higher than 74 %, even though they are classified with a low number of pixels.

Finally, the classification results with the four available classes yield relatively poor results, although the metrics of the unbalanced classifier seem very positive, what happens is that the samples of the N1 and N3 groups are classified as HOG but as these are very few and with few pixels. This situation can be seen in Table 18, the confusion matrix shows that the pixels of the N1 and N3 classes were predicted as HOG. Also, the model created with balanced classes provides only 66% of accuracy, with precision less than 50%. This situation gives the idea that the model generated is not accurate due to there were less samples than needed from N1 and N3 classes what is limiting the classification.

6.2. Future lines

This bachelor thesis demonstrates that further work can be done in this area. First, it is validating the use of SVM classifier for HSI as a non-destructive classification method to replace destructive methods such as those currently employed by El Museo Canario.

Then, it is necessary to increase the database of all samples available from El Museo Canario in order to generate new models more accurate. This recent work provides scripts to easily obtain new classes from the HSI using the same template that the recent database.

Finally, SVM presents good result, but another future line could be use Artificial Neural Networks (ANNs) as they present relevant results in other fields. ANNs can be compare with the recent results to decide future classifications.

7. BADGET

This chapter estimates the project budget based on the recommendations and guidelines established by the Colegio Oficial de Ingenieros Técnicos de Telecomunicación (COITT) established in 2008. In turn, this budget is divided into the following sections:

7.1. Materials resources

The material resources used for this final work includes hardware and software resources, which in turn may have costs associated with the requirement of licenses for their use. To estimate the amortization cost, a period of 4 years is stipulated, assuming a linear amortization system, in which fixed assets are disregarded constantly during the evaluation period. The following equation is used for this calculation:

$$Amortization\ Cost = \frac{Adquisition\ Value - Residual\ Value}{Useful\ lifetime\ in\ years}$$

As the present project has a duration of 300 hours distributed approximately in 12 weeks, less than 4 years stipulated, this cost will be the one derived from the 12 weeks in which the project is developed.

7.1.1. Hardware resources:

Table 35: Hardware resources total cost

Resource	Acquisition Value	Residual Value	Amortization Cost	Amount
MacBook Air	€1,129	0	€282.25	€70.56
			Total:	**€70.56**

7.1.2. Software resources:

Table 36: Software resources total cost

Resource	Acquisition Value	Residual Value	Amortization Cost	Amount
MATLAB 2022a Education	€250	0	€250	€62.50
Microsoft Office 365 Personal	€69	0	€69	€17.25
			Total:	**€79.75**

7.2. Human Resources

Human resource of this work is the engineer involved in the development of the project, and the costs are based on the hours of work that have been used in its execution. The amount of these working hours is calculated according to the recommendations of the Colegio Oficial de Ingenieros Técnicos de Telecomunicación (COITT) using the following equation.

$$Fees = H_n * 14.48 + H_e * 20.27$$

Where H_n are the hours in working day and H_e are extra hours out of working hours.

For this work 300 hours were needed, done in working days and no extra hours needed, so the final cost is:

$$Fees = 300 * 14.48 + 0 * 20.27 = €4{,}344.00$$

The total fees for time spent, tax free, is *four thousand three hundred and forty-four euros*.

7.3. Drafting of the document

In accordance with the COITT's guidelines, the amount for the writing of this final work is calculated using the following equation:

$$R = 0.005 * P * C_n$$

Where *P* is the budget value as the sum of all resources and C_n is the weighting coefficient as a budget cost function. On the other hand, for this project the weighting coefficient C_n has a value of unity because the total cost of the TFG does not exceed €30,050.00. Therefore:

$$R = 0.005 * (70.56 + 79.75 + 4344.00) * 1 = €224.72$$

The total cost for drafting the document is *two hundred and twenty-four euros* and seventy-two cents.

