NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTAMENT OF ELECTROTECHNICAL AND COMPUTER ENGINEERING

EDUARDO GONÇALVES FREITAS

BSc in Electrical and Computer Engineering

RIPENING ASSESSMENT CLASSIFICATION USING ARTIFICIAL INTELLIGENCE ALGORITHMS WITH ELECTROCHEMICAL IMPEDANCE SPECTROSCOPY DATA

MASTER IN ELECTRICAL AND COMPUTER ENGINEERING

NOVA University Lisbon

March, 2023

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTAMENT OF ELECTROTECHNICAL AND COMPUTER ENGINEERING

RIPENING ASSESSMENT CLASSIFICATION USING ARTIFICIAL INTELLIGENCE ALGORITHMS WITH ELECTROCHEMICAL IMPEDANCE SPECTROSCOPY DATA

EDUARDO GONÇALVES FREITAS

BSc in Electrical and Computer Engineering

Adviser: Rui Manuel Tavares
Auxiliar Professor, NOVA University Lisbon

Co-advisers: João Pedro Carvalho
Assistant Professor, ULHT

Examination Committee:

Chair: Paulo Pimentão,
Auxiliar Professor, NOVA University Lisbon

Rapporteur: João Pedro Oliveira,
Auxiliar Professor, NOVA University Lisbon

Adviser: Rui Manuel Tavares,
Auxiliar Professor, NOVA University Lisbon

MASTER IN ELECTRICAL AND COMPUTER ENGINEERING

NOVA University Lisbon
March, 2023

RIPENING ASSESSMENT CLASSIFICATION USING ARTIFICIAL INTELLIGENCE ALGORITHMS WITH ELECTROCHEMICAL IMPEDANCE SPECTROSCOPY DATA

ACKNOWLEDGMENTS

I would like to start by thanking my adviser Prof. Dr. Rui Tavares and co-adviser Prof. Dr. João Pedro Carvalho, for all the support and guidance given through this thesis' challenges. I would like to extend my gratitude to Prof. Dr. João Oliveira and Prof. Dr. Carla Quintão, for lending materials crucial for this research. I'd like also to thank researchers João Carmo and Dr. Ana Baptista for helping me better understand and test EIS. Lastly, I'd like to thank my colleague Miguel Marques, president of NEEC, for lending me materials and providing equipment necessary for this thesis' completion. I'd like to thank to the Department of Electrotechnical and Computer Engineering (DEEC) and FCT NOVA for the education provided over the past years.

On a personal note, I'd like to start by thanking to my family and sister Rita for supporting me over the years. To AEFCT and every member, for being part of my academic life from 2018 to 2022 and helping me grow as a person. A special thanks to Sónia, Susana, Edgar, Joana, and Rita for being there throughout the many challenges and adventures we faced, as well to Carolina, Maria, João and Laura. To Ana Raquel, Miguel, Onofre, and Isabel for believing in. A thanks to my housemate Luís, to Cátia and to my friends at FAL for always asking about my bananas in every meeting, especially to Inês. A special thanks for all the patience and support to my friends Marta, Margarida, Rúben, Leonor, Maria, Francisco, Gonçalo, Joana, Inês and Rita. And of course, to my overseas friend, Colin.

"There is not enough time to do all the nothing we want to do" (Bill Watterson)

Abstract

The increasingly growth of the world's population requires more efficient efforts to manage people's needs. One important problem lies on food distribution, its quality and sustainability throughout the chain, from the producer to the consumer's hands. It's more and more important to achieve intelligent non-destructive methods to assess food quality, at the crop, storage, and shelves in the supermarket. This dissertation approaches an application of machine learning on impedance data from bananas to assess whether the fruit is edible or not. To fulfill the proposed objective, impedance data of 10 bananas was acquired during 32 days through electrochemical impedance spectroscopy (EIS) with a two ECG electrode disposition, using an Impedance Analyzer Adapter on an Analog Discovery 2 device. A database was produced with impedance, humidity and temperature values from each measurement. After data pre-processing, several machine learning classifiers were trained and tested for several different feature combinations and data normalization methods. The XGB classifier achieved the best performance, with a F1-score of 98,36% and accuracy of 98,10%.

This study can be extrapolated to other fruits and vegetables to allow a better management in the food industry, improving food quality and preventing waste.

Keywords: Food waste, Sustainability, Fruit, Banana, Electrochemical Impedance Spectroscopy, EIS, ECG electrode, Analog Discovery 2, Machine Learning

Resumo

O crescente aumento da população mundial exige esforços mais eficientes para gerir as necessidades das pessoas. Um problema importante incide na distribuição de alimentos, na sua qualidade e sustentabilidade ao longo da cadeia industrial, desde o produtor até às mãos dos consumidores. É cada vez mais importante desenvolver métodos inteligentes e não destrutivos para avaliar a qualidade dos alimentos, desde a sua produção, ao armazenamento e prateleiras nos supermercados. Esta dissertação aborda uma aplicação de aprendizagem automática em dados de impedância de bananas para avaliar se a fruta é ou não comestível. Para cumprir o objetivo proposto, foram recolhidos dados de impedância de 10 bananas durante 32 dias, através de espectroscopia de impedância eletroquímica (EIS), através de dois elétrodos ECG, utilizando um Impedance Analyzer Adapter num Analog Discovery 2. Foi criada uma base de dados com valores de impedância, humidade e temperatura de cada medição. Após o pré-processamento dos dados, vários classificadores de aprendizagem automática foram treinados e testados para várias combinações de atributos e métodos de normalização de dados diferentes. O classificador XGB obteve o melhor desempenho, com um F1-score de 98,36% e uma exatidão de 98,10%.

Este estudo pode ser extrapolado para outros frutos e vegetais, a fim de permitir uma melhor gestão na indústria alimentar, melhorando a qualidade dos alimentos e prevenindo o desperdício.

Palavras chave: Desperdício alimentar, Sustentabilidade, Fruta, Banana, Espectroscopia de impedância eletroquímica, EIS, Elétrodo ECG, Analog Discovery 2, Aprendizagem automática

CONTENTS

LIST OF FIGURES

List of Tables

ACRONYMS

EIS	Electrochemical Impedance Spectroscopy
IS	Impedance Spectroscopy
ML	Machine Learning
AI	Artificial Intelligence
SVM	Support Vector Machine
DT	Decision Tree
NN	Neural Networks
MLP	Multi-Layer Perceptron
LG	Logistic Regression
GB	Gradient Boosting
XGB	Extreme Gradient Boosting
SCC	Soluble Solid Contents
AD2	Analog Discovery 2
SUT	Sample under test

SYMBOLS

Ω	International System unit for resistance and therefore, for impedance.
μ	Unit prefix in the metric system denoting a factor of 10^{-6}
ω	Angular frequency of a signal

1

INTRODUCTION

This chapter presents an introduction to the thesis' topic, its background, motivation, objectives and contributions.

1.1 Context and Motivation

The world's population has been growing and consequently food consumption increases as well. Managing food resources is becoming a pressing matter since losses become even more prejudicial, especially in foods like fruit, which make up for a large part of a healthy and nutritious diet.

In agriculture, large scale producers take in account several factors to maximize the product's quality and thus, the business revenue. Ripeness is one of the main indicators that allow farmers to make decisions about harvesting times, and therefore allow a better productive management through the rest of the industry chain.

However, ripeness of fruits and vegetables depends on many different factors, from unexpected climacteric changes to agricultural applications (e.g. irrigation, pruning, fertilizers) (Bruns, 2009), and even the natural biological mechanisms of the fruit. It's then necessary to collect information about the fruit's conditions throughout the various stages, from the crop to the supermarkets and other destinies. Besides the need of acquiring this data *in vivo* for an optimal harvesting process, often the ripening of the product happens post-harvest, in storage (e.g. tomatoes or bananas) (Prasanna et al., 2007) and in the supermarket shelves, which makes new efficient methods that allow fruit quality assessment necessary.

1.2 Problems and Objectives

The world faces new and unpredictable challenges managing food resources. From wars and natural disasters to an increasing impact of climatic changes in agriculture and the increase of the world's population, it is necessary to find ways to assess food quality and avoid waste.

Traditional methods of fruit quality assessment are outdated for today's reality - they're time consuming, require specialized manual intervention, depend on human subjectivity and are destructive. To give statistical relevance to the evaluation, a large amount of samples is needed, destroying the product and generating waste (Vanoli & Buccheri, 2012), which thus results in revenue losses for producers.

Newer assessment methods are needed. Not only should they be non-destructive, fast, and produce accurate measurements, but they also need to be simple and user friendly, enabling any worker to use them, making it accessible for all kinds of businesses.

Besides all the requirements stated, technology allow these methods to go even further. The hardware device that collects the data should be light and portable - as opposed to bench top machinery-, and complemented with a smart software, which processes the data and provides accurate answers and predictions.

This field is widely developed, with several different mechanisms and algorithms to measure and assess fruits and vegetables quality already in use. One of which is Electrical Impedance Spectroscopy (EIS), a growing technology in the agricultural community.

This dissertation aims to contribute with a comparative study of several machine learning classifiers to attain the most accurate and adequate model for fruit quality classification. For this purpose, the system will use impedance data acquired through a small impedance spectroscopy analyser device, and consequently, classify whether the fruit is good for consumption or not. This algorithm can be incorporated in a portable hardware device for pre-harvest, storage, and in-store applications.

The test subject will be the common commercialized banana from the *Cavendish* group.

1.3 Approach and Contributions

To accomplish the mentioned objectives, it was first necessary to constitute the data set.

The impedance data was acquired using an Impedance Analyzer Adapter on the Analog Discovery 2 device. The adapter was connected to two personalised wires, which connected to the ECG electrodes by alligator clips. The subject under test (SUT) - the banana -, had two ECG

electrodes oppositely disposed. With this setting, EIS was applied to the SUT on a frequency sweep from 10Hz to 1MHz and impedance data was acquired via the WaveForms software. The tests were carried out on 10 bananas throughout a 32-day period, alongside the respective day's room temperature and humidity.

Classification was done daily, based on the fruit's peel colour, using a metric present in the literature. Each day, for each banana, 10 EIS measurements were made. From this data, a pre-processing stage was executed to remove outliers.

To fulfil the objective of performing a comparative study of different machine learning approaches, 8 classification models from the Scikit-learn (or Sklearn) library, as well as the XGB library, were tested and parameterized to achieve the best performance.

In summary, this dissertation followed the stages presented in Figure 1.1.

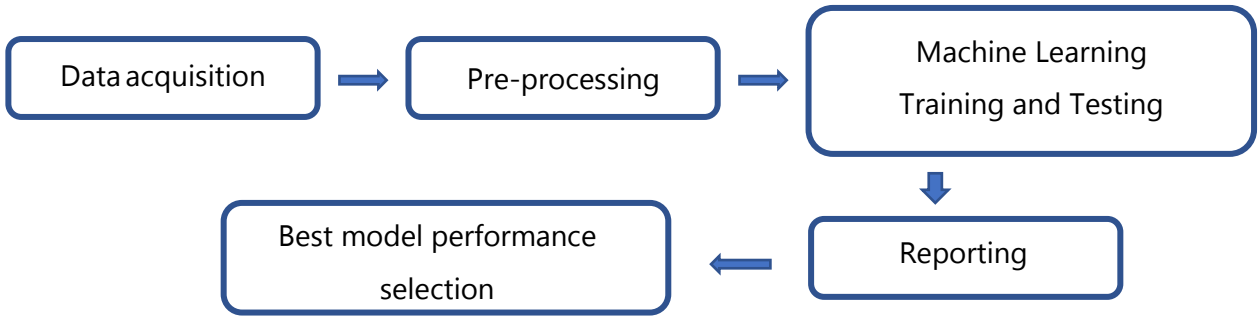

Figure 1.1 — Flowchart of the proposed stages for this thesis completion

From the resulting work produced by this dissertation, the following contributions can be drawn:

- A first known to date application on fruit quality assessment using EIS with the Analog Discovery 2 device;
- The constitution of a dataset with impedance data from several bananas, throughout a 32-day period;
- An extensive comparative study of several machine learning models;
- A submitted conference paper regarding this research.

1.4 Document Organization

To better understand the proposed work, this dissertation is divided into the following chapters:

• Chapter 1 - Introduction: a contextualization of the problems and motivations behind this dissertation are given, as well as an overview of the technical approach proposed.

• Chapter 2 - Fundamental Concepts: a theoretical explanation is presented about the topics discussed throughout this document.
• Chapter 3 - State of the Art: some examples are shown about non-destructive methods for fruit assessment and some results of the usage of machine learning for the same purpose.
• Chapter 4 - Technical Approach: in this chapter, the methodologies chosen to achieve the proposed objectives are presented and justified, alongside with the materials used.
• Chapter 5 - Results and Discussion: the obtained experimental results are presented and discussed.
• Chapter 6 - Conclusion: a summary of the developed work is presented, highlighting the main results. Furthermore, future work is suggested.

2

FUNDAMENTAL CONCEPTS

In this chapter, relevant concepts necessary to better understand this thesis work will be further explained in detail. An overview of fruit quality and its biochemical ripening process will be explored. Moreover, the basis of impedance spectroscopy and related topics are detailed, as well as the basis and definitions of the machine learning models applied.

A reader familiar with these topics can skip to Chapter 3.

2.1 Fruit Quality

Fruit is an indispensable element of the food wheel. It is a rich source of nutrients, minerals and anti-oxidants (Abasi et al., 2018). This importance of fruit in humans' diet reinforces the importance of fruit quality.

From the tree to the consumer in markets, fruit quality needs to be monitored for several purposes. Depending on the fruit, measuring its status is necessary to collect data about the fruit's ripening stage, to evaluate possible diseases or defects, to estimate the rest of the crop's status and predict optimal harvesting times (Abasi et al., 2018; Srivastava & Sadistap, 2018).

Post-harvest quality is also rather important to manage fruit's transportation and storage. Even though technology has developed agricultural storage facilities with refrigeration and controlled atmospheres, spoilage and fruit degradation is a frequent issue (Abasi et al., 2018). Thus, it is important to assure the fruit's quality before reaching the consumers and avoid waste.

In the markets, fruit quality reflects the preferences of the consumer since it's what brings the business's revenue, overcoming health and nutritional values. At this stage in the

production chain, external factors like colour, size, absence of defects and deformities are more important, but internal factors (e.g. texture, taste, sweetness, acidity) that can only be tested by the consumers after purchase are important as well, since they'll determine if the consumer will acquire the product again (Musacchi & Serra, 2018). In supermarkets, assessing fruit quality is also important since it'll determine the shelf-life of the product.

The definition of fruit quality is broad and can change depending on one's point of view and the final purpose of the fruit. This definition could be generalised as a *"dynamic synthesis of their [fruit] physicochemical properties and related to consumer perception*" (Musacchi & Serra, 2018).

It's clear that measuring the quality of fruits through its many stages is indispensable. These measurements should be fast, accurate and non-destructive. Traditional methods lack in these characteristics, resulting in fruit waste for each tested sample. This has brought attention to newer methods to test fruit quality, methods that are also cost-efficient, intelligent, portable and simple enough so unspecialized personnel can use them (Abasi et al., 2018; Ibba et al., 2020).

2.2 Fruit ripening

The fruit is a rather popularly known component of a plant's life cycle. In most cases, after the reproduction phenomenon, a seed is formed, covered by a fleshy pulp - the pericarp. The pericarp, among many other functions, acts as a food source and attracts living beings, assuring the distribution of seeds throughout nature and maintaining the offspring of the plants species, since the seeds are non-digestible, depending on the animal consuming them (Chacón & Bustamante, 2001).

To accomplish this task, fruits have a goal to attract as many distributors as possible. They must be ripe enough. A goal humans have found very attractive for millennia, since fruit has been part of humans' diet and part of the growing cultures of the sedentary *Homo sapiens* (Colledge et al., 2004).

Ripeness is a natural process which could be described as "*a highly coordinated, genetically programmed, and an irreversible phenomenon involving a series of physiological, biochemical, and organoleptic changes, that finally leads to the development of a soft edible ripe fruit with desirable quality attributes*" (Prasanna et al., 2007).

This field has been widely studied in recent years due to the interests of biochemists, geneticists and producers from the agricultural industry (Prasanna et al., 2007). Understanding

the ripening process is key to achieve an optimal harvesting time, which consequently leads to a higher fruit quality to reach the markets and reduce losses in waste or low fruit quality, which can happen due to early or late harvesting times (Abasi et al., 2018; Musacchi & Serra, 2018).

2.2.1 Climacteric and non-climacteric fruits

In terms of ripening mechanisms, fruit can be classified into two categories: climacteric and non-climacteric. Simply put, climacteric fruits are able to mature after being harvested, while non-climacteric fruits should only be harvested after reaching full maturity (Pua & Davey, 2010).

This is possible since the fruit's biological functions don't stop after being separated from the parent tree. Processes like respiration and biosynthesis of compounds that stimulate maturation (i.e. ethylene), keep functioning. The difference between these two categories is that climacteric fruits demonstrate an increase in respiration and ethylene production post-harvest, and in non-climacteric fruit it's residual to non-existent (Prasanna et al., 2007; Pua & Davey, 2010).

In Table 2.1, some examples of fruits from these two categories are given.

Table 2.1 — Classification of fruits (adapted from (Prasanna et al., 2007))

Climacteric fruits	Non-climacteric fruits
Banana	Cherry
Apple	Lemon
Mango	Grape
Peach	Pineapple
Kiwifruit	Orange

2.2.2 Physiological and physical changes

The phenomenon of fruit ripening is of physiological, biochemical and molecular nature (Pua & Davey, 2010). Within these natures, one can find more visible and physical changes such as skin colour, aroma, stiffness, taste, shape, or size (Abasi et al., 2018). However, many more changes occur to the fruit during ripening.

Perhaps one of the most famous biochemical evolutions in fruit ripening, is the increase in the biosynthetic production of ethylene, a plant hormone (phytohormones). This fruit ripening compound is able to trigger cell metabolism and senescence (aging) in a short period of time (Prasanna et al., 2007). As seen before, climacteric fruits can keep ripening post-harvest

due to the continuous production of this hormone. In the agricultural industry, the manipulation of ethylene quantities is used to control under and overripening of fruits and vegetables like tomatoes (Mansourbahmani et al., 2018).

Another major change in fruit revolves around pulp texture and cell wall structure. Enzymes present in fruit's cells are responsible for degrading the cell walls, resulting in changes in its "*thickness, permeability of plasma membrane, hydration of cell wall, decrease in the structural integrity, and increase in intracellular spaces*" (Prasanna et al., 2007). Depending on the fruit species, this enzyme-mediated process happens with different specifications and at different speeds, but overall, it ultimately results in fruit softening (Prasanna et al., 2007).

Due to the aforementioned factors and many other biochemical reactions, during the fruit ripening, changes also occur in the levels of cell respiration, cell organelles and elements variations - chlorophyll degradation, carotenoids production, anthocyanins -, essential oils, soluble solids content (SCC), acidic components -titratable and ascorbic-, starch and sugars quantities (Abasi et al., 2018; Prasanna et al., 2007).

