Università degli Studi di Genova

Scuola Politecnica & Facoltà di Ingegneria

Laurea Magistrale in Ingegneria Strategica

International MSc in Engineering Technology for Strategy and Security of Genoa University

---

Wine Qualification with an Electronic Nose

Advisors:

Prof. Francesco Masulli, Prof. Stefano Rovetta, Prof. Alberto Cabri

Candidate:

Ms**. Termeh Khani**

Academic Year 2025-2026 - STRATEGOS

Dedication

To my family, your unwavering belief in me, your encouragement through every challenge, your guidance as my constant source of strength, and your love have formed the foundation of this achievement.

To my friends, for their companionship and good wishes.

To my mentors and teachers, for sharing their knowledge and inspiring me throughout my academic life.

Acknowledgments

Firstly, I would like to express my gratitude to my academic advisors, Professors Francesco Masulli, Stefano Rovetta, and Alberto Cabri, from the University of Genova, for their valuable guidance during my master's thesis.

I would also like to thank Dr. Quaini for his essential support and assistance during the wine testing procedures.

Abstract

The evaluation of wine quality has traditionally relied on human sensory panels, a process that is inherently subjective, time-consuming, and expensive. This thesis presents the implementation and validation of a machine learning framework for an electronic nose (e-nose) system designed to objectively assess wine quality. The study, carried out at Vega Research Laboratories SRL in collaboration with the Associazione Italiana Sommelier, utilizes a prototype e-nose, featuring a Raspberry Pi and an array of six MQ-series gas sensors, to analyze a diverse dataset of 90 wine samples. Each sample was also evaluated by certified sommeliers, providing 19 key sensory descriptors. The core contribution of this work lies in the development of the data pipeline and predictive models: machine learning algorithms, including Linear SVM, Multi-Layer Perceptron (MLP), and ensemble methods (Voting and Bagging classifiers), were trained on this combined dataset to predict wine quality. Using robust Leave-One-Out Cross-Validation (LOOCV), the system demonstrated high efficacy, with the Voting Classifier achieving peak performance. Feature analysis revealed that a hybrid approach, integrating sensor data with human-rated features, significantly improved prediction accuracy for complex attributes. This work validates the feasibility of a cost-effective e-nose as a powerful tool for rapid, objective wine analysis, with significant potential for application in quality control within the wine industry.

TABLE OF FIGURES

TABLE OF CONTENTS

CHAPTER 1

Introduction of an Electronic Nose

1.1 What is an Electronic Nose?

An Electronic Nose (E-Nose) works as a sophisticated measurement apparatus engineered to emulate the olfactory sense of mammals, generating characteristic digital data, or "fingerprints," for various gases and odors [1]. The architecture of an E-Nose fundamentally relies on three principal elements: a sample delivery system, a detection system, and a computing system [2].

The sample delivery system enables the generation of the headspace (volatile compounds) of a sample, which represents the fraction to be analyzed. This unit injects the headspace into the detection mechanism under constant operating conditions, ensuring measurement reliability. The detection system constitutes the reactive core, typically comprising an array of chemical sensors that convert chemical information from the environment into analytical signals [3]. Since scents and odors generally consist of multiple chemical constituents, the E-Nose must detect combinations of chemicals rather than single compounds. This capability is realized through the integration of several unique sensors into a sensor array. When exposed to volatile compounds, these sensors experience changes in their electrical properties, which are then processed by the computing system using advanced pattern recognition algorithms to differentiate between various smells, correlating the electrical "fingerprint" with specific odors or substances [2] [4].

The foundational concept of the E-Nose emerged in the 1980s [1]. Over the ensuing decades, the technology has evolved significantly, transitioning from cumbersome and high-cost laboratory apparatus into affordable, mobile, and low-power devices [3]. Today, E-Noses are widely utilized across diverse sectors, including food analysis, healthcare, environmental monitoring, and security applications, making them valuable tools for both research and industry [4].

1.2 The Key Role of Electronic Noses in Several Industries

The versatility of E-Nose technology has led to its adoption across a diverse range of sectors, where it provides rapid, reliable solutions to complex odor and gas detection challenges. By offering efficient analysis, E-Noses enhance product quality, assure safety, and optimize processes in industrial settings. The following sections explore the key roles of this technology in several major industries.

1.2.1 Food and Beverage Industry

In the food and beverage sector, maintaining quality and safety is paramount. E-Noses serve as highly efficient, non-destructive tools for sensory evaluation, enabling manufacturers to assess the freshness and flavor of products while detecting contaminants or spoilage indicators [5].

Example: Food Quality Control

In this industry, E-Noses have become widely used tools across diverse products like milk, meat, bread, and beverages for quality control. They are employed by food manufacturers to assess the freshness and flavor profiles of processed foods like Nestlé and Kraft Heinz, which are food manufacturers.

For instance, commercial systems like the HERACLES E-Nose are used to profile the sensory attributes of products like cheese, ensuring consistency and detecting off-flavors [6]. In bakeries, devices such as the Cyranose 320 analyze the aroma profiles of bread and pastries, monitoring for volatile organic compound (VOC) changes that indicate staleness and thereby minimizing food waste [7]. Similarly, beverage companies utilize E-Noses to maintain product standards by analyzing the aroma profiles of coffee, tea, and soft drinks, ensuring they meet consumer expectations and uphold brand image [5].