7.4. COITT visa fees

For general projects, the COITT visa fees (year 2021) are charged using the following expression:

$$R = 0.006 * P_1 * C_1 + 0.003 * P_2 * C_2$$

Where P_1 is the general budget of the project, P_2 is the material execution budget corresponding to the civil works, C_1 is the reduction coefficient corresponding to P_1 and C_2 is the reduction coefficient corresponding to P_2. There is a minimum of €40. Therefore:

$$R = 0.006 * 4719.03 * 1 = €28.31 \rightarrow €40$$

The total cost of drafting the document is *forty euros*.

7.5. Processing and shipping costs

For general documents endorsed by telematic means, the cost is six euros and one cent (**€6.01**).

7.6. Final budget

The final budget with all the items listed in Table 37 amounts to *four thousand seven hundred sixty-five euros and four cents* (€4,765.04). To this amount is added the Canary Islands General Indirect Tax (IGIC, equivalent to 7%), obtaining the total cost of the work presented.

Table 37: Final Budget

Final Budget	
Items	**Amounts**
Material resources	
- Hardware	€70.56
- Software	€79.75
Human resources	€4,344
Subtotal	**€4,494.31**
Document Drafting	€224.72
COITT visa fees	€40
Processing and shipping cost	€6.01
Total	**€4,765.04**
IGIC	€333.55
TOTAL + TAX:	**€5,098.59**

Finally, the budget of this work is *five thousand ninety-eight euros and fifty-nine cents* (**€5,098.59**).

8. ANNEX I

Table 38: TAB group defined by El Museo Canario

TAB			
CHA-1	CHA-13-b	CHA-27	CHA-39
CHA-2	CHA-14	CHA-28	CHA-40
CHA-3	CHA-16	CHA-29	CHA-41
CHA-4	CHA-18	CHA-30	CHA-42
CHA-5	CHA-19	CHA-31	CPG-12-146
CHA-6	CHA-20	CHA-32	TAB-1
CHA-7	CHA-21	CHA-33	TAB-2
CHA-8	CHA-22	CHA-34	TAB-3
CHA-9	CHA-23	CHA-35*	TAB-OBS059
CHA-10	CHA-24	CHA-36	TAB-OBS060
CHA-12-a	CHA-25	CHA-37	TAB-OBS067
CHA-13	CHA-26	CHA-38	

* Two captured samples are called CHA-35.

Table 39: N1 group defined by El Museo Canario

N1			
MEL-33	CPG-0-151	DUM-89	RES-7-177
MEL-47-B	CPG-25-134	RES-7-173	RES-7-178
MEL-48	DUM-77	RES-7-176	

Table 40: N2 group defined by El Museo Canario

N2			
CPG-43-213	LLA-26	PAJ-57	PAJ-62-A
CPG-43-216	LLA-27	PAJ-59-G	PAJ-63
LLA-19	LLA-31	PAJ-60-C	TJR-109

Table 41: N3 group defined by El Museo Canario

N3			
BAR-11	LLA-21	PAJ-58-D	PAJ-62-B
BAR-6	LLA-22	PAJ-58-E	PAJ-62-C
BAR-9	LLA-23	PAJ-59-B	RES-10-186

CNB-149-B	LLA-25-A	PAJ-59-C	RES-10-190-II
CPG-43-208	LLA-25-B	PAJ-59-F-IV	TJR-104
CPG-43-210	LLA-28	PAJ-59-H	TJR-112
DUM-81	LLA-32	PAJ-60-A	TJR-99
DUM-92	MEL-50	PAJ-60-B	
LLA-20	PAJ-58-B	PAJ-61	