Understanding this fruit ripening development is relevant in this thesis to better understand how fruit can behave differently depending on its ripeness degree (Abasi et al., 2018). From an electrical point of view, fruit's impedance will vary as well with the ripening process. Thus, one can characterize fruit in terms of its effective resistance and effective capacitance (Rehman et al., 2011), which can be measured with various techniques, like spectroscopic impedance analysis, a good method for internal fruit quality investigation (Abasi et al., 2018).

2.2.3 Bananas

The banana production has a great impact in tropical economies and the world's food market, cultivated in over 130 countries. It is the second largest fruit production, corresponding to 16% of the world's total fruit production, with India the main contributor to this industry (Mohapatra et al., 2010).

Biologically, bananas are classified as a berry. Selective breeding led to the production of edible seedless bananas. The most common internationally traded group of this fruit are the *Cavendish* bananas, also referred to the dessert banana, due to its sweetness and a reduced amount of starch contents (Mohapatra et al., 2010; Seymour, 1993).

As stated in Table 2.1, bananas are climacteric fruits and producers make use of its ripening mechanisms to commercialize and export the product with good quality. After reaching full maturity whilst still unripe, the bananas are harvested and stored in refrigerators, where they remain during shipping. The temperatures of these stages vary from banana species and

purpose. Usually, the closer to the industry chain end, the higher the temperature since it'll help trigger the ripening mechanisms. Besides the natural phytohormone ethylene can be applied in a synthetic exogenous form, fast forwarding the ripening process accordingly with the intended timing (Seymour, 1993).

From a biochemical perspective, banana ripening leads to an enzymatic conversion of starch contents into sugar and a transference of moisture from peel to pulp. This explains the increased sweetness and softer texture of a ripe banana. Softening of the peel and pulp's texture is also related to cell wall degradation caused by the increased solubility of pectins. Banana ripening can be observed by its peel colouration change, from green, to yellow, with dark spots developing and ending in a brown tonality. Bananas are rich in antioxidants which, in its oxidation process, lead to the production of the tannins responsible for the dark spots. The overall change in peel colour is due to chlorophyl degradation, as it happens with oranges or tree leaves in Autumn, exposing the remaining carotenoid components (Mohapatra et al., 2010).

Lastly, bananas are rich in vitamins and minerals, with large amounts of potassium, followed by magnesium and calcium. Considering their carbohydrates and antioxidant properties, banana based foods are healthy and important for the human diet, especially in infants (Ruales et al., 1990). The peel can be used as cattle food, but also as a base material for alcohol and bio-gas production (Mohapatra et al., 2010).

Due to its high importance in agricultural markets, producers are always looking forward for strategies and techniques that provide the best quality for the consumers, increasing the business earnings.

2.3 Basics of Impedance

Every engineer has learnt what impedance is to some extent. However, given the importance of this electrotechnical concept for this thesis, a quick overview of impedance fundamentals will be given. Liu (2006) summarized this concept very neatly.

As popularly known, all materials have an electric resistance. This resistance is the literal resistance presented by the materials nature against the flow of electrons, the electric current. Ohm's law states a relation between resistance and current, known as the voltage. This law translates this relation into the following equation:

$$v(t) = R \times i(t) \tag{2.1}$$

R is the resistance, measured in ohms (Ω), and only assumes values equal or greater than zero. *v(t)* is the potential difference between the two points of a resistance. *i(t)* is the current flow through the resistance, in DC.

However, in circuit theory, this law applies only to this circuit element: the ideal resistor. When using an ideal resistor, one can assume that Ohm's law is valid for all values of current and voltage and its phases (for AC current) are equal, and the resistance presented by the resistor is independent of frequency (T. Frelink, 2006).

In real life, a material's resistance is presented with more complex nuances. Thus, the concept of impedance is born. It still represents the resistance against the flow of electrons but has more properties that allow electric circuit representation of real-life materials. Impedance is a frequency-dependent quantity, also expressed in ohms, usually represented by the letter Z.

Then, in the AC domain, Ohm's law can be rewritten as:

$$V = I \times Z \tag{2.2}$$

Since impedance is a complex quantity, it can be expressed in rectangular form as

$$Z = R + jX \tag{2.3}$$

where R is the real component (resistance), and X is the imaginary component (reactance). The reactance may be positive or negative. The real and imaginary parts of impedance can be represented graphically like in Figure 2.1.

Impedance can also be expressed in polar form as:

$$Z = |Z| \angle \theta \tag{2.4}$$

Therefore,

$$Z = R + jX = |Z| \angle \theta \tag{2.5}$$

where

$$j = \sqrt{-1} \tag{2.6}$$

$$|Z| = \sqrt{R^2 + X^2} \tag{2.7}$$

$$\theta = \tan^{-1} \frac{X}{R} \tag{2.8}$$

and

$$R = |Z| \cos \theta \tag{2.9}$$

$$X = |Z| \sin \theta \tag{2.10}$$

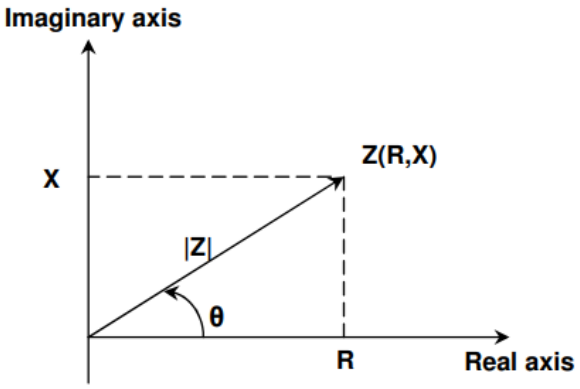

Figure 2.1 — Impedance complex plane (adapted from (Liu, 2006))

Besides resistors, there are two other passive elements - capacitors and inductors. These elements, besides other properties, are able to store energy (Liu, 2006) and depend on the frequency of the AC signal. Both these elements have an impedance, which can be represented by:

$$Z_C = \frac{1}{j\omega C} \tag{2.11}$$

$$Z_L = j\omega L \tag{2.12}$$

C represents the capacitance of the element, expressed in farads (F), L the inductance, expressed in henry (H), and ω corresponds to the angular frequency of the applied signal. In the DC domain, capacitors and inductors behave differently. A capacitor will act as an open circuit whereas an inductor will behave as a short circuit (Liu, 2006).

Ohm's law described in (2.2) can be applied to inductors and capacitors as well, with the respective impedance.

2.4 Conductivity

Conductivity is a property of materials that describes their ability to conduct the flow of electricity. Some materials are good conductors due to their atomic nature, while others are not. Those materials are considered insulators due to their very low conductivity. The SI unit of conductivity is expressed in siemens per meter (S/m).

Biological tissues act as conductors as well, depending on the chemical composition. For example, a high concentration of hydrogen ions in a solution allows electrons to travel through particles more freely. This example is relevant since this phenomenon occurs in fruit,

which are rich in organic acids. With ripening, biochemical reactions with these acids variate hydrogen ions concentrations, thus influencing the fruit's conductivity (Liu, 2006).

2.5 Electrochemical Impedance Spectroscopy (EIS)

Every material has electrical properties that can be studied through tests and techniques that interact with the electric state of the studied object. One of these techniques is Impedance Spectroscopy (IS), which can be used to measure the impedance of a system as a function of frequency. This method has been used to study the electrical behaviour of solid-liquid systems, like fuel cells and chemical sensors, or biological tissues and membranes, like fruit (Al-Ali et al., 2017; Macdonald & Johnson, 2005). When applied to the latter, it can be referred to as Electrochemical Impedance Spectroscopy (EIS), where the impedance of an electrochemical cell as a function of frequency is measured.

Impedance spectroscopy studies the small-signal electrical response of an object and its changes in the current flow. According to the object's characteristics, this response will vary and allow one to draw conclusions. Useful information from this response can be obtained, from "*mass transport, rates of chemical reactions, corrosion, and dielectric properties, to defects, microstructure, and compositional influences on the conductance of solids*" (Al-Ali et al., 2017; Macdonald & Johnson, 2005).

Depending on the purpose of the EIS application, the applied signal can have different parameters. An AC voltage can be applied, resulting in a current response and vice-versa. The ratio between the applied signal and the response provides the impedance of the measured system, respectively according to equation 2.2. Since organic cells behave differently to different frequencies, these tests are often applied over a wide range of frequencies (Hussain et al., 2021). The signal and respective response are measured via electrodes connected to a hardware responsible for the EIS execution. The types of electrodes and their disposition influence the EIS experiment. This topic is further explained in chapter 2.5.1.

Biological tissues have complex constitutions. There are extra and intracellular fluids, membranes, organelles and cell walls, all characteristics of vegetable cells. And if EIS is applied to a whole fruit, there are different parts of the fruit: peel, pulp, seeds. All these details will affect the conductivity of the EIS applied signal. Electrically, fruit could be seen as an aggregation of capacitors and resistors, in series and parallel. Studies show that for lower frequencies, current flows mainly through extracellular fluids that contain higher concentrations of ions, which results in a higher impedance value since cell walls and membranes become obstacles

for the current path. For higher frequencies, cell organelles and intracellular fluids conduct current as well, which leads to lower impedance values (Ibba et al., 2020; Liu, 2006), as exampled by Figure 2.2.

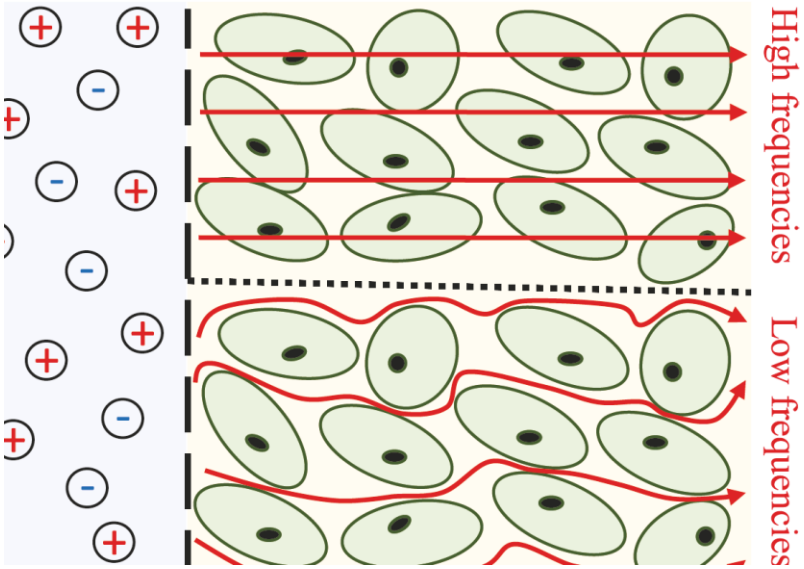

Figure 2.2 — Current flow through plant cells (adapted from (Ibba, 2020))

These resistive and capacitive behaviours of cells in EIS can be described by equivalent circuits and help characterize experimental frequency responses. This provides information about the samples, about its properties and behaviours under the applied signal (Ibba et al., 2020). Equivalent circuits may be designed for a specific process, or previously studied and versatile equivalent circuits could be used (Liu, 2006), like the Cole, Hayden or Double-shell models, which can both represent a single cell as well as a whole tissue since they're made from a grid of cells. While more complex equivalent circuits can better represent a specific cell behaviour, they also increase the computational resources and risk data overfitting. Equivalent circuits also often lack in representing the influence of the electrode-electrolyte interface in the impedance measurements (Ibba et al., 2020). Ibba et. al (2020) proposed the inclusion of a Warburg element in equivalent circuits to emulate this interface. Overall, equivalent circuits are a resourceful tool to simulate EIS experiments and draw further conclusions about cell behaviour.

Even though IS has been growing popular in many fields, there are still some issues limiting the measurements' accuracy. For example, the reliability of electrode disposition in bulk impedance tests or the materials constituting the electrodes. Hwang et al. (1997) noted the presence of spreading resistance effects (poor contact) due to lack of polishing in pressed metal electrodes.

EIS is widely used in biomedical, chemical, material engineering and similar fields. One of its main advantages is the non-invasive implications for impedance measurements. This allows researchers to study biological tissues *in-vivo*, without damaging the sample (Chowdhury et al., 2018). Besides the research fields, EIS is also used in industries. In the agricultural industry, EIS can be applied for evaluating fruit ripeness, according to its biochemical and physical changes, the measured values of impedance will vary as well (Al-Ali et al., 2017). As seen previously, IS techniques can detect defects in the samples. This also applies to EIS in the detection of bruises and other anomalies in fruits, bringing advantages to producers in storage centres, since they're able to early detect damaged fruits before browning and ruining other fruits (Bakr et al., 2016).

In Figure 2.3 is possible to observe an example of an EIS experiment setup on a mango fruit. The application of EIS in practical cases will be further detailed in section 3.1.4.

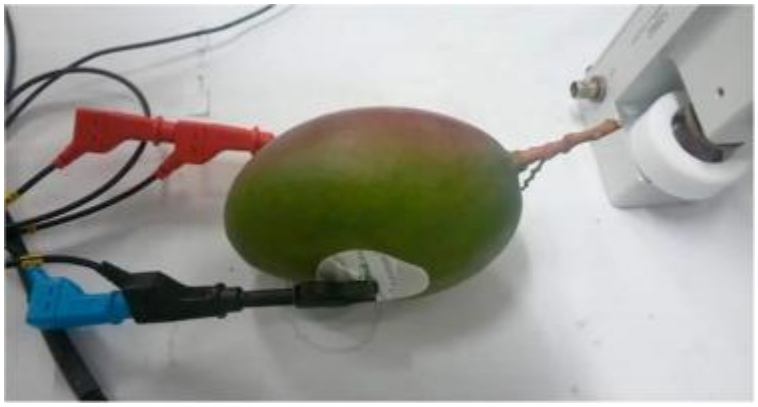

Figure 2.3 — Two-electrode EIS experiment setup on a mango (adapted from (Figueiredo Neto et al., 2017)

2.5.1 Electrodes

Regardless of the type of electrodes, this element is fundamental for EIS experiments. There are several types of electrodes that can be used to measure fruit impedance. Some are inserted inside the SUT, referred as "invasive" or "implantable" electrodes. Others are attached on the body's surface, such as pre-gelled ECG, fabric/textile electrodes or even clamp electrodes, which do not affect the samples integrity. In EIS experiments, standard Ag/AgCl ECG electrodes are commonly used due to their availability, low-cost and the pre-gelled characteristic, which means these electrodes come with an electrolyte gel layer to facilitate the signal's conduction between the sample's surface and the metallic part of the electrode (Ibba, 2020).

The interface between sample-electrode usually presents an issue to measurement accuracy, since this junction can affect charge transfer. For example, when using ECG electrodes

in humans, the natural skin sweat can be an inconvenience since the saline contents will affect the electrical signal being conducted through the electrodes. To overcome this issue, an offset can be applied to compensate the input's signal (Corbellini & Vallan, 2014). However, for fruit applications, this doesn't present a problem as observed in the literature.

Electrode configuration also plays a role in EIS performance. The measurements can be made using electrical conducting electrodes on the sides of the sample under test (SUT), directly in front of each other. Two or four electrodes can be used, forming a circular cylinder or rectangular parallelepiped between each pair of electrodes. The main difference relies on the fact that with a two-electrode disposition, both the signal excitation and the respective response are measured, whereas with four-electrode disposition, these tasks are separated and thus there is no risk of electrode polarization (Liu, 2006). Depending on the object studied, these geometries might have to be adapted.

2.6 Machine Learning

Machine Learning (ML) and other branches of artificial intelligence (AI) are a very common topic in contemporaneous technology. These sciences make use of large amounts of data to teach and allow systems to be autonomous, making decisions and arriving at conclusions without being specifically programmed to do so. This is what Machine Learning is about: training intelligent models with data and enabling the machines to learn over time, identify patterns and make decisions or predictions, without human intervention. And the more data available, the more a ML model can learn.

The growth in computational power and technology accessibility led to the expansion of machine learning in several fields, from robotics, medicine, chemical sciences, to meteorology, economics and agriculture (Liakos et al., 2018). Whether it's daily tasks or more complex processes, the autonomous decision-making ability of machines makes things safer, more accessible and easier for people and industries, as it happens with self-driving vehicles or image recognition of abnormalities in production lines (Batta, 2018).

ML can also be divided into two types, according to the training process: supervised and unsupervised learning. Supervised learning trains the models from data in which classes are already known. This means the input and the expected output already exist within the provided data. Thus, the model will learn and be able to correctly identify the correspondent classes in new data. On the other hand, unsupervised learning uses unlabelled data. During training, the

models will learn and identify patterns and paths within the data, providing its own deductions (Batta, 2018).

Within these two types of machine learning, there are dozens of different algorithms, each with different possible parameters and applications. Depending on the objective of the user and specifics of the study, some models might be more adequate than others. Generally, you can divide ML algorithms in Classification and Regression models, regarding the wanted output.

Classification models aim to predict a discrete output. The class can be binary or have several labels, but they are limited. Classification models can be used to predict if an e-mail is spam or not, or for example, whether a fruit is unripe, ripe, or overripe. In contrast, regression models' outputs are continuous or numeric. These models are used to predict real estate prices or to predict the weather's temperature. To apply machine learning to a problem, many steps and decisions must be considered. The raw data must be pre-processed, relevant features must be selected and the output has to be reported in order to conclude which ML model should be applied.

The next chapters will explore the classification models chosen for this dissertation and other relevant ML steps.

2.6.1 Pre-processing

Data can come in many forms from the source. To use this information for ML, it is necessary to prepare the data first to attain the best results. Of course, the adequate pre-processing changes for each scenario, but there are some common steps.

It is first necessary to clean the raw data. This mainly means removing outliers and any other values that might affect the quality of the data, including handling missing values.

Then, if necessary, a data transformation step can be applied. This means converting all necessary data into a type that can be analysed by ML algorithms, or even create new features from the original features. For example, in supervised learning with classification models, if the class is in words, it should be converted into numbers. The same can happen if the input is constituted by text. There are functions that convert words into numerical forms that ML model can learn easier.

It is often necessary to normalize the data as well. Data normalization is the process of scaling features into a common interval, allowing the model to better compare and correlate the information and converge. This is useful when the value range of each feature has a wide amplitude comparing to other features. For example, weather temperature might range only

in a short interval whereas the population of a country's cities will be in the order of magnitude of millions. There are several types of data normalization that can be applied.

Sometimes there's a lot of data available, but it doesn't mean every bit of it is useful for a good ML application. Thus, feature selection and feature extraction steps are important to select only the useful information needed. This is also known as dimensionality reduction of the dataset.

Despite the name, pre-processing is a recurring step in ML application that'll help improve and optimize the algorithm's results.

2.6.2 Classification models

For the purpose of this thesis, only supervised learning classification models will be studied. The following chapters present a summary of the applied ML algorithms.

2.6.2.1 Support Vector Machines

Support Vector Machines (SVM) are a binary classifier that can be used both for classification and regression applications. The linear SVM kernel aims to find a hyperplane that best separates and classifies the data instances. The radial basis function (RBF) SVM uses higher-dimensional hyperplanes to separate more complex data, non-linearly (Liakos et al., 2018).