Table 42: HOG group defined by El Museo Canario

HOG			
ANT-158	CPG-0-147	CPG-43-221	MEL-56
BAR-1	CPG-0-148	CPG-43-222	RES-10-180
BAR-10	CPG-0-149	CPG-43-223	RES-10-181
BAR-13	CPG-0-150	DUM-78	RES-10-183
BAR-2-A	CPG-12-140	DUM-80	RES-10-184
BAR-2-B	CPG-12-141	DUM-82	RES-10-185
BAR-4	CPG-12-142	DUM-83	RES-10-187
BAR-5	CPG-12-143	DUM-85	RES-10-189
BAR-7	CPG-12-144	DUM-88-1	RES-10-190-I
BOG-15	CPG-12-145	DUM-90	RES-10-192
BOG-16	CPG-25-133	DUM-91	RES-10-193
BOJ-17	CPG-25-135	HOG-38-1368-73	RES-7-174
CED-18-114	CPG-25-136	HOG-38-816-65	RES-7-175
CED-5-118-I	CPG-43-205	HOG-38-818-65	TJR-100
CED-5-118-II	CPG-43-206	HOG-OBS001	TJR-102
CED-5-118-III	CPG-43-207	HOG-OBS009	TJR-105
CED-C-155	CPG-43-209	HOG-OBS013	TJR-107-II
CED-T-113	CPG-43-211	LLA-24	TJR-111
CNB-147-A	CPG-43-212	MEL-36	TJR-94
CNB-147-B	CPG-43-214	MEL-37	TJR-97
CNB-149-A	CPG-43-215	MEL-38	TJR-98
CNB-151	CPG-43-217	MEL-40	VAC-1-119
CNB-152	CPG-43-218	MEL-42	VAC-2-120-A
CNB-153	CPG-43-219	MEL-47-A	VAC-2-120-B
CNB-155	CPG-43-220	MEL-53	VAC-2-120-C

Table 43: Undefined samples by El Museo Canario

Undefined			
ANT-157	DUM-88-2	PAJ-58-A	RCH-OBS024
BAR-8	DUM-93	PAJ-58-C	RES-7-179
CHA-17	FOR-OBS029	PAJ-58-F	RES-10-194
CPG-12-138	FOR-OBS031	PAJ-58-G	TJR-96
CPG-12-139	LLA-30	PAJ-59-A	TJR-106
DUM-79	MEL-51	PAJ-59-F-II	

Table 44: Samples selected for train and test subclasses of HOG set

HOG		
TRAIN		
RES10-180 (TOMA1)	RES10-181 (TOMA1)	RES10-183 (TOMA1)
RES10-184 (TOMA1)	RES10-187 (TOMA1)	RES7-174 (TOMA1)
RES7-175 (TOMA1)	CED-18-114 (TOMA2)	CED-C-115 (TOMA2)
HOG-38-818-65 (TOMA2)	HOG-38-816-65 (TOMA2)	ANT-158 (TOMA2)
VAC-2-120-B (TOMA2)	VAC-1-119 (TOMA3)	CNB-151(TOMA3)
CNB-149-A (TOMA3)	CNB-147-A (TOMA3)	CNB-147-B (TOMA3)
CNB-152 (TOMA3)	DUM-78 (TOMA4)	DUM-80 (TOMA4)
DUM-82 (TOMA4)	DUM-83 (TOMA4)	DUM-85 (TOMA4)
TEST		
RES10-185 (TOMA1)	RES10-192 (TOMA1)	RES10-193 (TOMA1)
CED-T-113 (TOMA2)	HOG-38-1368-73 (TOMA2)	VAC-2-120-C (TOMA2)
VAC-2-120-A (TOMA3)	CNB-153 (TOMA3)	CNB-155 (TOMA3)
DUM-90 (TOMA4)	DUM-91 (TOMA4)	

Table 45: Samples selected for train and test subclasses of TAB set

TAB		
TRAIN		
CHA-33 (TOMA5)	CHA-28 (TOMA5)	CHA-36 (TOMA5)
CHA-39 (TOMA5)	CHA-31 (TOMA5)	CHA-27 (TOMA5)
TAB-1(1) (TOMA5)	TAB-1(3) (TOMA5)	
TEST		
CHA-30 (TOMA5)	CHA-35 (TOMA5) * 2	TAB-1(2) (TOMA5)

Table 46: Samples selected for train and test subclasses of N1 set

N1		
TRAIN		
RES7-178 (TOMA1)	DUM-77 (TOMA4)	DUM-89 (TOMA4)
TEST		
RES7-173 (TOMA1)		

Table 47: Samples selected for train and test subclasses of N3 set

N3		
TRAIN		
CNB-149-B (TOMA3)	DUM-81 (TOMA4)	
TEST		
DUM-92 (TOMA4)		

Table 48: Rest of the captured samples

Undefined		
RES10-194 (TOMA1)	RES10-190* (TOMA1)	RES10-191* (TOMA1)
RES-7-179 (TOMA2)	ANT-156** (TOMA2)	ANT-157 (TOMA2)
HOG-38-1575-69** (TOMA2)	NO ETIQ (820-1) (TOMA 3)	NO ETIQ (820-1) (TOMA 3)

NO ETIQ (811-4) (TOMA 3)	NO ETIQ (811-2) (TOMA 3)	DUM-79 (TOMA4)
DUM-88-2 (TOMA4)	DUM-93 (TOMA4)	

* Samples that need to be confirmed as HOG or N3

** Samples that are not in the list

Table 49: Confusion Matrix with HOG and TAB

CLASS	**Predicted HOG**	**Predicted TAB**
HOG	5585	0
TAB	0	2288

Table 50: Confusion Matrix with HOG and TAB balanced classes

CLASS	**Predicted HOG**	**Predicted TAB**
HOG	2288	0
TAB	0	2288

Table 51: Confusion Matrix with TAB and N1

CLASS	**Predicted TAB**	**Predicted N1**
TAB	2288	0
N1	0	264

Table 52: Confusion Matrix with TAB and N1 balanced classes

CLASS	**Predicted TAB**	**Predicted N1**
TAB	264	0
N1	0	264

Table 53: Confusion Matrix with TAB and N3

CLASS	**Predicted TAB**	**Predicted N3**
TAB	2288	0
N3	0	216

Table 54: Confusion Matrix with TAB and N3 balanced classes

CLASS	**Predicted TAB**	**Predicted N3**
TAB	216	0
N3	0	216

Table 55: Confusion Matrix with HOG, TAB and N1 balanced classes

CLASS	Predicted HOG	Predicted TAB	Predicted N1
HOG	1609	0	679
TAB	0	2288	0
N1	215	0	49

Table 56: Confusion Matrix with HOG, TAB and N3 balanced classes

CLASS	Predicted HOG	Predicted TAB	Predicted N3
HOG	1570	0	718
TAB	3	2268	17
N3	177	0	39

Table 57: Confusion Matrix with HOG, N1 and N3 balanced classes

CLASS	Predicted HOG	Predicted N1	Predicted N3
HOG	123	40	101
N1	196	19	49
N3	165	25	26

Table 58: Confusion Matrix with TAB, N1 and N3 balanced classes

CLASS	Predicted TAB	Predicted N1	Predicted N3
TAB	264	0	0
N1	0	59	205
N3	0	45	171

9. ANNEX II

The source codes related to this final work are available through the supervisors.

10. BIBLOGRAPHY

[1] J. M. H. Rodríguez, "Analyzing Canarian Archeology Heritage through Hyperspectral Image Analysis." 2019.

[2] K. F. Khan, "Application, principle and operation of ICP-OES in pharmaceutical analysis," vol. 8, pp. 281–282, Oct. 2019.

[3] S. Feroz, "Induced Couple Plasma - Optical Emission Spectroscopy (ICP-OES)." Oct. 2020.

[4] T. Alphiya *et al.*, "Kidney stones analysis by ICP-OES," *Journal of Physics: Conference Series*, vol. 1611, p. 12055, Oct. 2020, doi: 10.1088/1742-6596/1611/1/012055.

[5] J. Jagodić *et al.*, "Elemental profiling of adrenal adenomas in solid tissue and blood samples by ICP-MS and ICP-OES," *Microchemical Journal*, vol. 165, p. 106194, 2021, doi: https://doi.org/10.1016/j.microc.2021.106194.

[6] O. Y. Song *et al.*, "Elemental composition of pork meat from conventional and animal welfare farms by inductively coupled plasma-optical emission spectrometry (ICP-OES) and ICP-mass spectrometry (ICP-MS) and their authentication via multivariate chemometric analysis," *Meat Science*, vol. 172, p. 108344, 2021, doi: https://doi.org/10.1016/j.meatsci.2020.108344.