2.6.2.2 Decision Tree

As the name indicates, Decisions Trees (DT) have a tree-like architecture and can be both used for classification and regression applications. Through sequential comparisons between features, the dataset is decomposed into organized subsets. Each comparison is a node that forms a certain number of branches (the subsets). The predictions will be the leaves of these branches after the necessary number of comparisons. Thus, from the original dataset (the root), several branches will form, ending in the leaves (the predictions), forming a tree (Ibba, 2020).

2.6.2.3 Neural Networks

The Neural Networks model (NN) (also referred as Multi-Layer Perceptron (MLP)) is based on the human's brain behaviour. This means that the model will be based on a layered network of neurons which connect and interact with each other. During the training process, the model will provide different weights to the neurons according to which combination provides the best score. The model itself applies a feedback mechanism (or back propagation) to ensure this weight adjustment (Voyant et al., 2017).

2.6.2.4 Adaptive Boosting

The Adaptive Boosting algorithm, or more commonly known as AdaBoost, consists of the use of several different simpler classifiers, referred as weak classifiers, usually decision tree like models with very few nodes, and trains them with subsets of the dataset. Then, based on these outputs, the model assigns weights to each weak classifier according to its score, focusing the higher weights on the more poorly classified samples. A new training cycle is executed, and importance values are attributed to each weak classifier according to its score. It then combines the output of these classifiers and provides a final prediction, as a strong classifier (Carreira-Perpiñán & Zharmagambetov, 2020).

2.6.2.5 Logistic Regression

Despite the name, a Logistic Regression (LG) is a classification ML model. Through a linear combination of the input data, the model applies a logistic function to predict the probability of a certain class occurring as the output. Similarly to previous models, LG also sets importance weights to the input features in order to maximize the model's score (Ibba, 2020).

2.6.2.6 Gradient Boosting

Like Adaboost, Gradient Boosting (GB) is a model that makes use of the output of multiple weak classifiers to build a strong classifier. The differences rely on the handling of losses - GB applies a new weak classifier to the residual data of a previous weak classifier -, and the higher complexity of the weak learners used, which can lead to a higher computing time, but makes GB a more robust model in comparison (Carreira-Perpiñán & Zharmagambetov, 2020).

2.6.2.7 Extreme Gradient Boosting

Extreme Gradient Boosting, or XGBoost (XGB), as the name indicates, is an improved version of the GB model. Overall, the key improvements of XGB are the handling of overfitting by penalizing more complex weak classifiers when dealing with loss functions; while GB ignores missing data, XGB assigns missing values to one of the branches of the weak classifier. Depending on the computing device, XGB supports parallel boosting, functioning on distributed systems. These improvements allow XGB to be faster and more flexible (T. Chen & Guestrin, 2016).

2.6.2.8 Naïve Bayes

Naïve Bayes models (NB) are probabilistic algorithms based on the Bayes' theorem, which states that the conditional probability of an event occurring can be determined based on the

probabilities of another events. Applied to ML, this means that the probability of a class can be determined through the probability of individual input features, which are considered independent from each other. The probability of each class is also considered independent from the other classes. This "naïve" assumption simplifies the calculations of this model, making it a fast classifier for large datasets (Liakos et al., 2018).

2.6.3 Model reports

The success of a machine learning algorithm can be assessed by several different metrics that provide different information about the model's behaviour. Here's an overview of the most common metric reports for classification models:

- **Accuracy**: corresponds to the ratio between the correct predictions and the total number of predictions. It's the overall metric to evaluate a ML classifier.
- **Precision**: corresponds to the ratio between the true positive predictions and the total number of positive predictions considered by the algorithm. This metric helps measure the amount of false positives.
- **Recall**: corresponds to the ratio between the true positive predictions considered by the algorithm and the real number of positive predictions. This metric helps measure the amount of false positives.
- **F1 score**: combines precision and recall values through their harmonic mean. It is also an overall metric to evaluate a ML model.
- **Confusion matrix**: corresponds to a matrix which can be interpreted as a table. It shows quantitively the number of true and false negatives, and true and false positives. It's useful to quickly analyse the predictions made.

2.6.4 Model tuning

After pre-processing the data, training, and testing the ML models, one can also tune the model. This simply means testing different parameterizations of the algorithms to achieve a better accuracy. The parameters are referred as hyperparameters and only influence the model's approach to the data, they don't change the input.

This process can be done through automatic tools where the user provides a range of hyperparameters, and the model is trained and tested for each possibility and a report will show the best achievable performance.

3

STATE OF THE ART

Chapter 3.1 will focus on previously studies and implementations of non-destructive fruit quality analysis, with an emphasis on state-of-the-art EIS techniques. Further, chapter 3.2 will present machine learning application results on fruit assessment using data acquired through non-destructive methods.

3.1 Fruit ripeness - non-destructive assessment methods

Fruit quality parameters can be obtained through a myriad of techniques. More antiquated methods (e.g. impedance with electrodes inside the pulp and pressure measurements; opening fruits) not only destroy the samples tested and generate waste, but also require a certain level of expertise handling the devices and large amounts of data processing, involving bench-top and lab materials, which are time consuming and consequently expensive due to all the commodities necessary for their application. Furthermore, large amounts of samples have to be destroyed in order for this data acquisition to have statistical relevance to aid producers decisions about the crop's management (Abasi et al., 2018; Kitinoja & AlHassan, 2012; Vanoli & Buccheri, 2012).

These challenges increase the desirability for modern, user-friendly, cheap, and non-destructive techniques to assess fruit quality. Since fruit ripening and following harvesting are key for a successful agricultural business, it's also required that these new techniques enable fast and accurate measurements, allowing several readings throughout the crops, storage and supermarkets (Abasi et al., 2018; Srivastava & Sadistap, 2018; Torres et al., 2017).

All these requirements in the last decades led to the development of several different devices that measure different characteristics in fruit, with different pros and cons as well.

Despite these advancements, there's still a long technological path to be explored and improved (Srivastava & Sadistap, 2018).

In the following chapters, state-of-the-art non-destructive methods will be presented, along with a few examples of the respective devices. Many other experimental examples can be found in the review literatures of Abasi et al., 2018 and Srivastava & Sadistap, 2018.

3.1.1 Electronic Nose

An electronic nose (e-nose) in simple words, is a device integrated with sensors which read aerial chemical substances released by fruits. This technology, born from chemical engineering and related fields, is able to automatically "*detect, measure and characterize volatile organic compounds*" such as "*odours, vapours and gases*" (Srivastava & Sadistap, 2018). These measurements evaluate chemical compounds derived from fruit ripening such as soluble solids variations, acidic and starch contents, ethylene production, determining optimal harvesting times based on the collected data (Abasi et al., 2018).

The high performance of these devices even allows them to differentiate between distinct fruits since the biochemicals released will also be different (El Hadi et al., 2013). The system

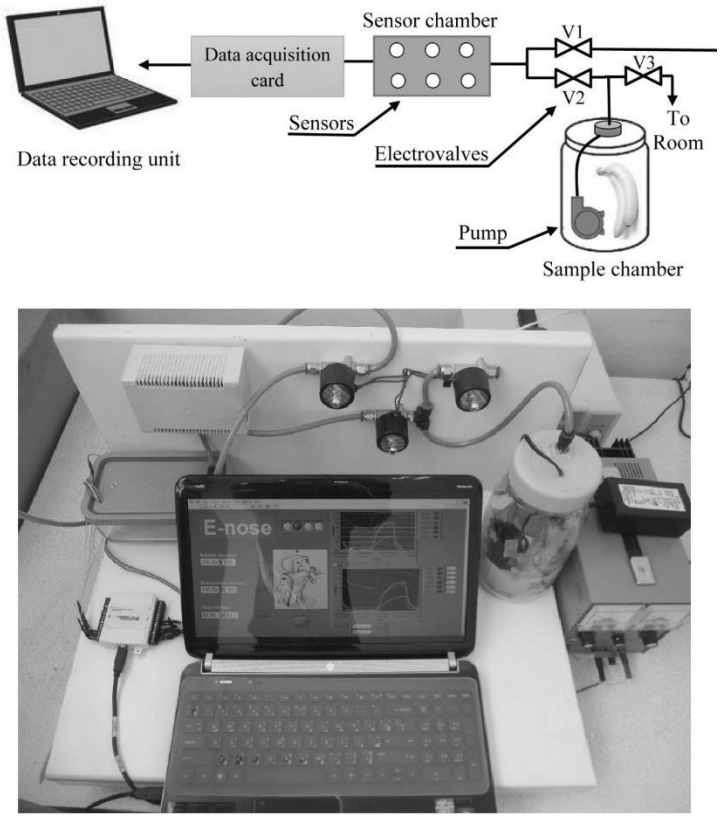

Figure 3.1 — Electric nose experiment setup (adapted from (Sanaeifar et al., 2014))

gathers the data from the various sensors and interprets them into specific unique aroma fruit signatures (Srivastava & Sadistap, 2018).

Li-Ying Chen et. al tested an electronic nose based on metal oxide semiconductors sensors on bananas. The samples were stored inside glass bottles to preserve the odours originated from the banana. These gases are then passed through a chamber with the sensors, which will perform the measurements. The experience occurs over several days, collecting data from the bananas over their different ripening stages. Machine learning classification algorithms were trained and tested with the acquired data, obtaining 96,66% accuracy for SVM classifier and even 100% for the k-nearest neighbour (KNN) model (L. Y. Chen et al., 2018). A similar experimental setup is presented in Figure 3.1.

Nonetheless, its indubitable accuracy makes electronic noses a good tool for science, having applications in other fields as well (e.g. "*biomedical, cosmetics, environmental, food, manufacturing, military, pharmaceutical, regulatory, and various scientific research fields*") (Wilson & Baietto, 2009).

3.1.2 Optical methods

There are several different types of non-destructive devices and techniques that use optical properties to assess the fruit's ripening stage, which are amongst the most popular methods. Two different types of optical fruit quality evaluation can be enumerated: UV-VIS-NIR Spectroscopy and Image Processing.

3.1.2.1 UV-VIS-NIR Spectroscopy

UV-VIS-NIR stands for "ultraviolet-visible-near-infrared". This technique relies on spectroscopy and wave reflection/absorption concepts. A light signal is emitted by the device onto the fruit's surface. Depending on the wavelength of the beam emitted and the fruit's composition, the light will be absorbed in some parts, reflected, and refracted by others. Then, the device will read and interpret the reflected radiation and draw conclusions. In Figure 3.2 there's a schematic exemplifying the possible light distribution from an incident beam on a fruit sample. These optical phenomena change according to the fruits internal levels of soluble solids content, sugar, acidity, texture and other characteristics that evolve with fruit's natural maturation process (Gao et al., 2010; Srivastava & Sadistap, 2018).

Within the three wavelength types referred, the most used are visible and near-infrared signals. NIR spectroscopy has the advantage of being able to penetrate better the fruits elements, namely the skin, disregarding of any sample preparation, allowing instantaneous

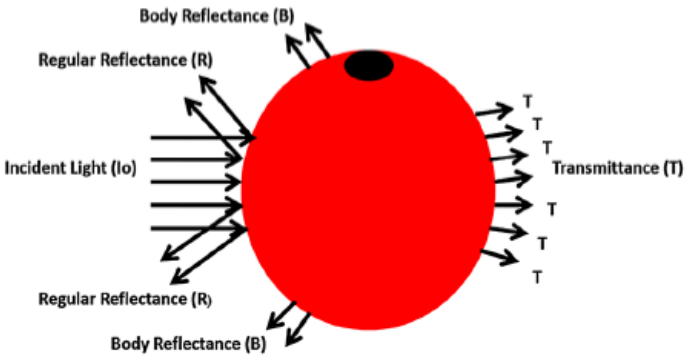

Figure 3.2 — Light distribution scheme (adapted from Srivastava & Sadistap, 2018)

measurements. NIR also allows more accuracy in the evaluation of fruit's internal composition quantities (Srivastava & Sadistap, 2018).

Manuela Zude (2016) from the Institute of Agricultural Engineering Bornim tested visible and near-infrared spectroscopy in bananas to detect and evaluate chlorophyll and sugar contents on the sample's skin surface. The experience was performed on bananas with different peel colourations, corresponding to different ripening stages. It was concluded that VIS spectroscopy was able to better analyse chlorophyll contents, whereas NIR performed better in detecting sugar contents (Zude, 2016).

3.1.2.2 Image processing

Fruit quality can be assessed through images as well, with the aid of digital cameras and machine learning and deep learning methods. Within fruit quality, this method allows fruit sorting based on its shape and size, skin colour evaluation, detection of deformities and diseases and analysis of fruit's composition through hyperspectral imaging (Brosnan & Sun, 2004; Lorente et al., 2013).

This technique alongside the ML experiment has been widely studied. In 2019, Sabilla et. al tested KNN, SVM and DT models in a dataset with over 5000 images of different banana species, in different ripening stages and framed in different positions as well. Figure 3.3 shows conceptually the setup used to acquire the images. Each image was pre-processed to count each pixel by the respective Red, Green and Blue (RGB) values. The SVM and KNN models achieved the highest accuracy, both with 96,6% (Sabilla et al., 2019).

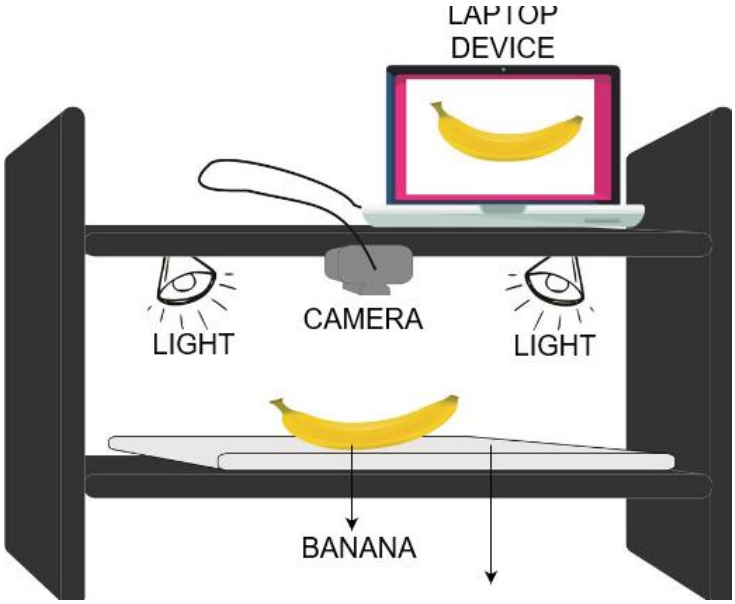

Figure 3.3 — Image collecting setup (adapted from (Sabilla et al., 2019))

While these methods ease off manual labour and reduce the risk of human error in the assessment (Srivastava & Sadistap, 2018), its accuracy depends on the quality of the images acquired (e.g. illumination, orientation, shape/size of the object) (Cubero et al., 2011).

3.1.3 Sound methods

Similar to optical methods, sound-based devices use sound waves and mechanical/acoustic properties in order to draw conclusions about the fruit's physiology and chemical characteristics.

3.1.3.1 Ultrasounds

Relying on physics laws, when an ultrasound wave is applied to an object, there will be reflections, refractions, and absorption of the different mediums of the fruit (skin, pulp, seeds). These behaviours will also vary depending on the fruit's composition (e.g. thickness, density, defects) at the time of the measurement.

Jagannath et. al (2005) tested both banana ripening and bruising with acoustic resonance spectroscopy. An impulse hammer would excite vibrations at a certain frequency onto the fruit bed and thus the banana vibration responses were recorded using a specific type of microphone. Through Fast Fourier Transform, it was possible to extract the banana's resonance frequency and amplitude, from the recorded signal. The samples were then manually bruised, and the experiment was repeated. A relation between the softening of the pulp and peel was

detected between the resonances, corroborated by the decrease in resonance's frequency on bruised bananas (Jagannath et al., 2005).

Amongst the advantages brought by ultrasound-devices, one should highlight the portability of the devices, allowing fast and high sensitive measurements on-field and in every environment (Srivastava & Sadistap, 2018).

3.1.3.2 Impact force

During the ripening process, changes in fruit's density may lead to weight variations and pulp firmness. Penetrometers can draw information about fruit quality through pressure-based measurements, however, the samples are damaged.

Surprisingly enough, impact force techniques rely on the same concepts without destroying the fruit. The sample is dropped onto a load cell that'll measure the impact force of the fall, which through data processing, gives information about the fruit's state. The falling heights are minimal (e.g. 10mm), not causing any damage to the samples (Mireei et al., 2015).

3.1.4 Electrochemical Impedance Spectroscopy - EIS

The fundamentals of EIS were explained in chapter 2.5. Despite having a lower accuracy compared to bench-top laboratory equipment, the small and portable devices of this technique provide simplicity, low-cost application, and can still provide a broad variety of information on the samples. Most importantly, it can be used to measure fruit in different environments (in vivo, in storage, at the shelf) (Ibba et al., 2020; Rehman et al., 2011).

Despite many different applications of EIS, it's often seen in experiments setups using "*frequency response analyzers and LCR meters featuring a wide range of test frequencies, high accuracy, and the possibility to use different electrode configurations*" (Ibba et al., 2018).

Ibba has produced much research on this topic. Ibba (2020) studied the bio-impedance evolution of 9 bananas over a 12-day period with EIS, at room temperature. An Analog Devices AD5933 evaluation board was used for the input signal and response measurement, over a frequency sweep between 100 Hz and 85 kHz. The contact between the microcontroller and the banana was done with pre–gelled ECG (Ag/AgCl) electrodes, on a 4-electrode disposition, as observed in Figure 3.4.

To check the accuracy of this EIS classification, a parallel analysis was executed through image processing of the samples' peel colour. Besides this, the device's accuracy was studied comparing the impedance measured on a 10nF capacitor and the theoretic impedance values.

Ibba detected an increase in the banana's impedance over the first 6 days, a tendency most noticeable at lower frequencies, where the impedance is higher. This analysis happened on the range of 400-85 kHz, where the signal response was less noisy. However, the impedance started decreasing for the rest of the experiment's period, until 40% of the original value. With the image processing analysis, it was possible to create a relation between the different ripening stages of the banana and its corresponding impedance values (Ibba, 2020).

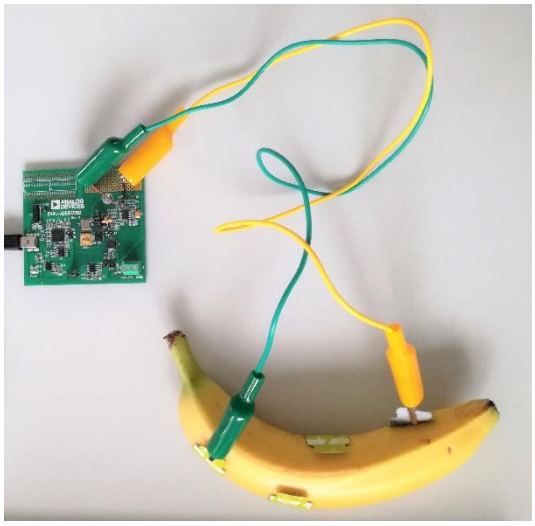

Figure 3.4 — Experimental setup using a four-electrode disposition (adapted from (Ibba, 2020))

A similar experiment was conducted by Kulkarni et. al (2017). For 3 months, hundreds of bananas' impedance were measured with EIS and classified according to its ripening stage. The microcontroller-based approach used an AD5933 evaluation board as well. However, the selected frequency sweep ranged from 100Hz to 12kHz. In this experiment, image processing was also used to classify the banana in accordance with the peel's colour. Given the large number of samples measured, Kulkarni used clamp electrodes that can be reused. The experimental setup can be observed in Figure 3.5.