[7] H. Liu, Q. Meng, X. Zhao, Y. Ye, and H. Tong, "Inductively coupled plasma mass spectrometry (ICP-MS) and inductively coupled plasma optical emission spectrometer (ICP-OES)-based discrimination for the authentication of tea," *Food Control*, vol. 123, p. 107735, 2021, doi: https://doi.org/10.1016/j.foodcont.2020.107735.

[8] G. Habte *et al.*, "Elemental profiling and geographical differentiation of Ethiopian coffee samples through inductively coupled plasma-optical emission spectroscopy (ICP-OES), ICP-mass spectrometry

(ICP-MS) and direct mercury analyzer (DMA)," *Food Chemistry*, vol. 212, pp. 512–520, 2016, doi: https://doi.org/10.1016/j.foodchem.2016.05.178.

[9] I. Sharma, "ICP-OES: An Advance Tool in Biological Research," *Open Journal of Environmental Biology*, pp. 27–33, Oct. 2020, doi: 10.17352/ojeb.000018.

[10] M. Khushaim, "Investigation of the Precipitation Behavior in Aluminum Based Alloys," 2015. doi: 10.13140/RG.2.1.2987.3521.

[11] N. Abood, "The 30-Minute Guide to ICP-MS." Oct. 2020.

[12] F. Iaquinta, L. Fialho, J. Nóbrega, M. Pistón, and I. Machado, "Determination of Cd, Pb and Se in beef samples using aerosol dilution by ICP-MS," *Journal of Food Measurement and Characterization*, vol. 15, Oct. 2021, doi: 10.1007/s11694-021-00999-3.

[13] Y. Yu, M. Zhao, S. Li, Z. Li, and Y. Zhu, "Simulaneous Determination of Heavey Metals in 6 Kind of Mineral Water by ICP-MS," *IOP Conference Series: Earth and Environmental Science*, vol. 814, p. 12009, Oct. 2021, doi: 10.1088/1755-1315/814/1/012009.

[14] I. Hwang *et al.*, "Determination of Toxic Elements and Arsenic Species in Salted Foods and Sea Salt by ICP–MS and HPLC–ICP–MS," *ACS Omega*, vol. XXXX, Oct. 2021, doi: 10.1021/acsomega.1c01273.

[15] W. Al-Onazi, A. Al-Mohaimeed, M. Amina, and M. El-Tohamy, "Identification of Chemical Composition and Metal Determination of Retama raetam (Forssk) Stem Constituents Using ICP-MS, GC-MS-MS, and DART-MS," *Journal of Analytical Methods in Chemistry*, vol. 2021, pp. 1–9, Oct. 2021, doi: 10.1155/2021/6667238.

[16] W. Li, N. Niu, N. Guo, H. Zhou, J. Bu, and A. Ding, "Comparative Study on the Determination of heavy metals in Soil by XRF and ICP-MS,"

Journal of Physics: Conference Series, vol. 2009, p. 12075, Oct. 2021, doi: 10.1088/1742-6596/2009/1/012075.

[17] I. Liritzis and N. Zacharias, "X-Ray Fluorescence Spectrometry (XRF) in Geoarchaeology," 2011, pp. 109–142. doi: 10.1007/978-1-4419-6886-9_6.

[18] National Aeronautics and Space Administration, "The Electromagnetic Spectrum," *https://imagine.gsfc.nasa.gov/science/toolbox/emspectrum1.html*, Mar. 2013.

[19] R. Leon *et al.*, "VNIR–NIR hyperspectral imaging fusion targeting intraoperative brain cancer detection," *Scientific Reports*, vol. 11, p. 19696, Dec. 2021, doi: 10.1038/s41598-021-99220-0.

[20] B. Regeling *et al.*, "Hyperspectral Imaging Using Flexible Endoscopy for Laryngeal Cancer Detection," *Sensors*, vol. 16, p. 1288, Dec. 2016, doi: 10.3390/s16081288.