Alongside with the EIS measurement, soluble solid contents (SCC) were measured as well with a refractometer. Thus, a relationship between impedance values, SCC values and the class of the banana was established. In summary, it was noted a decrease of impedance over time, where the major impedance variations were observed in lower input frequencies. Besides, it was detected that generally, a banana from the same hand (a group of bananas originating from the same node) behaves similarly to the rest of the elements. This means that a single banana from a hand must be tested to evaluate the ripening state of the rest of the hand (Kulkarni, 2017).

Figure 3.5 — Experimental setup using clamp electrodes (adapted from Kulkarni, 2017)

These experiments were proven simple, cheap, efficient and didn't cause any harm to the fruit. Several other examples can be found in the literature of EIS applications in fruit's maturation and quality assessment.

3.2 Machine Learning applications for fruit quality assessment

As seen in 3.1, there is already a large variety of developed work with non-destructive methods for fruit quality assessment, some using intelligent models to predict and classify fruit maturity. However, to date it wasn't possible to find in the literature applications of all the models used in this dissertation on EIS data from bananas.

In Table 3.1 there's a summary of machine learning classification models' accuracy using different non-destructive method analysis. Moreover, other examples using different fruits will be given in the same table as well.

Table 3.1 — Machine learning models' performance on fruit quality assessment

Fruit	Non-destructive method	Model	Accuracy	Reference
Banana	EIS	SVM - linear	100%	(Bertemes-Filho et al., 2020)
		SVM -polynomial	92,9%	
Banana	E-nose	KNN	100%	(L. Y. Chen et al., 2018)
		SVM	96,7%	
Banana	Image Processing	KNN	96,6%	(Sabilla et al., 2019)
		SVM	96,6%	
		DT	95,5%	
Avocado	EIS	SVM	90%	(Islam et al., 2018)
Cape Gooseberry	Image Processing	NN	89,9%	(Castro et al., 2019)
		SVM	93%	
		KNN	90%	
Tomato	Image Processing	SVM	84,8%	(El-Bendary et al., 2015)

4

TECHNICAL APPROACH

This chapter presents the executed technical approach to accomplish the proposed objectives in two stages: Data acquisition in chapter 4.1 and the ML application in 4.2.

4.1 Data Acquisition

To complete the goal of this dissertation, it was first necessary to create a dataset since to the date there wasn't any datasets available online with banana impedance data. Thus, a complete and fully detailed dataset was created, including EIS values, as well as temperature, humidity, the number of days passed since the beginning of the experiment, image examples and the respective classification of each sample.

The EIS experiment was performed with a small portable device, a Digilent Analog Discovery 2 (AD2). It was added to the device an Impedance Analyzer Adapter to facilitate the impedance measurement. Technical details of this device will be given on the chapter 4.1.1.

The contact with the bananas was made through a two-electrode disposition, using circular pre–gelled ECG (Ag/AgCl) electrodes.

Connecting the AD2 to the electrodes, two copper wires were used. Alligator clips were welded on one end of the wire connected to the electrodes while on the other end, terminal wire-wire male-to-wire Raster Signal 2mm pins were welded into the wire and connected to the AD2's adapter. This wire setup was built and improved over time through trial and error to achieve the least amount of interference with the EIS signal. A calibration step was executed before each measurement, which will be explained in chapter 4.1.1.3.

Lastly, EIS data was measured from 10 store-bought bananas from the *Cavendish* group, from different hands. The bananas were bought as green (unripe) as possible. Then, 10 measurements of EIS were performed for each banana, daily, over a 32-day period. The input voltage was 1V and the frequency sweep ranged from 10Hz to 1MHz. Since the data was collected automatically by the AD2, it was possible for some errors to occur, so the 10-measurement acted as a fail-safe to not lose any data. The acquisition lasted 32 days since it was most fruitful

to collect as much EIS information possible of every ripening stage of the bananas. The resulting experimental setup can be observed in Figure 4.1.

Figure 4.1 — Experimental setup, with the AD2 connected to the banana's ECG electrodes

The classification of each sample will be explained in chapter 4.1.2.

4.1.1 Analog Discovery 2

Developed by Analog Devices, the Digilent Analog Discovery 2 is a multi-function instrument sometimes referred as a small laboratory that can fit in one's hand due to its size. Despite costing roughly 370 euros, it is considered cheap when compared to all the bench-top equipment needed for the same purposes this multi-tool can achieve. It allows the user to measure, visualize, generate, record, and control mixed signal circuits of all kinds. The device can be observed in Figure 4.2.

One of its features is the Impedance Analyzer function, used to analyse capacitive and inductive elements. It operates with the integrated waveform generator and oscilloscope, can perform a frequency sweep ranging from 100 μHz to 25 MHz and measure impedance, admittance, inductance, or capacitance.

To measure the impedance of an element, it's also necessary to add an external reference resistor. This resistor should be chosen according to a pre-defined table that relates to best reference resistance to the capacitance/inductance under measurement, or sample under measurement. This reference resistor is used by the AD2 to measure the voltage at the end of

the element by the oscilloscope. However, when measuring samples which inductive/capacitance values are not known, the choice of the reference resistor becomes more difficult. Besides this obstacle, the reference resistor also depends on the selected frequency sweep interval, since the SUT's impedance will also vary.

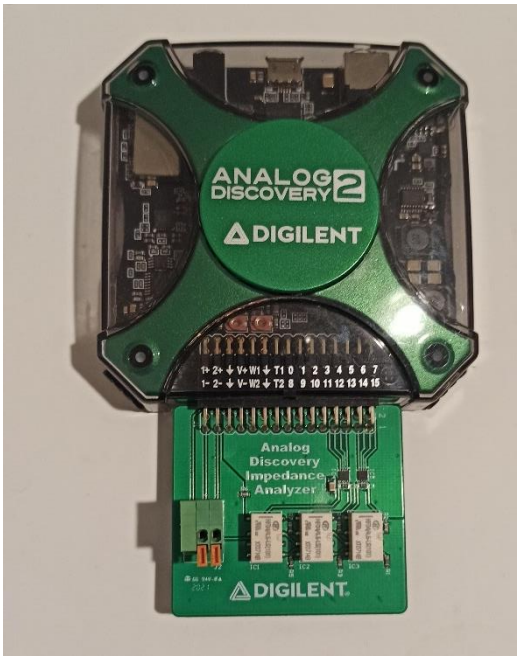

Figure 4.2 — Analog Discovery 2 with the Impedance Analyzer Adapter attached

Due to this problem, the Impedance Analyzer Adapter was used. This addon is coupled to the AD2 and has a two terminal block where pin wires can connect. This adapter allows autoscaling of the measured impedance since it disposes of a set of relays to automatically switch between the most adequate reference resistor. The adapter can be observed attached to the AD2 on Figure 4.2.

The software interface used to interact with the device was the WaveForms (beta version 3.19.11). Every information regarding the AD2 device and the WaveForms software comes from the respective reference manual.

4.1.1.1 Impedance Analyzer parameters

The AD2's Impedance Analyzer tool with the respective adapter could be configured in several ways. The parameters used are shown in Table 4.1 — .

Table 4.1 — Parameters of the AD2's impedance analyser tool

Parameter	Value
Start frequency	10Hz
Stop frequency	1MHz

Steps	151
/Decade	30
Topology	Adapter
Mode	CV Resistor
Voltage	1V
Offset	0V
Trace	|Z| (impedance)

The frequency sweep can be chosen differently depending on the test subject, the device properties, and the objective of the experiment. Usually, the reference resistor has a fixed value which limits the frequency sweep since the SUT's impedance will vary with the frequency. An inadequate frequency range can lead to a noisy response or to inaccurate impedance values (Ibba, 2020). For this dissertation, a wide range was chosen since the AD2 can adapt the reference resistor to adequate values for the banana's impedance variation in function of frequency.

The "Steps" parameter corresponds to the number of samples per acquisition and the "/Decade" is the division of those samples per decade on the plot plane. These were the automatically selected values by the software for chosen frequency sweep.

The selected "Topology" and "Mode" options referred to the use of the adapter.

The signal input voltage was set to 1V since not only was the standard value used in EIS experiments found in the literature, but also the voltage that incurred in a less noisy response. This comparison can be seen in Figure 4.3. The yellow continuous line represents the impedance (|Z|). A lower voltage would result in a lack of conduction through the sample, preventing the correct impedance measurement.

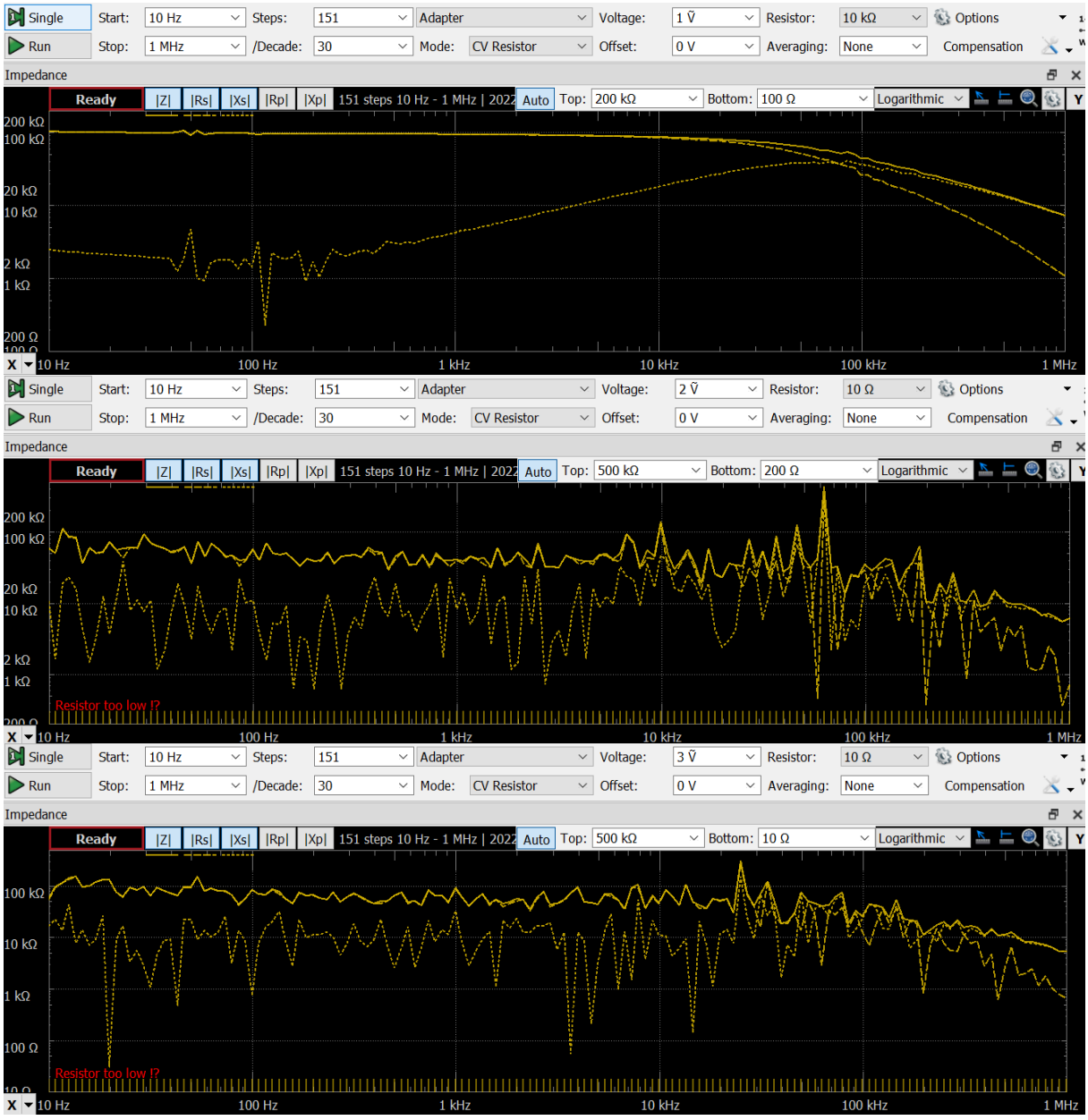

Figure 4.3 — Impedance behaviour of the same sample for the input's voltage of 1V, 2V and 3V (from top to bottom)

4.1.1.2 WaveForms Script

The WaveForms software provides a Script function where it's possible to program the AD2 activity. One could manually select the desired parameters in the Impedance tool, and then activate it through a script.

This was used to loop 10 measurements on the same sample, without touching the setup. For each iteration, the collected data was also exported to a .csv file.

4.1.1.3 AD2 calibration

Despite the AD2 capabilities, there is always an accuracy subjectivity. This could be due to faulty contact with the adapter's terminals, alligator clipper-electrode contact or even the adapters auto-scaling assumption. However, for the completion of the purposed objectives, precision is more important than accuracy for the success of an intelligent algorithm.

To ensure the AD2 was calibrated between data collections, an impedance measurement was performed on an electrolytic 1µF capacitor. After conferring the response, the device was considered calibrated, and the EIS experiments proceeded. This also allowed an overview on the accuracy error of the AD2, as seen in Figure 4.4.

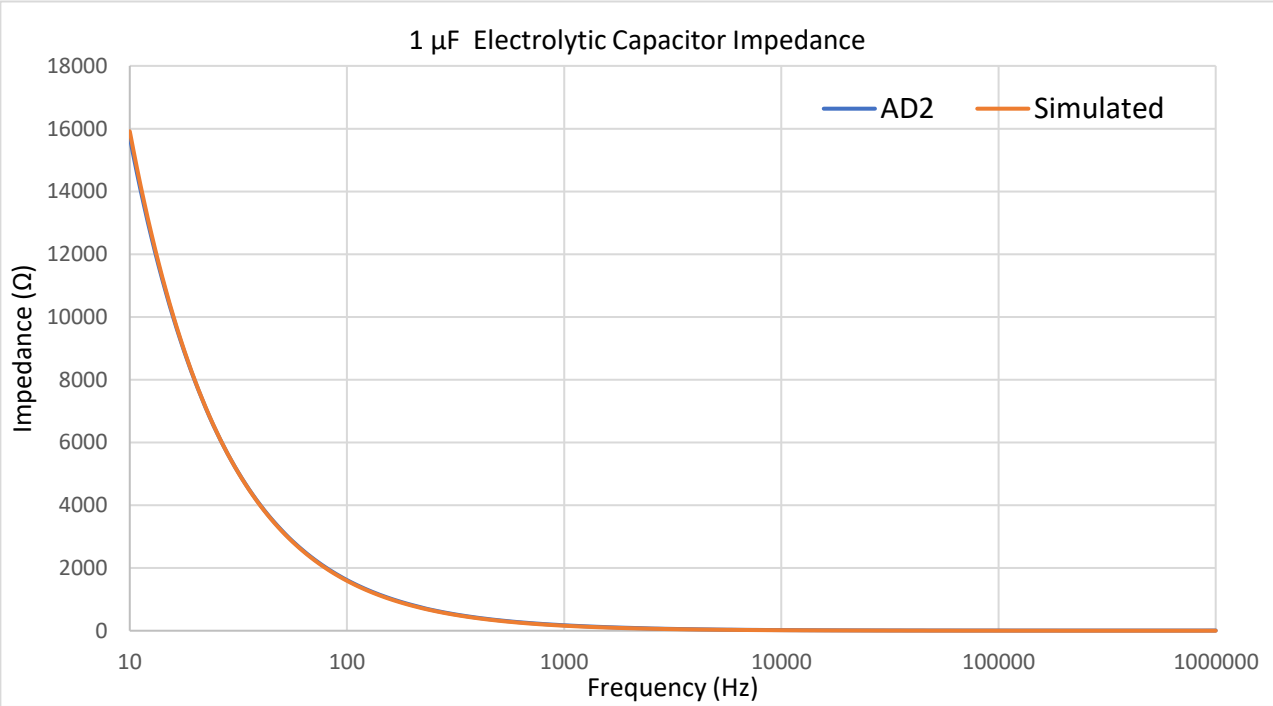

Figure 4.4 — Comparison of an IS experiment on a 1µF capacitor and the respective simulated value

As observed, there are no significant differences between the measured and the simulated values. Furthermore, a comparison was made between the banana's impedance measurement performed by the AD2 setup and by a bench-top lab equipment, using a GAMRY Instruments Potentiostat (Reference 3000). The result of this comparison can be observed in Figure 4.6. and the experiment's setup is shown in Figure 4.5. Despite the discrepancy, the

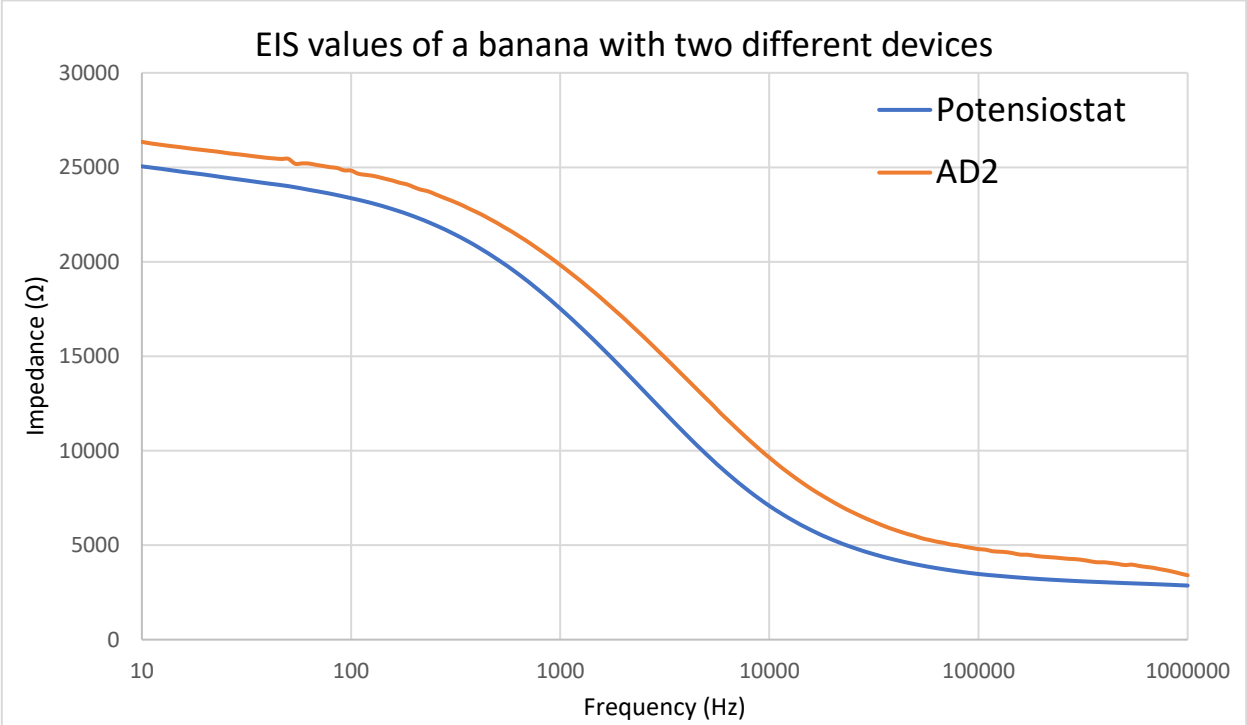

Figure 4.6 — Comparison of an EIS experiment on a banana with the AD2 device and the Potentionstat

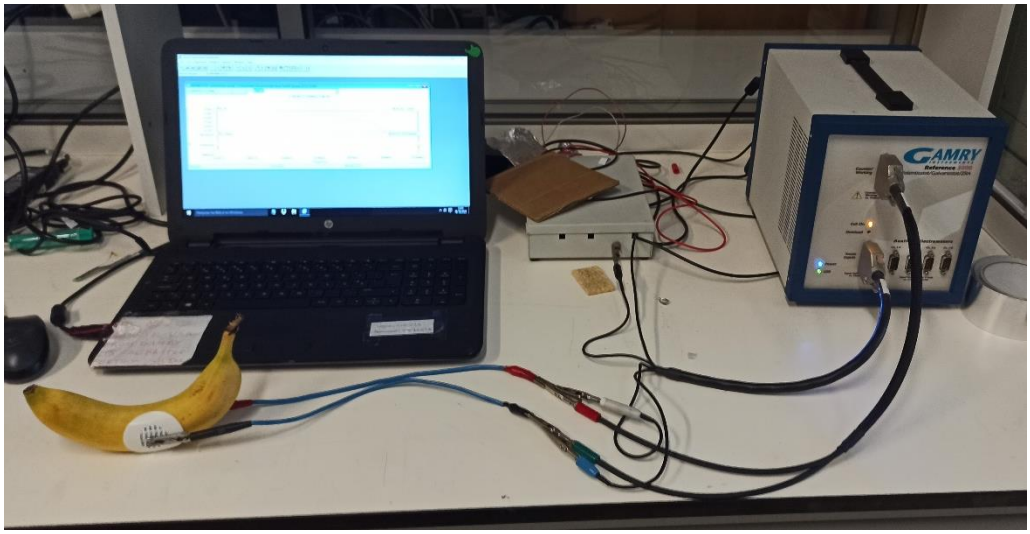

Figure 4.5 — EIS experiment on a banana with the Potentionstat

difference between the impedances is constant over the frequency sweep, which is probably due to the reference resistor auto-scaling differences.