[21] B. Regeling *et al.*, "Development of an image pre-processor for operational hyperspectral laryngeal cancer detection," *Journal of Biophotonics*, vol. 9, Dec. 2015, doi: 10.1002/jbio.201500151.

[22] H. Fabelo *et al.*, "Spatio-spectral classification of hyperspectral images for brain cancer detection during surgical operations," *PLOS ONE*, vol. 13, p. e0193721, Dec. 2018, doi: 10.1371/journal.pone.0193721.

[23] D. Ravì, H. Fabelo, G. Marrero Callico, and G. Yang, "Manifold Embedding and Semantic Segmentation for Intraoperative Guidance With Hyperspectral Brain Imaging," *IEEE Transactions on Medical Imaging*, vol. PP, p. 1, Dec. 2017, doi: 10.1109/TMI.2017.2695523.

[24] S. Ortega, H. Fabelo, R. Camacho, M. Plaza, G. Marrero Callico, and R. Sarmiento, "Detecting brain tumor in pathological slides using

hyperspectral imaging," *Biomedical Optics Express*, vol. 9, p. 818, Jul. 2018, doi: 10.1364/BOE.9.000818.

[25] H. Fabelo *et al.*, "A Novel Use of Hyperspectral Images for Human Brain Cancer Detection using in-Vivo Samples," Jul. 2016, pp. 311–320. doi: 10.5220/0005849803110320.

[26] D. Wu and D.-W. Sun, "Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review - Part II: Applications," *Innovative Food Science & Emerging Technologies*, vol. 19, pp. 15–28, Jul. 2013, doi: 10.1016/j.ifset.2013.04.016.

[27] R. Lu, "Quality Evaluation of Fruit by Hyperspectral Imaging," *Computer Vision Technology for Food Quality Evaluation*, Jul. 2008, doi: 10.1016/B978-012373642-0.50017-X.

[28] P. Menesatti, C. Costa, and J. Aguzzi, "Quality Evaluation of Fish by Hyperspectral Imaging," in *Hyperspectral Imaging for Food Quality Analysis and Control*, 2010, pp. 273–294. doi: 10.1016/B978-0-12-374753-2.10008-5.

[29] P. Xiao, Q. Zhu, M. Huang, and K. Yin, "Hyperspectral image detection of pork freshness based on Active Learning," *American Society of Agricultural and Biological Engineers Annual International Meeting 2014, ASABE 2014*, vol. 3, pp. 2277–2285, Dec. 2014.

[30] R. M. Wentworth, J. Neiss, M. Nelson, and P. Treado, "Standoff Raman hyperspectral imaging detection of explosives," Dec. 2007, pp. 4925–4928.

[31] Y. Liu, L. Qunbo, J. Wang, L. Pei, and W. Li, "UAV-based hyperspectral imaging detection for explosives and contaminants," Dec. 2019, p. 58. doi: 10.1117/12.2532716.

[32] J. Luypaert, D. L. Massart, and Y. Heyden, "Near-infrared spectroscopy applications in pharmaceutical analysis," *Talanta*, vol. 72, pp. 865–883, Jul. 2007, doi: 10.1016/j.talanta.2006.12.023.

[33] Y. Chen, N. van Berkel, C. Luo, Z. Sarsenbayeva, and V. Kostakos, "Application of miniaturized near-infrared spectroscopy in pharmaceutical identification," *Smart Health*, vol. 18, p. 100126, Jul. 2020, doi: 10.1016/j.smhl.2020.100126.

[34] I. Paris *et al.*, "Near infrared spectroscopy and process analytical technology to master the process of busulfan paediatric capsules in a university hospital," *J Pharm Biomed Anal*, vol. 41, pp. 1171–1178, Jul. 2006, doi: 10.1016/j.jpba.2006.02.049.

[35] C. Daffara, E. Pampaloni, L. Pezzati, and M. Barucci, "Scanning Multispectral IR Reflectography SMIRR: An Advanced Tool for Art Diagnostics," *Acc Chem Res*, vol. 43, pp. 847–856, Jul. 2010, doi: 10.1021/ar900268t.