4.1.2 Classification

To be possible to apply a supervised learning approach to the data, a classification of each banana for each day of the data collection was necessary. The classification was based on a binary system, to classify whether the banana was good for consumption (edible) or not (non-

edible). Besides this binary system, a subclass could be used according to the different ripening stages of the sample.

This classification was performed empirically by observing the peel colour of the banana and comparing it to a scaled provided by Saranya et. al (2022) experiment, as seen in Figure 4.7. Unripe (green) and over ripe (brown) bananas were considered non-edible while both ripe stages (yellow and yellow with spots) were considered edible. In some cases, the bananas didn't mature in a homogeneous form, meaning that some parts of the banana could have greenish tones, while others were already rather yellow. To overcome this issue, only the skin colour near the electrodes was considered for classification.

Figure 4.7 — Classification metric used (adapted from Saranya et al., 2022)

Table 4.2 summarizes this classification attribution.

Table 4.2 — Banana classification and its ripening stage and colour

Ripening stage Colour	Class
Unripe Green / Yellowish green	Non-edible
Ripe Yellow/Yellow with dark spots	Edible
Over ripe Brown	Non-edible

4.1.3 Temperature and Humidity

The experiment was performed at room temperature in a non-controlled environment. Each day, temperature and humidity values were measured, in Celsius degrees ($°C$) and percentage, respectively.

The measurements were executed with a DHT22 sensor connected to an Arduino Due and saved into a Microsoft Excel file. It was assumed the sensor was correctly calibrated. The sensor was left near the banana's area throughout the experiment period, so it was also considered not necessary an adjustment period before registering the room temperature and humidity values. Figure 4.8 shows both the Arduino device and the DHT22 sensor.

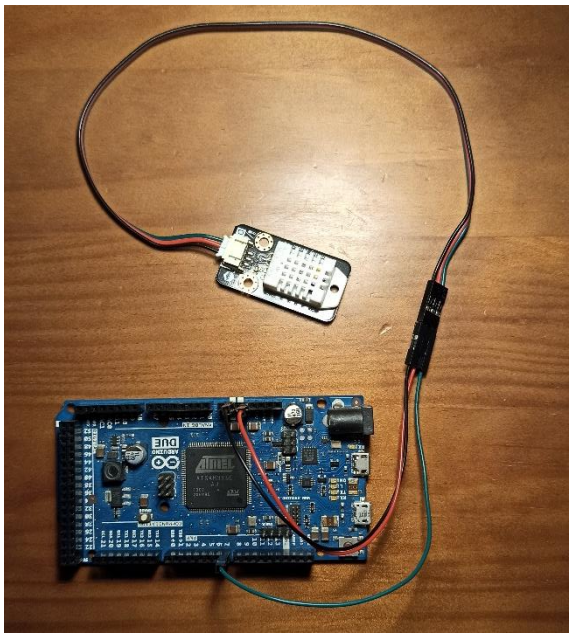

Figure 4.8 — DHT22 sensor connected to the Arduino Due device

4.2 Machine Learning approach

The ML approach to achieve this dissertation's proposed objectives followed a model evaluation structure. This means that after pre-processing and splitting the data, different models were trained, tested and assessed based on their performance. The same approach was used for feature selection, where available features were permuted, always including the impedance values. Furthermore, these features were normalised using different techniques and the best model's performance was evaluated with and without normalised data.

Table 4.3 and Table 4.4 present the different classification models and the normalization methods used, respectively, and the corresponding libraries.

Table 4.3 — Classification models

Name	Library
SVM	Scikit-learn
DT	
MLP	
AdaBoost	
LG	
GB	
NB	
XGB	XGB

Table 4.4 — Normalization methods

Name	Library
MinMaxScaler	Scikit-learn
StandardScaler	
Normalizer	
RobustScaler	
L1 normalization	NumPy
L2 normalization	

Table 4.5 shows all the used feature combinations. In this table, "I" corresponds to the banana impedance values, "T" and "H" to the corresponding room temperature and room humidity values, respectively, and "D" corresponds to the number of days passed since the beginning of the measurements.

Table 4.5 — Feature permutations

Label	Feature
1	I
2	I + T
3	I + T + H
4	I + T + H + D
5	I + H
6	I + D

7	I + T + D
8	I + H + D

After training and testing all the models, with all the different feature combinations, with all the different types of data normalization, the model with the best accuracy was selected and tuned, using the Grid Search tool from the Scikit-learn library.

The ML application was executed using Python coding language (version 3.8.10), on the Jupyter Notebook platform. Data was also manipulated using Microsoft Excel.

4.2.1 Data Pre-processing

After the 32-day data collection, graphs containing all the measurements per day of each sample were plotted separately. Ideally, all 10 EIS sweeps should be equal. However, in these plots some outliers were found. Figure 4.9 depicts this example fairly, where peaks can be observed for example, in data frames represented by colours brown, green, blue and red.

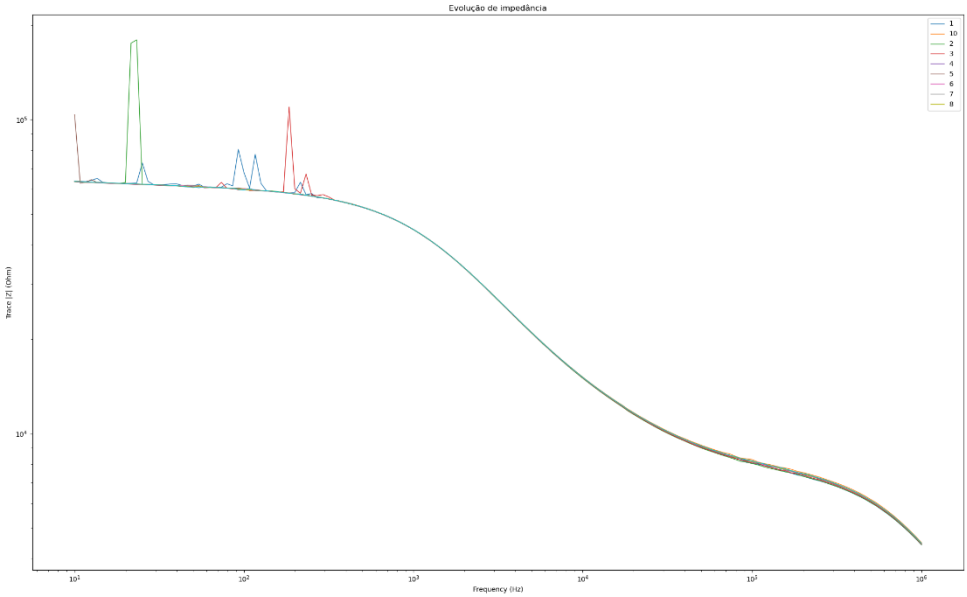

Figure 4.9 — 10 impedance data frames of the same sample with peaks

It was also possible to observe some cases without relevant peaks, but where the impedance measurements were rather different between each other. Sometimes, due to poor contact between the alligator-clips and the ECG electrodes, the impedance would reach values way above 1MΩ and the AD2's reference resistor would be fixed on the 1MΩ value.

To counter these problems, an algorithm was created to detect in each data frame every local max and the next minimum low. If the difference between these two points was over a predefined threshold, the whole data frame would be considered an outlier and would be

excluded. Besides this mechanism, it was also implemented that the impedance measured values shouldn't be above 1MΩ.

After removing the faulty data frames, the median of the remaining data frames was selected and exported as the data frame representing the impedance values of the sample for that day. Then, new plots were created with the selected data frames for each day, representing the banana's impedance evolution over a course of 32 days, on the chosen frequency sweep. These plots will be presented in chapter 5.1.

Data was split into training and test sets, considering a proportion of 67% and 33%, respectively.

4.2.2 Reporting

After training and testing, the performance of each scenario was exported using the Scikit-learn reporting tools. The F1-score, accuracy, precision, recall and confusion matrix were exported into Microsoft Excel files, divided by feature combination. Table 4.6 and Table 4.7 show an example of these reports.

Table 4.6 — DT model report without normalization for "I" feature

	Precision	Recall	F1-score	Support
Class 0	0,860	0,961	0,907	51
Class 1	0,958	0,852	0,902	54
Accuracy			0,905	0,905
Macro average	0,909	0,906	0,905	105
Weighted average	0,910	0,905	0,905	105

Table 4.7 — DT model confusion matrix without normalization for "I" feature

	Predicted Negative - 0	Predicted Positive - 1
Actual Negative - 0	49	8
Actual Positive - 1	2	46

To compare reports, only the F1-score and accuracy metrics were considered, since they're the ones which provide a general idea of a ML classifier performance. These metrics of each scenario were gathered into a single Excel file, in which the best performances were selected. In section 5.2, these tables will be presented and discussed.

5

Results and Discussion

This chapter presents all the results obtained during this thesis. In section 5.1, the results of the 32-day impedance measurement will be shown. Then, section 5.2 will focus on the results of every machine learning test. Moreover, a detailed analysis will be given on the scenarios with the best performances from the comparative study.

5.1 Banana impedance evolution

After pre-processing the data as described in section 4.2.1, the banana impedance evolution over the course of the 32-day experiment was plotted for each banana. Some of the produced graphs can be seen from Figure 5.1 to Figure 5.5. The rest of the plots can be observed in Appendix A.

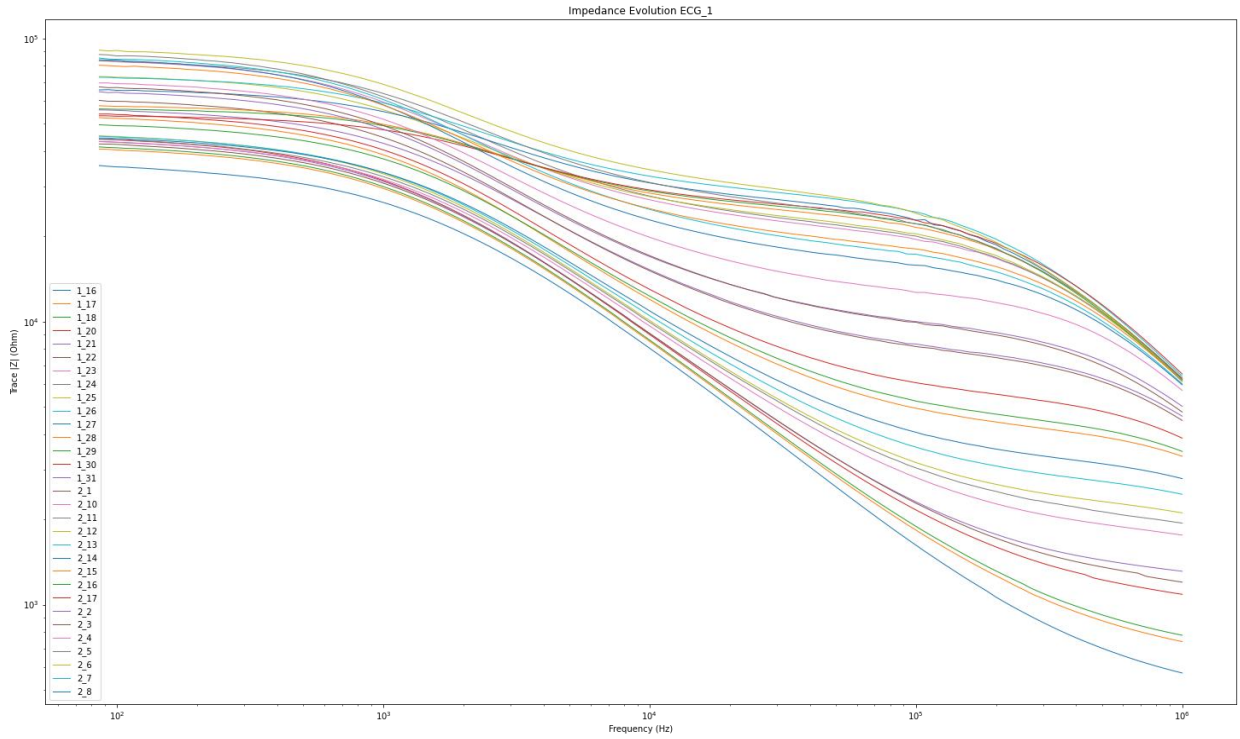

Figure 5.1 — Impedance evolution for sample 1

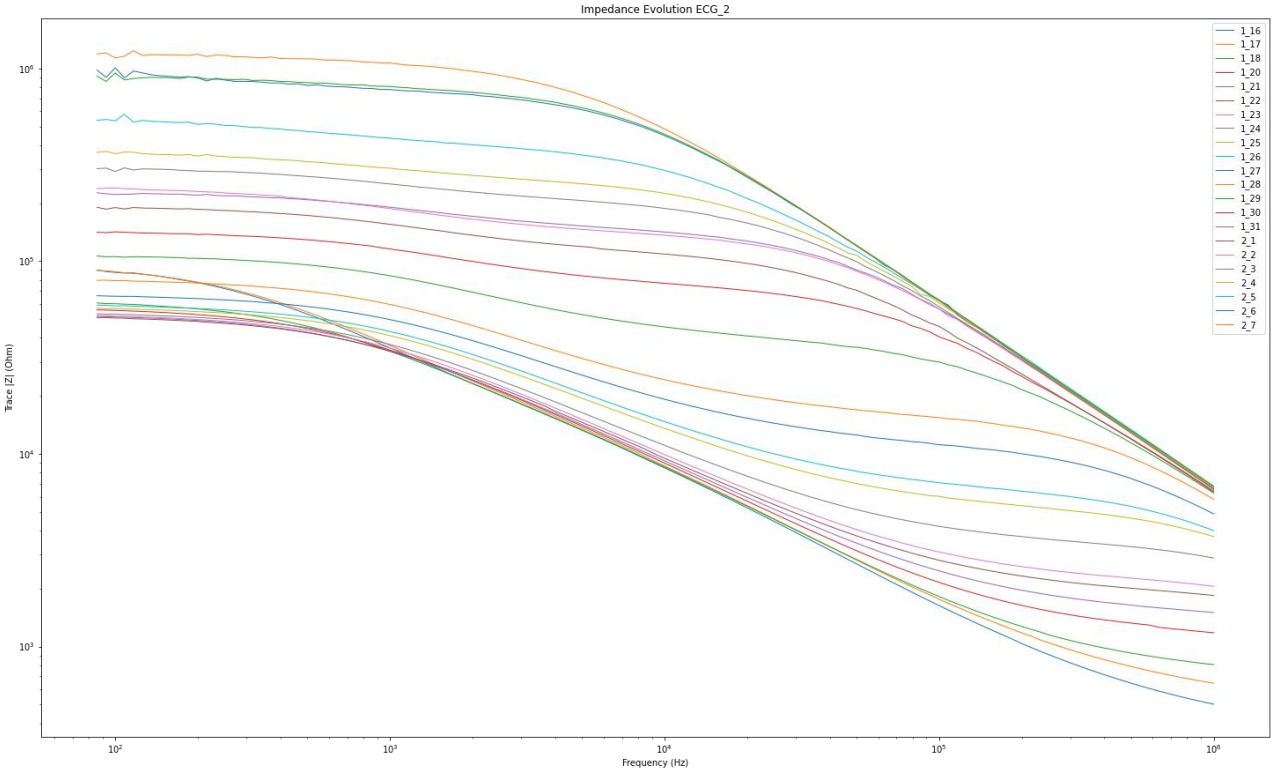

Figure 5.2 — Impedance evolution for sample 2

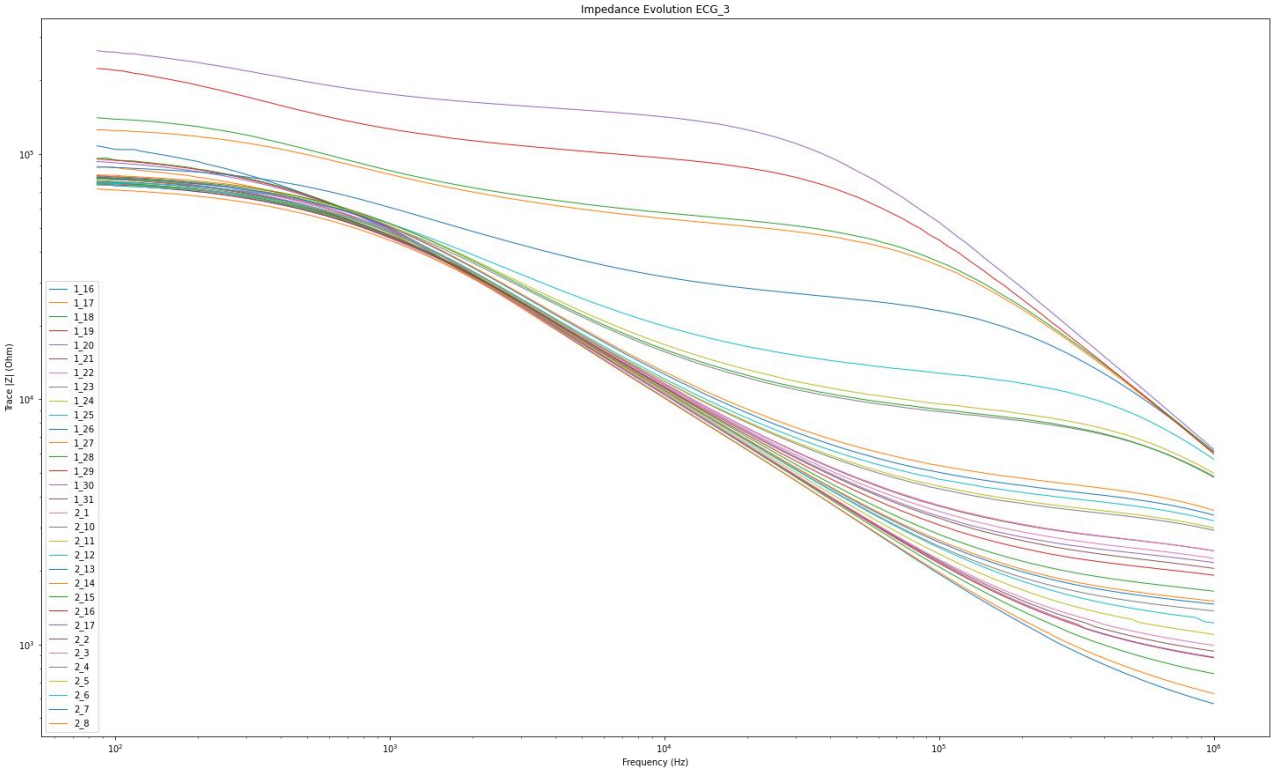

Figure 5.3 — Impedance evolution for sample 3

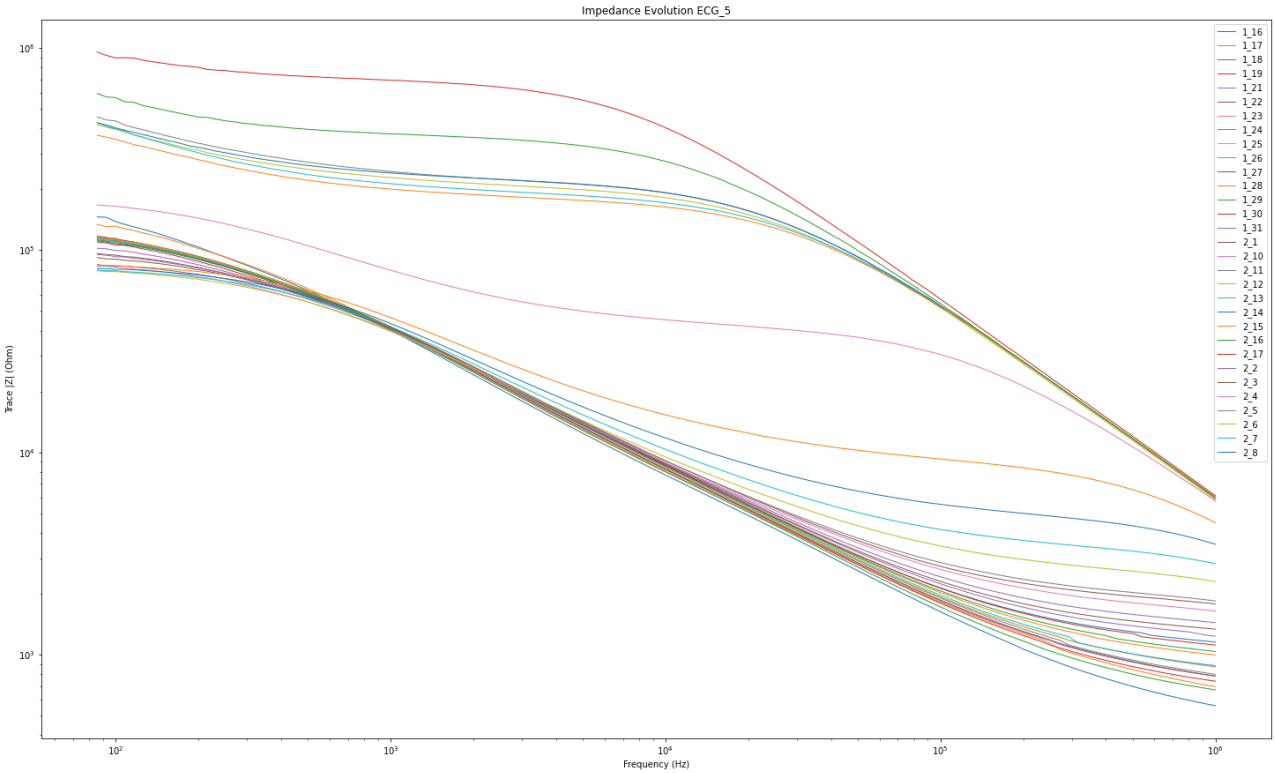

Figure 5.4 — Impedance evolution for sample 5

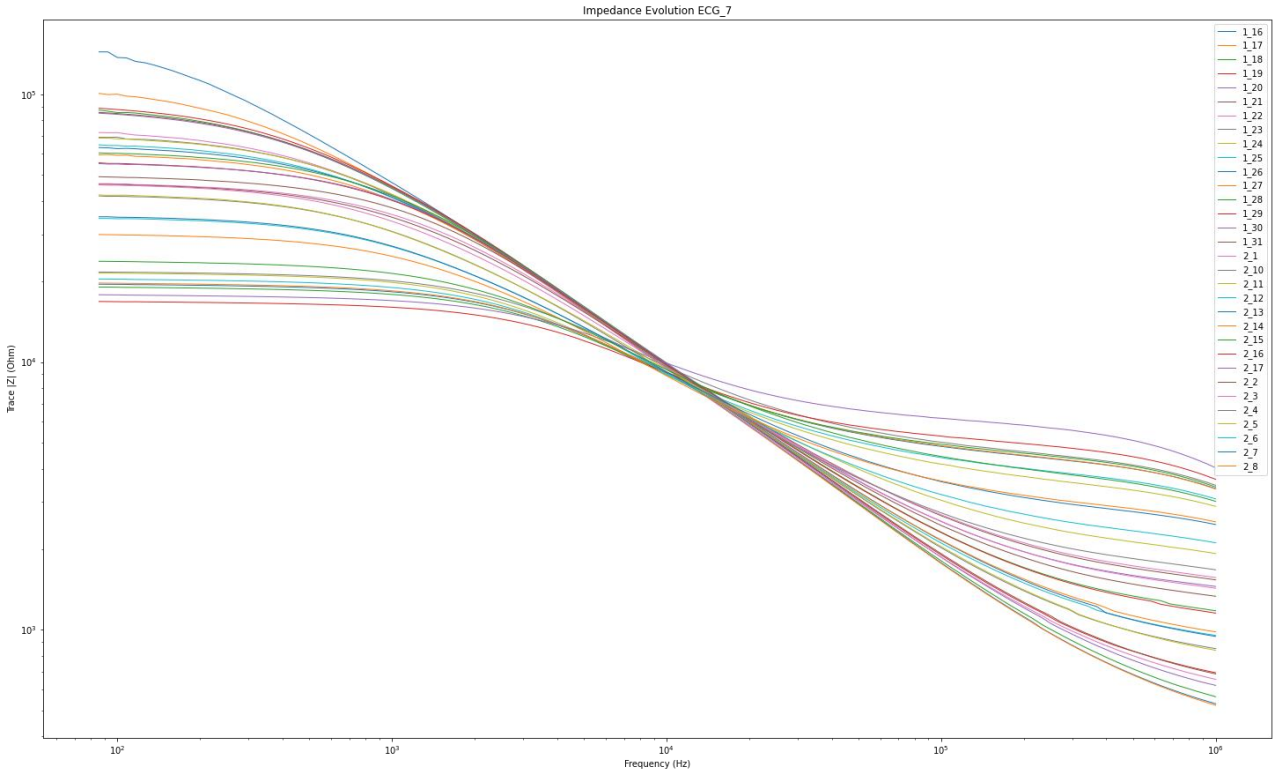

Figure 5.5 — Impedance evolution for sample 7

In each figure, each line corresponds to a data frame of the impedance collected in one day. The legend represents each selected data frame by colour, described by day and month in a "MONTH_DAY" format. As observed, in some cases there are less than 32 data frames as should be expected. This indicates that in some days, all the 10 data collections contained errors and were eliminated by the pre-processing stage. Nonetheless, globally a good amount of data was collected to represent the banana's ripeness evolution in terms of impedance.

Impedance varied in function of the input's frequency in accordance with equation 2.11. Simply put, for lower frequencies impedance was higher, and decrease as the frequency sweep proceeded to higher frequencies.

Despite being tested under the same conditions, there are some differences between the samples. It is possible to note a discrepancy within the data frames on Figure 5.2, Figure 5.3 or Figure 5.4. This was due to the AD2's selection of the reference resistor. It was most noticeable on sample 5, in which during a large portion of the data collection period, the chosen reference resistor was higher, thus the impedance measured was also higher. It's also possible to notice a difference in each sample's impedance evolution. Figure 5.5 presents a "S"-type curve during the entirety of the data collection, whereas Figure 5.1 only presents this behaviour during the first days of the experiment. This could be due to the sample's different physiological characteristics such as size or girth, and the fact that these samples came from different banana hands.

It is also relevant to note that the impedance values measured between samples were rather different. Table 5.1 shows a short example of the impedances measured in three different days, for three different frequencies. Despite these discrepancies, the bananas matured in a similar way, as shown in Figure 5.6 for samples 1 and 9.

Table 5.1 — Short example of impedance values for specific input frequencies

		Day 1			Day 15			Day 30		
		10 Hz	10 kHz	1 MHz	10 Hz	10 kHz	1 MHz	10 Hz	10 kHz	1 MHz
Sample	1	38689.17	8047.81	573.29	58634,82	15243.02	4629.28	59543.96	28412.14	6307.34
	3	127030.59	10152.55	572.79	84698.84	11261.11	2041.59	150513.20	57656.15	6105.46
	5	193647.29	7735.44	559.61	115670.61	8752.98	1234.57	494690.19	163675.11	6016.79
	7	172043.88	9491.02	530.44	50980.93	9815.98	1335.73	19582.64	9114.85	3417.95
	9	72292.58	9482.22	512.27	50943.09	10076.31	1440.14	26931.89	12650.29	4522.94

Through this short example it is possible to find some correlation between samples for some frequencies, mainly for higher frequency values. However, rather large discrepancies can also be found. Overall, it is possible to observe a tendency of the increase of impedance over time, especially for higher frequencies, even though for some cases, the opposite happened for lower frequencies. It is even possible to find a decrease followed by an increase in impedance, as seen for the 10Hz input in sample 5. These tendencies go against the experiences referred in section 3.1.4. These factors present a challenge for the machine learning models to overcome.

Figure 5.6 — Samples 1 and 9 in three different moments of the experiment, showing different ripening stages

5.2 Machine learning results

The ML part of this dissertation followed the steps described in section 4.2. Through loops, all possibilities of features, models and types of normalization were tested and reported. The F1-score and accuracy metrics were used to evaluate the success of the model. Each report was gathered into tables, differentiated by feature combination, as presented from Table 5.3 to Table 5.10 in subsection 5.2.1.

From these tables, it is possible to observe a wide variation for each model performance. Overall, SVM, NB, NN and LG behaved poorly, with accuracies ranging from 40% to 70%. On the other hand, Decision Tree based algorithms such as DT, GB, XGB and AdaBoost performed better, reaching accuracies and F1-scores above 90%.

Normalization didn't always improve the score of the model's test. In some cases, normalization decreased the performance. In fact, for features combination "I", "I+T+H" and "I+H+D", the best F1-scores and accuracies resulted in the tests without normalization, with percentages above or equal to 95%.

From each table, the best performance was selected and summarized in Table 5.2.

Table 5.2 — Best ML performance by feature combination

Feature combination	**I**	**I + T**	**I + T + H**	**I + T+ H + D**	**I + H**	**I + D**	**I + T + D**	**I + H + D**
ML Model	XGB	XGB	XGB	GB	GB	XGB	XGB	XGB
Normalization method	None	Sklearn Normalizer	None	Sklearn StandardScaler	Sklearn MinMaxScaler	Sklearn Normalizer	Sklearn MinMaxScaler	None
F1-score	95,33%	96,43%	98,36%	94,00%	92,31%	93,22%	94,83%	95,41%
Accuracy	95,24%	96,19%	98,10%	94,29%	91,43%	92,38%	94,29%	95,24%

As highlighted in Table 5.2, the best model to predict whether a banana is good for consumption or not is the XGB model, without data normalization, using the features Impedance, Temperature and the Number of days since the beginning of the experiment (or in other words, how many days since the banana was bought).

5.2.1 Performance tables

This subsection presents the F1-score and accuracy of every ML model tested, with different types of data normalizations, for every feature combination. In each table, the best scores are highlighted in green. The best score was considered the model report with both the highest F1-score and accuracy.

Table 5.3 — Performance table for feature "Impedance"

Feature: I

	None		Sklearn MinMaxScaler		Sklearn StandardScaler		Sklearn Normalizer		Sklearn RobustScaler		NumPy L1 normalization		NumPy L2 normalization	
	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy
SVM	0,600	0,467	0,695	0,657	0,604	0,600	0,696	0,533	0,696	0,533	0,522	0,581	0,525	0,552
DT	0,902	0,905	0,893	0,886	0,912	0,905	0,803	0,781	0,889	0,886	0,814	0,790	0,889	0,867
NB	0,152	0,467	0,185	0,495	0,410	0,562	0,563	0,600	0,368	0,543	0,557	0,590	0,544	0,552
GB	0,927	0,924	0,911	0,905	0,887	0,876	0,875	0,867	0,907	0,905	0,893	0,876	0,897	0,886
NN	0,582	0,438	0,810	0,790	0,774	0,733	0,696	0,533	0,752	0,724	0,704	0,543	0,584	0,552
LG	0,526	0,571	0,661	0,638	0,690	0,667	0,696	0,533	0,732	0,714	0,704	0,543	0,618	0,552
XGB	0,953	0,952	0,936	0,933	0,879	0,867	0,887	0,876	0,909	0,905	0,904	0,895	0,879	0,867
ADA	0,887	0,886	0,907	0,905	0,844	0,838	0,885	0,876	0,877	0,867	0,845	0,829	0,836	0,829

Table 5.4 — Performance table for feature "Impedance + Temperature"

Feature: I + T

	None		Sklearn MinMaxScaler		Sklearn StandardScaler		Sklearn Normalizer		Sklearn RobustScaler		NumPy L1 normalization		NumPy L2 normalization	
	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy
SVM	0,654	0,486	0,724	0,695	0,687	0,705	0,696	0,533	0,700	0,543	0,671	0,505	0,671	0,505
DT	0,907	0,905	0,929	0,924	0,839	0,829	0,885	0,876	0,912	0,905	0,844	0,838	0,897	0,886
NB	0,138	0,524	0,319	0,552	0,131	0,495	0,583	0,619	0,197	0,457	0,611	0,648	0,566	0,562
GB	0,942	0,943	0,899	0,895	0,860	0,848	0,922	0,914	0,914	0,905	0,897	0,895	0,887	0,876
NN	0,258	0,562	0,860	0,857	0,780	0,790	0,696	0,533	0,823	0,790	0,671	0,505	0,515	0,552
LG	0,637	0,686	0,642	0,638	0,673	0,686	0,696	0,533	0,748	0,705	0,671	0,505	0,672	0,610
XGB	0,962	0,962	0,899	0,895	0,850	0,838	0,964	0,962	0,898	0,886	0,863	0,867	0,889	0,876
ADA	0,936	0,933	0,873	0,867	0,847	0,838	0,915	0,905	0,891	0,876	0,811	0,810	0,852	0,829

Table 5.5 — Performance table for feature "Impedance + Temperature + Humidity"

Feature: I + T + H

	None		Sklearn MinMaxScaler		Sklearn StandardScaler		Sklearn Normalizer		Sklearn RobustScaler		NumPy L1 normalization		NumPy L2 normalization	
	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy
SVM	0,579	0,514	0,746	0,695	0,655	0,629	0,542	0,581	0,748	0,629	0,533	0,600	0,533	0,600
DT	0,917	0,905	0,935	0,933	0,915	0,895	0,889	0,867	0,855	0,829	0,852	0,838	0,852	0,829
NB	0,662	0,514	0,215	0,514	0,455	0,543	0,610	0,610	0,381	0,505	0,522	0,581	0,618	0,600
GB	0,967	0,962	0,917	0,914	0,929	0,914	0,896	0,876	0,912	0,895	0,903	0,895	0,881	0,867
NN	0,735	0,581	0,775	0,762	0,871	0,848	0,495	0,552	0,789	0,771	0,688	0,524	0,602	0,571
LG	0,524	0,533	0,758	0,714	0,733	0,695	0,709	0,648	0,758	0,714	0,688	0,524	0,586	0,543
XGB	0,984	0,981	0,936	0,933	0,937	0,924	0,873	0,848	0,896	0,876	0,904	0,895	0,881	0,867
ADA	0,933	0,924	0,909	0,905	0,884	0,857	0,836	0,810	0,898	0,876	0,845	0,829	0,847	0,829

Table 5.6 — Performance table for feature "Impedance + Temperature + Humidity + No. of Days"

Feature: I + T + H + D

	None		Sklearn MinMaxScaler		Sklearn StandardScaler		Sklearn Normalizer		Sklearn RobustScaler		NumPy L1 normalization		NumPy L2 normalization	
	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy
SVM	0,437	0,533	0,737	0,714	0,672	0,629	0,636	0,467	0,662	0,495	0,645	0,476	0,645	0,476
DT	0,887	0,867	0,879	0,876	0,913	0,914	0,838	0,838	0,873	0,867	0,826	0,819	0,787	0,743
NB	0,111	0,390	0,641	0,476	0,375	0,619	0,615	0,667	0,632	0,467	0,524	0,619	0,533	0,533
GB	0,860	0,838	0,816	0,819	0,940	0,943	0,904	0,905	0,926	0,924	0,870	0,867	0,857	0,838
NN	0,750	0,600	0,808	0,819	0,811	0,810	0,598	0,590	0,772	0,752	0,645	0,476	0,500	0,524
LG	0,505	0,533	0,708	0,686	0,741	0,733	0,636	0,467	0,789	0,771	0,645	0,476	0,617	0,514
XGB	0,876	0,857	0,827	0,829	0,902	0,905	0,889	0,886	0,925	0,924	0,838	0,838	0,874	0,857
ADA	0,831	0,810	0,852	0,848	0,879	0,876	0,895	0,895	0,897	0,895	0,862	0,857	0,797	0,762

Table 5.7 — Performance table for feature "Impedance + Humidity"