[36] M. Kubik, "Chapter 5 Hyperspectral Imaging: A New Technique for the Non-Invasive Study of Artworks," *Physical Techniques in the Study of Art, Archaeology and Cultural Heritage*, vol. 2, pp. 199–259, Jul. 2007, doi: 10.1016/S1871-1731(07)80007-8.

[37] L. Macdonald *et al.*, "Assessment of multispectral and hyperspectral imaging systems for digitisation of a Russian icon," *Heritage Science*, vol. 5, Jul. 2017, doi: 10.1186/s40494-017-0154-1.

[38] F. Daniel *et al.*, "Hyperspectral imaging applied to the analysis of Goya paintings in the Museum of Zaragoza (Spain)," *Microchemical Journal*, vol. 126, pp. 113–120, Jul. 2015, doi: 10.1016/j.microc.2015.11.044.

[39] V. Miljkovic and D. Gajski, "ADAPTATION OF INDUSTRIAL HYPERSPECTRAL LINE SCANNER FOR ARCHAEOLOGICAL

APPLICATIONS," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLI-B5, pp. 343–345, Jul. 2016, doi: 10.5194/isprs-archives-XLI-B5-343-2016.

[40] P. Kolokoussis, M. Skamantzari, S. Tapinaki, V. Karathanassi, and A. Georgopoulos, "3D AND HYPERSPECTRAL DATA INTEGRATION FOR ASSESSING MATERIAL DEGRADATION IN MEDIEVAL MASONRY HERITAGE BUILDINGS," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLIII-B2-2021, pp. 583–590, Dec. 2021, doi: 10.5194/isprs-archives-XLIII-B2-2021-583-2021.

[41] S. Sabins and K. Lulla, "Remote sensing: Principles and interpretation," *Geocarto International*, vol. 2, no. 1, 1987, doi: 10.1080/10106048709354087.

[42] M. K. Griffin, S. May-Hsu, H. K. Burke, S. M. Orloff, and C. a. Upham, "Examples of EO-1 Hyperion data analysis," *Lincoln Laboratory Journal*, vol. 15, no. 2, 2005.

[43] N. Memarsadeghi and T. Doggett, "NASA computational case study, hyperspectral data processing: Cryospheric change detection," *Computing in Science and Engineering*, vol. 14, no. 4, 2012, doi: 10.1109/MCSE.2012.43.

[44] R. Raj, J. P. Walker, V. Vinod, R. Pingale, B. Naik, and A. Jagarlapudi, "Leaf water content estimation using top-of-canopy airborne hyperspectral data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 102, 2021, doi: 10.1016/j.jag.2021.102393.

[45] V. Vinod, R. Raj, R. Pingale, and A. Jagarlapudi, "Estimating Leaf Water Content using Remotely Sensed Hyperspectral Data." Nov. 2021.

[46] J. Liu, E. Pattey, J. R. Miller, H. McNairn, A. Smith, and B. Hu, "Estimating crop stresses, aboveground dry biomass and yield of corn using multi-temporal optical data combined with a radiation use efficiency model," *Remote Sensing of Environment*, vol. 114, no. 6, 2010, doi: 10.1016/j.rse.2010.01.004.

[47] N. Muselimyan *et al.*, "Seeing the Invisible: Revealing Atrial Ablation Lesions Using Hyperspectral Imaging Approach," *PLOS ONE*, vol. 11, p. e0167760, Dec. 2016, doi: 10.1371/journal.pone.0167760.

[48] H. Asfour *et al.*, "Comparison between Autofluorescence and Reflectance-Based Hyperspectral Imaging for Visualization of Atrial Ablation Lesions," *Biophysical Journal*, vol. 110, pp. 493a–494a, Dec. 2016, doi: 10.1016/j.bpj.2015.11.2639.

[49] C. Wang, W. Zheng, Y. Bu, S. Chang, S. Zhang, and R. Xu, "Multi-scale hyperspectral imaging of cervical neoplasia," *Arch Gynecol Obstet*, vol. 293, Dec. 2016, doi: 10.1007/s00404-015-3906-8.

[50] C. Wang *et al.*, "In vivo and in vitro hyperspectral imaging of cervical neoplasia," *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*, vol. 8951, Dec. 2014, doi: 10.1117/12.2041046.

[51] F. Manni *et al.*, "Hyperspectral Imaging for Glioblastoma Surgery: Improving Tumor Identification Using a Deep Spectral-Spatial Approach," *Sensors*, vol. 20, p. 6955, Dec. 2020, doi: 10.3390/s20236955.

[52] J. Sandak *et al.*, "Nondestructive Evaluation of Heritage Object Coatings with Four Hyperspectral Imaging Systems," *Coatings*, vol. 11, p. 244, Dec. 2021, doi: 10.3390/coatings11020244.

[53] M. Zucco, M. Pisani, and T. Cavaleri, "Fourier Transform Hyperspectral Imaging for Cultural Heritage," 2017. doi: 10.5772/66107.

[54] MathWorks, "What Is Image Segmentation?" https://es.mathworks.com/discovery/image-segmentation.html (accessed Dec. 09, 2021).

[55] MathWorks, "3-D Brain Tumor Segmentation Using Deep Learning." https://es.mathworks.com/help/deeplearning/ug/segment-3d-brain-tumor-using-deep-learning.html (accessed Dec. 09, 2021).

[56] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979, doi: 10.1109/TSMC.1979.4310076.

[57] JeongYong Hwang, "Theory5. SVM(1)," Apr. 25, 2020. https://wjddyd66.github.io/machine%20learning/Theory(5)SVM/ (accessed Dec. 09, 2021).

[58] S. Ananna, A. Miah, and M. Ahmad, "Abnormality Detection of Heart by using SVM Classifier." Dec. 2020.

[59] A. J. and J. D, "Detection of Knee Joint Disorders using SVM Classifier," *International Journal of Scientific Research in Science and Technology*, pp. 261–271, Dec. 2021, doi: 10.32628/IJSRST218535.

[60] N. Dimmita, "Detection of Lung Cancer using SVM Classifier," *International Journal of Emerging Trends in Engineering Research*, vol. 8, pp. 2177–2180, Dec. 2020, doi: 10.30534/ijeter/2020/113852020.

[61] F. Amara and F. Mohamed, "Voice Pathologies Classification Using GMM And SVM Classifiers," *International Journal of Mathematics and Computers in Simulation*, vol. 15, pp. 110–114, Dec. 2021, doi: 10.46300/9102.2021.15.21.

[62] Baeldung, "F-1 Score for Multi-Class Classification," *https://www.baeldung.com/cs/multi-class-f1-score*, Oct. 19, 2020.

[63] R. Sen, S. Goswami, and B. Chakraborty, "Jeffries-Matusita distance as a tool for feature selection," in *2019 International Conference on Data Science and Engineering (ICDSE)*, 2019, pp. 15–20. doi: 10.1109/ICDSE47409.2019.8971800.

[64] Y. Wang, Q. Qi, and Y. Liu, "Unsupervised Segmentation Evaluation Using Area-Weighted Variance and Jeffries-Matusita Distance for Remote Sensing Images," *Remote Sensing*, vol. 10, no. 8, 2018, doi: 10.3390/rs10081193.

[65] M. Dabboor, S. Howell, M. Shokr, and J. Yackel, "The Jeffries–Matusita distance for the case of complex Wishart distribution as a separability criterion for fully polarimetric SAR data," *International Journal of Remote Sensing*, vol. 35, no. 19, pp. 6859–6873, 2014, doi: 10.1080/01431161.2014.960614.

[66] P. Srinivasaperumal and S. Sanjeevi, "Jeffries Matusita based mixed-measure for improved spectral matching in hyperspectral image analysis," *International Journal of Applied Earth Observation and Geoinformation*, vol. 32, pp. 138–151, Jul. 2014, doi: 10.1016/j.jag.2014.04.001.