Feature: I + H

	None		Sklearn MinMaxScaler		Sklearn StandardScaler		Sklearn Normalizer		Sklearn RobustScaler		NumPy L1 normalization		NumPy L2 normalization	
	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy
SVM	0,662	0,495	0,788	0,790	0,796	0,800	0,627	0,457	0,658	0,495	0,532	0,581	0,688	0,524
DT	0,868	0,867	0,904	0,895	0,904	0,895	0,796	0,819	0,891	0,876	0,857	0,857	0,849	0,848
NB	0,188	0,505	0,373	0,552	0,405	0,581	0,565	0,648	0,416	0,571	0,583	0,619	0,621	0,629
GB	0,911	0,914	0,923	0,914	0,920	0,914	0,851	0,867	0,893	0,876	0,873	0,867	0,881	0,876
NN	0,662	0,495	0,899	0,895	0,852	0,848	0,480	0,505	0,771	0,762	0,688	0,524	0,688	0,524
LG	0,602	0,648	0,702	0,676	0,704	0,695	0,612	0,457	0,660	0,657	0,688	0,524	0,649	0,505
XGB	0,911	0,914	0,914	0,905	0,920	0,914	0,809	0,829	0,883	0,867	0,891	0,886	0,906	0,905
ADA	0,904	0,905	0,874	0,857	0,912	0,905	0,788	0,800	0,867	0,848	0,830	0,829	0,860	0,857

Table 5.8 — Performance table for feature "Impedance + No. of Days"

Feature: I + D

	None		Sklearn MinMaxScaler		Sklearn StandardScaler		Sklearn Normalizer		Sklearn RobustScaler		NumPy L1 normalization		NumPy L2 normalization	
	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy
SVM	0,696	0,533	0,775	0,762	0,825	0,810	0,522	0,581	0,247	0,419	0,671	0,505	0,671	0,505
DT	0,920	0,914	0,880	0,857	0,881	0,876	0,869	0,848	0,887	0,857	0,891	0,886	0,847	0,829
NB	0,324	0,524	0,602	0,648	0,619	0,695	0,586	0,610	0,569	0,581	0,517	0,590	0,606	0,590
GB	0,873	0,867	0,884	0,857	0,857	0,848	0,900	0,886	0,910	0,886	0,925	0,924	0,860	0,848
NN	0,712	0,552	0,882	0,857	0,827	0,829	0,531	0,562	0,837	0,800	0,671	0,505	0,580	0,600
LG	0,495	0,552	0,705	0,657	0,763	0,733	0,593	0,581	0,750	0,714	0,671	0,505	0,683	0,619
XGB	0,870	0,867	0,896	0,876	0,883	0,876	0,932	0,924	0,919	0,895	0,862	0,857	0,881	0,867
ADA	0,836	0,829	0,900	0,886	0,897	0,895	0,918	0,905	0,917	0,895	0,874	0,876	0,832	0,819

Table 5.9 — Performance table for feature "Impedance + Temperature + No. of Days"

Feature: I + T + D

	None		Sklearn MinMaxScaler		Sklearn StandardScaler		Sklearn Normalizer		Sklearn RobustScaler		NumPy L1 normalization		NumPy L2 normalization	
	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy
SVM	0,658	0,495	0,736	0,733	0,736	0,733	0,630	0,619	0,743	0,590	0,579	0,514	0,579	0,514
DT	0,914	0,905	0,897	0,886	0,860	0,857	0,850	0,838	0,847	0,829	0,826	0,819	0,844	0,838
NB	0,161	0,505	0,395	0,533	0,541	0,629	0,606	0,629	0,562	0,629	0,604	0,638	0,574	0,619
GB	0,906	0,895	0,904	0,895	0,870	0,867	0,844	0,838	0,906	0,895	0,891	0,886	0,904	0,905
NN	0,472	0,362	0,860	0,848	0,780	0,790	0,583	0,619	0,790	0,762	0,679	0,514	0,688	0,524
LG	0,642	0,638	0,782	0,771	0,620	0,638	0,662	0,533	0,743	0,724	0,679	0,514	0,688	0,524
XGB	0,922	0,914	0,948	0,943	0,870	0,867	0,815	0,810	0,924	0,914	0,909	0,905	0,893	0,895
ADA	0,904	0,895	0,887	0,876	0,811	0,810	0,804	0,800	0,857	0,838	0,857	0,857	0,885	0,886

Table 5.10 — Performance table for feature "Impedance + Humidity + No. of Days"

Feature: I + H + D

	None		Sklearn MinMaxScaler		Sklearn StandardScaler		Sklearn Normalizer		Sklearn RobustScaler		NumPy L1 normalization		NumPy L2 normalization	
	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy
SVM	0,645	0,486	0,717	0,714	0,734	0,724	0,468	0,524	0,662	0,495	0,500	0,562	0,500	0,562
DT	0,885	0,886	0,874	0,857	0,867	0,848	0,917	0,914	0,907	0,905	0,793	0,781	0,902	0,886
NB	0,188	0,505	0,243	0,467	0,213	0,438	0,490	0,524	0,333	0,543	0,527	0,590	0,583	0,590
GB	0,936	0,933	0,877	0,867	0,898	0,876	0,917	0,914	0,923	0,924	0,891	0,886	0,898	0,876
NN	0,179	0,476	0,843	0,819	0,813	0,781	0,495	0,533	0,796	0,810	0,688	0,524	0,574	0,619
LG	0,708	0,733	0,673	0,657	0,685	0,667	0,585	0,514	0,789	0,771	0,688	0,524	0,643	0,610
XGB	0,954	0,952	0,881	0,867	0,935	0,924	0,885	0,876	0,895	0,895	0,875	0,867	0,898	0,876
ADA	0,887	0,886	0,800	0,781	0,831	0,810	0,857	0,838	0,902	0,905	0,830	0,829	0,871	0,848

6

Conclusion

This dissertation aimed to find an intelligent, quick, and accurate way to evaluate fruit quality, more precisely, whether a banana is good for consumption or not, to aid one's decision to better manage this fruit and prevent spoilage. Given the efficient fruit management, waste prevention and sustainability goals, a non-destructive method of quality analysis was studied: Electrochemical Impedance Spectroscopy.

Since to the date of this dissertation's completion there weren't any fruit impedance datasets available, an experimental set up was built and impedance data was acquired, as described in 4.1. Thus, new impedance data was collected from 10 bananas of the same species, over a 32-day period. Each day, it was also collected room temperature and humidity data, plus the number of days of maturation was registered. Each sample was classified through a metric from the literature, based on the banana's peel colour. After a pre-processing step to exclude outliers, these features composed the dataset to be used. The resulting impedance evolution data of the 10 bananas showed many behaviours, with a general tendency to an increase of impedance with ripening. Although rare, it was also possible to notice some flaws in the selected data frames, which decreased the dataset quality.

Then, in order to reach the proposed objective, an extensive comparative machine learning classification study was conducted, training and testing 8 different ML models, for different feature combinations and using different types of data normalisation. Despite the many behaviours observed in the samples' impedance evolution, it was possible to achieve model classifications with high performances. Generally, models based on decision trees, such as DT, GB, XGB, and ADA performed better, with F1-scores and accuracies above 90%. On the other hand, SVM, LG, NB and NN classifiers performed poorer, with F1-scores and accuracies ranging from

40% and 70%. It wasn't noticed any tendency with the types of data normalisation used nor with the feature selection combinations.

The model which achieved the highest performance was the Extreme Gradient Boosting (XGB), without data normalization, with the features "Impedance + Temperature + No. of days", with a F1-score of 98,36% and accuracy of 98,10%.

Despite the good results achieved, this dissertation's research has some flaws to be improved. Generally, the hardware set up brought difficulties for the EIS measurements, mainly on the contact between the alligator clips and the ECG electrodes. Consequently, this problem led to an unstable behaviour of the Impedance Analyzer Adapter, resulting on impedance values with discrepancies, on the same sample's impedance evolution and between samples. The affected data frames should be considered as outliers by the pre-processing step, which then needs to be reviewed.

Nevertheless, this dissertation produced a complete dataset, a first known-to-date use of the AD2 for fruit assessment, a paper regarding this research, and an extensive comparative study between ML classifiers which resulted on a possible smart solution to be used on the quality assessment of bananas from the *Cavendish* group.

This study is another step forward to the application of artificial intelligence and non-destructive methods towards a more sustainable future.

6.1 Future work

This dissertation is the base of many projects that can be developed in this field, including an improvement of the dissertation itself. The proposed next steps are:

- Improvement of the quality of the EIS data acquired. It is necessary to create a more robust dataset, with more impedance data, not only from the Cavendish group, but from other banana types to universalize the proposed solution;
- Improvement of the comparative study, using different and more adequate Machine Learning classifiers for the volume and characteristics of the dataset;
- Tuning of the ML models for a more optimal solution;
- This study can be extrapolated to other fruits and vegetables. It is only needed to collect the corresponding impedance data over a maturation cycle;
- This study can be extrapolated to other uses besides fruit quality. EIS data can be used to assess fruit bruising, viruses and parasites, as well as biochemical composition;

- Development of a simple, low-cost, hand-held customised hardware focused on collecting EIS data, as well as other information relevant for fruit quality assessment. The hardware could have in its software the whole ML algorithm, or simply connect to an outside device, such as a smartphone.

BIBLIOGRAPHY

Abasi, S., Minaei, S., Jamshidi, B., & Fathi, D. (2018). Dedicated non-destructive devices for food quality measurement: A review. *Trends in Food Science and Technology*, *78*(May), 197–205. https://doi.org/10.1016/j.tifs.2018.05.009

Al-Ali, A., Elwakil, A., Ahmad, A., & Maundy, B. (2017). Design of a portable low-cost impedance analyzer. *BIODEVICES 2017 - 10th International Conference on Biomedical Electronics and Devices, Proceedings; Part of 10th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2017*, *2017-Janua*(Biostec), 104–109. https://doi.org/10.5220/0006121901040109

Bakr, A. A., Radwan, A. G., Madian, A. H., & Elwakil, A. S. (2016). Aging effect on apples bio-impedance using AD5933. *2016 3rd International Conference on Advances in Computational Tools for Engineering Applications, ACTEA 2016*, *35*(1), 158–161. https://doi.org/10.1109/ACTEA.2016.7560131

Batta, M. (2018). Machine Learning Algorithms - A Review. *International Journal of Science and Research (IJSR)*, *18*(8), 381–386. https://doi.org/10.21275/ART20203995

Bertemes-Filho, P., Laus Bertemes, W., Cavalieri, R., Torres Paré, A., Spessatto, J., & Savi, D. (2020). Ripening classification of bananas (Musa acuminate) using electrical impedance spectroscopy and support vector machine. *International Journal of Biosensors & Bioelectronics*, *6*(4), 99–101. https://doi.org/10.15406/ijbsbe.2020.06.00195

Brosnan, T., & Sun, D. W. (2004). Improving quality inspection of food products by computer vision - A review. *Journal of Food Engineering*, *61*(1 SPEC.), 3–16. https://doi.org/10.1016/S0260-8774(03)00183-3

Bruns, H. A. (2009). A survey of factors involved in crop maturity. *Agronomy Journal*, *101*(1), 60–66. https://doi.org/10.2134/agronj2007.0271R

Carreira-Perpiñán, M. A., & Zharmagambetov, A. (2020). Ensembles of Bagged TAO Trees Consistently Improve over Random Forests, AdaBoost and Gradient Boosting. *FODS 2020 - Proceedings of the 2020 ACM-IMS Foundations of Data Science Conference*, *section 3*, 35–46. https://doi.org/10.1145/3412815.3416882

Castro, W., Oblitas, J., De-La-Torre, M., Cotrina, C., Bazan, K., & Avila-George, H. (2019). Classification of Cape Gooseberry Fruit According to its Level of Ripeness Using Machine Learning Techniques and Different Color Spaces. *IEEE Access*, *7*, 27389–27400. https://doi.org/10.1109/ACCESS.2019.2898223

Chacón, P., & Bustamante, R. O. (2001). The effects of seed size and pericarp on seedling recruitment and biomass in Cryptocarya alba (Lauraceae) under two contrasting moisture

regimes. *Plant Ecology*, *152*(2), 137–144. https://doi.org/10.1023/A:1011463127918

Chen, L. Y., Wong, D. M., Fang, C. Y., Chiu, C. I., Chou, T. I., Wu, C. C., Chiu, S. W., & Tang, K. T. (2018). Development of an electronic-nose system for fruit maturity and quality monitoring. *Proceedings of 4th IEEE International Conference on Applied System Innovation 2018, ICASI 2018*, 1129–1130. https://doi.org/10.1109/ICASI.2018.8394481

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-August-2016*, 785–794. https://doi.org/10.1145/2939672.2939785

Chowdhury, A., Datta, S., Bera, T. K., Ghoshal, D., & Chakraborty, B. (2018). Design and development of microcontroller based instrumentation for studying complex bioelectrical impedance of fruits using electrical impedance spectroscopy. *Journal of Food Process Engineering*, *41*(1), 1–13. https://doi.org/10.1111/jfpe.12640

Colledge, S., Conolly, J., & Sherman, S. (2004). Archaeobotanical evidence for the spread of farming in the eastern Mediterranean. *Current Anthropology*, *45*(4 SUPPL.), 35–58. https://doi.org/10.1086/422086

Corbellini, S., & Vallan, A. (2014). Arduino-based portable system for bioelectrical impedance measurement. *IEEE MeMeA 2014 - IEEE International Symposium on Medical Measurements and Applications, Proceedings*. https://doi.org/10.1109/MeMeA.2014.6860044

Cubero, S., Aleixos, N., Moltó, E., Gómez-Sanchis, J., & Blasco, J. (2011). Advances in Machine Vision Applications for Automatic Inspection and Quality Evaluation of Fruits and Vegetables. *Food and Bioprocess Technology*, *4*(4), 487–504. https://doi.org/10.1007/s11947-010-0411-8

El-Bendary, N., El Hariri, E., Hassanien, A. E., & Badr, A. (2015). Using machine learning techniques for evaluating tomato ripeness. *Expert Systems with Applications*, *42*(4), 1892–1905. https://doi.org/10.1016/j.eswa.2014.09.057

El Hadi, M. A. M., Zhang, F. J., Wu, F. F., Zhou, C. H., & Tao, J. (2013). Advances in fruit aroma volatile research. *Molecules*, *18*(7), 8200–8229. https://doi.org/10.3390/molecules18078200

Figueiredo Neto, A., Cárdenas Olivier, N., Rabelo Cordeiro, E., & Pequeno de Oliveira, H. (2017). Determination of mango ripening degree by electrical impedance spectroscopy. *Computers and Electronics in Agriculture*, *143*(September), 222–226. https://doi.org/10.1016/j.compag.2017.10.018

Gao, H., Zhu, F., & Cai, J. (2010). A review of non-destructive detection for fruit quality. *IFIP Advances in Information and Communication Technology*, *317*, 133–140. https://doi.org/10.1007/978-3-642-12220-0_21

Hussain, M. I., El-Keblawy, A., Akhtar, N., & Elwakil, A. S. (2021). *Electrical Impedance Spectroscopy in Plant Biology*. *August*, 395–416. https://doi.org/10.1007/978-3-030-73245-5_12

Ibba, P. (2020). *PhD in Food Engineering and Biotechnology Fruit Quality Evaluation Using Electrical Impedance Spectroscopy*. Libera Università di Bolzano.

Ibba, P., Falco, A., Abera, B. D., Cantarella, G., Petti, L., & Lugli, P. (2020). Bio-impedance and circuit parameters: An analysis for tracking fruit ripening. *Postharvest Biology and Technology*, *159*(May 2019), 110978. https://doi.org/10.1016/j.postharvbio.2019.110978

Ibba, P., Falco, A., Rivadeneyra, A., & Lugli, P. (2018). Low-Cost Bio-Impedance Analysis System

for the Evaluation of Fruit Ripeness. *Proceedings of IEEE Sensors*, *2018-Octob*, 1–4. https://doi.org/10.1109/ICSENS.2018.8589541

Islam, M., Wahid, K., & Dinh, A. (2018). Assessment of ripening degree of avocado by electrical impedance spectroscopy and support vector machine. *Journal of Food Quality*, *2018*. https://doi.org/10.1155/2018/4706147

Jagannath, J. H., Das Gupta, D. K., Bawa, A. S., Sebastin, R., & Vishnu, B. (2005). Assessment of ripeness/damage in banana (musa paradisiacal) by acoustic resonance spectroscopy. *Journal of Food Quality*, *28*(3), 267–278. https://doi.org/10.1111/j.1745-4557.2005.00008.x

Kitinoja, L., & AlHassan, H. Y. (2012). Identification of appropriate postharvest technologies for small scale horticultural farmers and marketers in sub-Saharan Africa and South Asia -Part 1. Postharvest losses and quality assessments. *Acta Horticulturae*, *934*(January 2012), 31–40. https://doi.org/10.17660/ActaHortic.2012.934.1

Kulkarni, M. N. (2017). Data driven modelling for banana ripeness assessment. *ICICCS*, 662–665.

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors (Switzerland)*, *18*(8), 1–29. https://doi.org/10.3390/s18082674

Liu, X. (2006). *Electrical Impedance Spectroscopy Applied in Plant Physiology Studies Xing Liu Master of Engineering RMIT Electrical Impedance Spectroscopy Applied in Plant Physiology Studies*. 116.

Lorente, D., Blasco, J., Serrano, A. J., Soria-Olivas, E., Aleixos, N., & Gómez-Sanchis, J. (2013). Comparison of ROC Feature Selection Method for the Detection of Decay in Citrus Fruit Using Hyperspectral Images. *Food and Bioprocess Technology*, *6*(12), 3613–3619. https://doi.org/10.1007/s11947-012-0951-1

Macdonald, J. R., & Johnson, W. B. (2005). Fundamentals of Impedance Spectroscopy. *Impedance Spectroscopy: Theory, Experiment, and Applications, Second Edition*, 1–26. https://doi.org/10.1002/0471716243.ch1

Mansourbahmani, S., Ghareyazie, B., Zarinnia, V., Kalatejari, S., & Mohammadi, R. S. (2018). Study on the efficiency of ethylene scavengers on the maintenance of postharvest quality of tomato fruit. *Journal of Food Measurement and Characterization*, *12*(2), 691–701. https://doi.org/10.1007/s11694-017-9682-3

Mireei, S. A., Sadeghi, M., Heidari, A., & Hemmat, A. (2015). On-line firmness sensing of dates using a non-destructive impact testing device. *Biosystems Engineering*, *129*(31), 288–297. https://doi.org/10.1016/j.biosystemseng.2014.10.012

Mohapatra, D., Mishra, S., & Sutar, N. (2010). Banana and its by-product utilisation: An overview. *Journal of Scientific and Industrial Research*, *69*(5), 323–329.

Musacchi, S., & Serra, S. (2018). Apple fruit quality: Overview on pre-harvest factors. *Scientia Horticulturae*, *234*(July), 409–430. https://doi.org/10.1016/j.scienta.2017.12.057

Prasanna, V., Prabha, T. N., & Tharanathan, R. N. (2007). Fruit ripening phenomena-an overview. *Critical Reviews in Food Science and Nutrition*, *47*(1), 1–19. https://doi.org/10.1080/10408390600976841

Pua, E. C., & Davey, M. R. (2010). Plant developmental biology. *Plant Developmental Biology*, *1*, 1–497. https://doi.org/10.1007/978-3-642-02301-9

Rehman, M., Abu Izneid, B. A. J. A., Abdullah, M. Z., & Arshad, M. R. (2011). Assessment of quality of fruits using impedance spectroscopy. *International Journal of Food Science and*

Technology, *46*(6), 1303–1309. https://doi.org/10.1111/j.1365-2621.2011.02636.x

Ruales, J., Pólit, P., & Nair, B. M. (1990). Evaluation of the nutritional quality of flakes made of banana pulp and full-fat soya flour. *Food Chemistry*, *36*(1), 31–43. https://doi.org/10.1016/0308-8146(90)90005-O

Sabilla, I. A., Wahyuni, C. S., Fatichah, C., & Herumurti, D. (2019). Determining banana types and ripeness from image using machine learning methods. *Proceeding - 2019 International Conference of Artificial Intelligence and Information Technology, ICAIIT 2019*, 407–412. https://doi.org/10.1109/ICAIIT.2019.8834490

Sanaeifar, A., Mohtasebi, S. S., Ghasemi-Varnamkhasti, M., Ahmadi, H., & Lozano, J. (2014). Development and application of a new low cost electronic nose for the ripeness monitoring of banana using computational techniques (PCA, LDA, SIMCA, and SVM). *Czech Journal of Food Sciences*, *32*(6), 538–548. https://doi.org/10.17221/113/2014-cjfs

Seymour, C. B. (1993). *Banana. Biochemistry of fruit ripening* (pp. 83–106).

Srivastava, S., & Sadistap, S. (2018). Non-destructive sensing methods for quality assessment of on-tree fruits: a review. *Journal of Food Measurement and Characterization*, *12*(1), 497–526. https://doi.org/10.1007/s11694-017-9663-6

T. Frelink. (2006). *5657-Application-Note-EIS*.

Torres, I., Pérez-Marín, D., De la Haba, M. J., & Sánchez, M. T. (2017). Developing universal models for the prediction of physical quality in citrus fruits analysed on-tree using portable NIRS sensors. *Biosystems Engineering*, *153*, 140–148. https://doi.org/10.1016/j.biosystemseng.2016.11.007

Vanoli, M., & Buccheri, M. (2012). Overview of the methods for assessing harvest maturity. *Stewart Postharvest Review*, *8*(1), 1–11. https://doi.org/10.2212/spr.2012.1.4

Voyant, C., Notton, G., Kalogirou, S., Nivet, M. L., Paoli, C., Motte, F., & Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, *105*, 569–582. https://doi.org/10.1016/j.renene.2016.12.095

Wilson, A. D., & Baietto, M. (2009). Applications and advances in electronic-nose technologies. *Sensors*, *9*(7), 5099–5148. https://doi.org/10.3390/s90705099

Zude, M. (2016). Original article Non-destructive prediction of banana fruit quality using VIS / NIR spectroscopy. *Institute of Agricultural Engineering Bornim*, *58*(March), 1–8. https://doi.org/10.1051/fruits

A

APPENDIX

In this chapter there are relevant documents collected or produced during this dissertation's execution.

A.1 Impedance evolution graphics

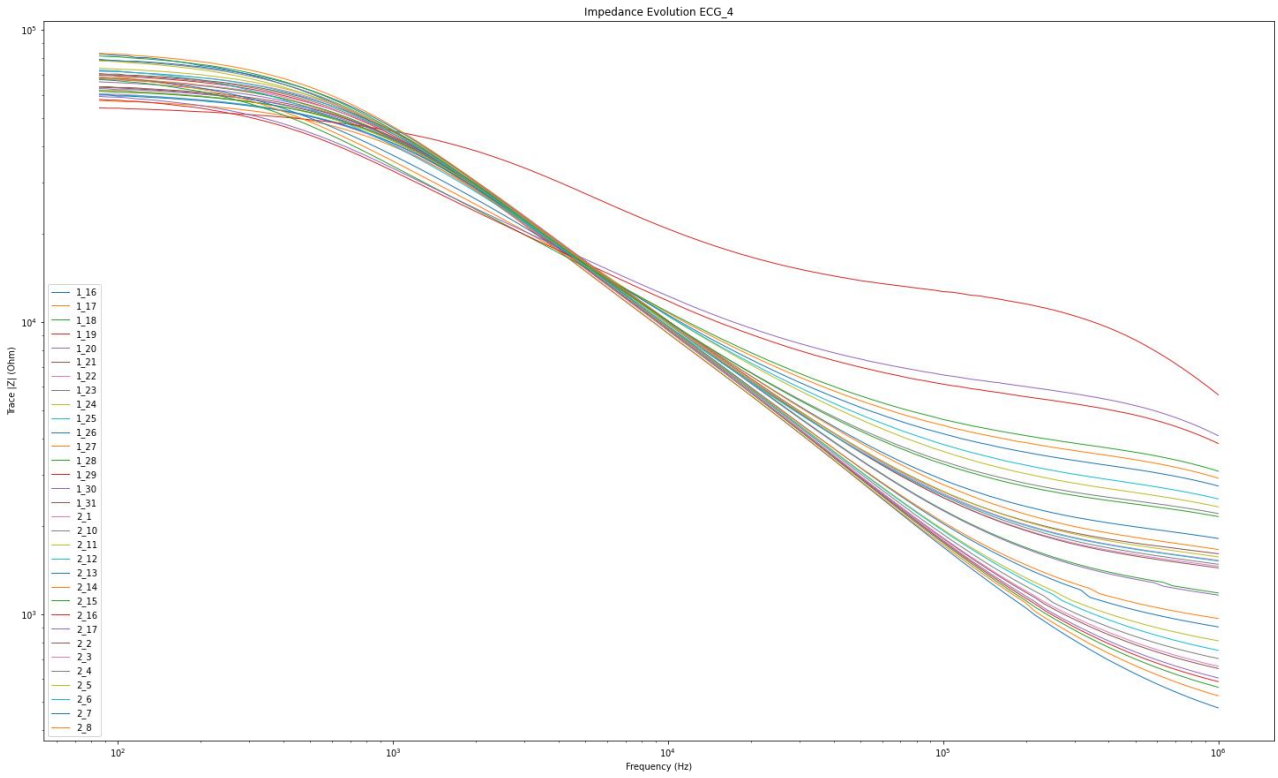

Figure 6.1 — Impedance evolution for sample 4

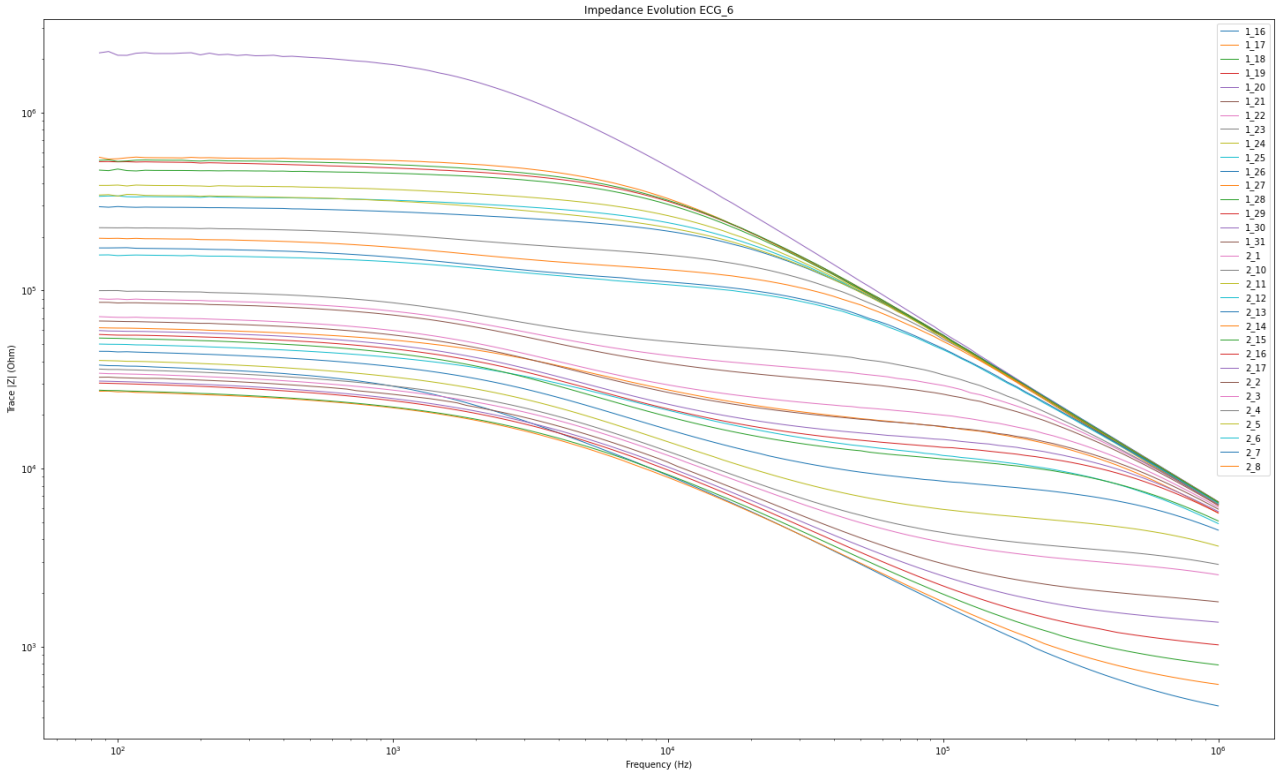

Figure 6.3 — Impedance evolution for sample 6

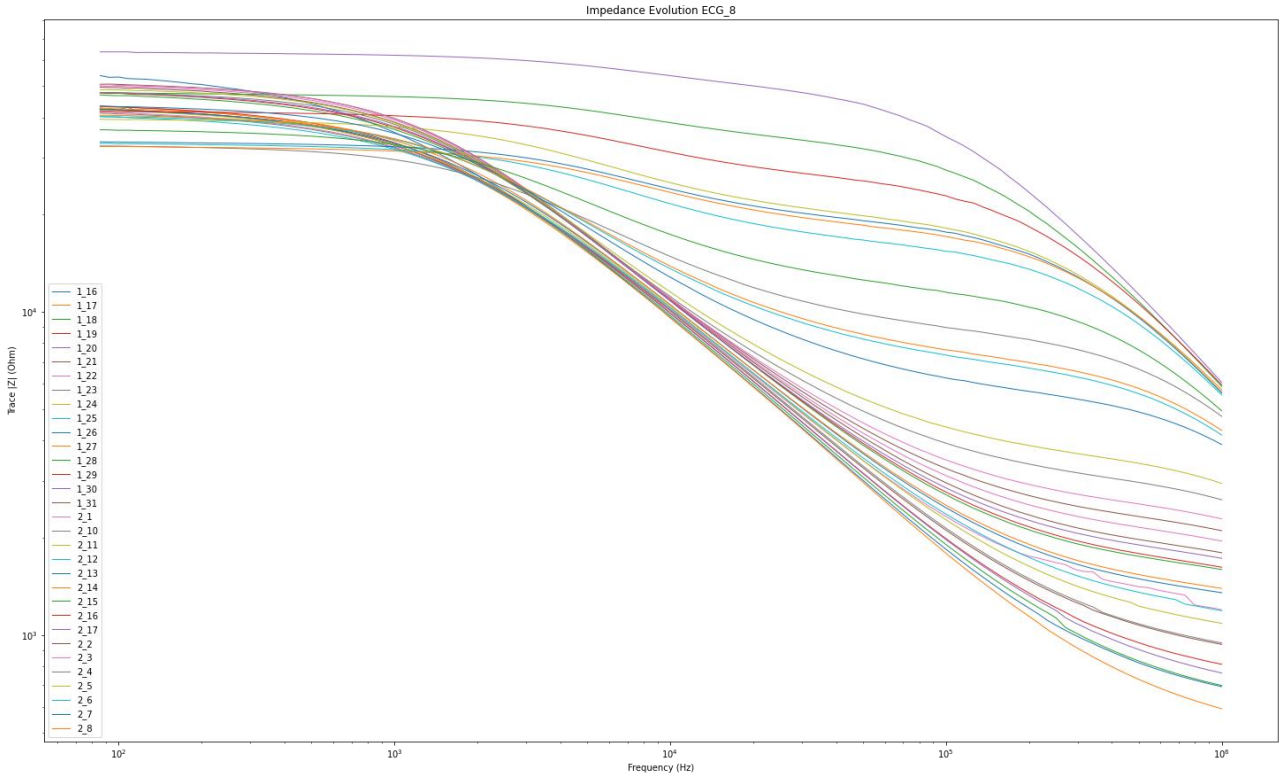

Figure 6.2 — Impedance evolution for sample 8

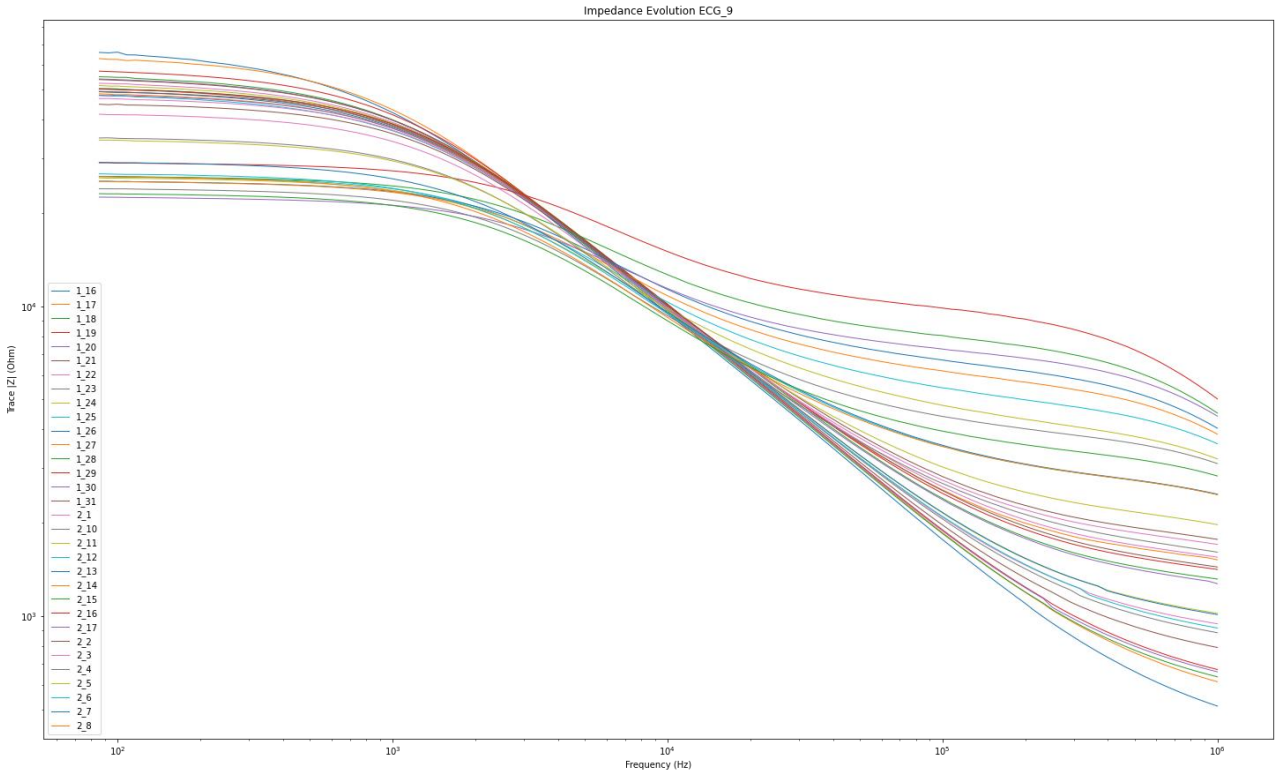

Figure 6.4 — Impedance evolution for sample 9

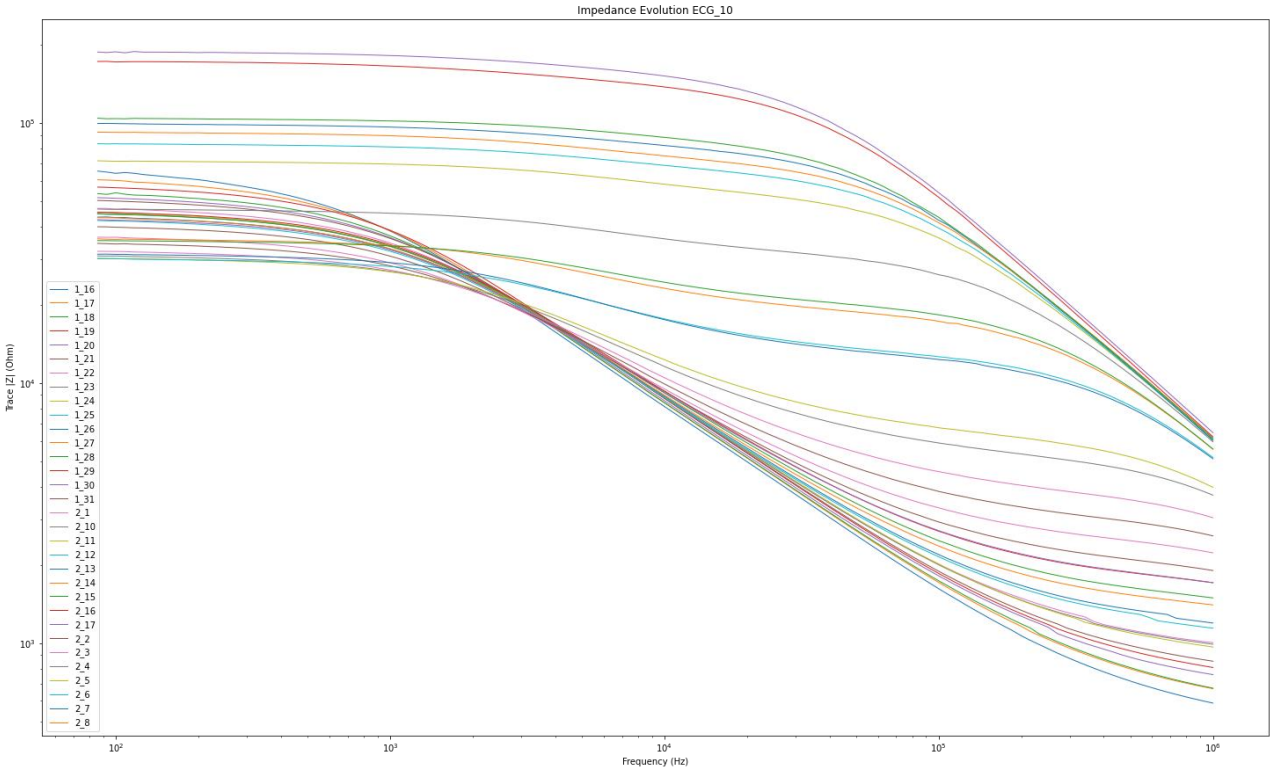

Figure 6.5 — Impedance evolution for sample 10

2023 Eduardo Gonçalves Freitas RIPENING ASSESSMENT CLASSIFICATION USING AI ALGORITHMS WITH EIS DATA

NOVA SCHOOL OF SCIENCE & TECHNOLOGY