Figure1. Food Quality Control [23]

1.2.2 Environmental Monitoring

Environmental pollution poses significant risks to human health and ecosystems. E-Noses play a critical role in air quality monitoring by providing real-time, continuous detection of toxic substances and their emission sources. These systems are essential components of monitoring stations in urban, industrial, and residential areas [8].

Example: Air Quality Monitoring Stations

E-Nose units designed by companies like Sensigent LLC are deployed to track key air toxins, including VOCs, Nitrogen Oxides (NOx), and Sulfur Dioxide (SO_2). By automatically analyzing samples at regular intervals, these systems provide real-time pollution metrics, help identify local pollution sources, and allow environmental agencies to implement timely control measures and monitor long-term air quality changes [8], [9].

Figure2. Air Quality Monitoring [24]

1.2.3 Healthcare and Medical Diagnostics

In healthcare, early and non-invasive screening significantly improves patient outcomes. E-Noses function as rapid diagnostic tools by detecting disease biomarkers present in bodily fluids, such as the volatile organic compounds (VOCs) in exhaled breath. The principle is that metabolic changes caused by diseases alter the chemical composition of breath, creating specific VOC patterns that an E-Nose can identify [10].

Example: Disease Detection

The Cyranose 320, for example, is used to analyze breath samples for signatures associated with conditions like asthma, Chronic Obstructive Pulmonary Disease (COPD), and lung cancer [11]. This capability supports early disease detection, enables personalized treatment strategies, and aids in remote patient monitoring, ultimately improving healthcare efficiency.

Figure3. Cyranose Electronic Nose [24]

1.2.4 Pharmaceutical Industry

The pharmaceutical industry requires stringent quality control to ensure the safety, efficacy, and uniformity of drug products. E-Noses are crucial tools in this process, used to detect odor contaminants, assess drug stability, and monitor product quality [12]. They help manufacturers identify off-odors or off-tastes in formulations, which can indicate issues in manufacturing or storage.

Example: Drug Quality Control

Pharmaceutical companies like Pfizer and GlaxoSmithKline use portable analyzers like the Cyranose 320 to measure the odor profiles of tablets, capsules, and liquids, ensuring batch-to-batch consistency and patient acceptability [13]. Furthermore, by monitoring VOC profiles over time, E-Noses assist in stability testing and shelf-life determination, helping manufacturers establish proper storage conditions and expiry dates.

1.2.5 Safety and Security

Guaranteeing public safety is a global priority, and E-Noses are increasingly applied for security screening to detect explosives, chemical weapons, and other hazardous substances. These devices enhance threat detection speed and improve the effectiveness of preventive security measures [14].

Example: Explosives Detection

A notable example is the MOSES (Mobile Olfactory Sensing and Explosives System), a portable sensor array designed to detect volatile compounds associated with explosives. Used in airports, border crossings, and by bomb squads, MOSES can rapidly analyze air samples or swabs from suspicious objects. It provides security personnel with a fast and accurate method to distinguish dangerous materials like nitroaromatics and peroxides from harmless substances, creating greater real-time situational awareness [14].

1.2.6 Agriculture and Agro-Industry

In agriculture, maximizing yield and ensuring food security depend on the early detection of pests and diseases. E-Noses offer a fast, non-destructive method for screening crop health and evaluating environmental conditions [15].

Example: Crop Disease Detection

Technologies like the AROMA (Agricultural Remote Olfactory Monitoring Apparatus) are fitted with VOC-sensing arrays to detect the specific odorous changes emitted by plants under stress from pathogens or pests. For example, AROMA can monitor high-value crops for signs of powdery mildew or bacterial wilt, alerting farmers to problems before visible symptoms appear. This enables timely interventions, supports precision agriculture practices, and helps reduce yield losses while promoting environmental sustainability [16].

1.2.7 Consumer Goods and Product Development

Sensory attributes are critical for influencing consumer preferences in industries like cosmetics and personal care. E-Noses are instrumental in product development and quality control, helping to determine fragrance compositions, detect off-odors, and maintain product consistency [17].

Example: Fragrance Evaluation

Systems from companies like Alpha MOS are used in fragrance laboratories to rapidly evaluate perfume compositions, optimize scent profiles, and ensure batch-to-batch consistency. By analyzing the chemical patterns of volatile aroma compounds, these systems help manufacturers refine products, benchmark against competitors, and prevent brand tarnishing caused by variations in fragrance, ultimately leading to increased consumer satisfaction [5].

1.3 Limitations and Challenges of Electronic Nose

While electronic noses represent a significant technological breakthrough with broad applications, their effectiveness is constrained by several key limitations. These challenges, which span operational, analytical, and practical domains, must be addressed to ensure reliable performance and wider adoption. The following sections detail these primary limitations.

1.3.1 Operational and Environmental Sensitivity

A core challenge for E-Nose technology is its high sensitivity to operational conditions. Sensor performance can be significantly suppressed by environmental factors such as high humidity and the presence of interfering substances like alcohol [18]. This reduces the sensors' ability to accurately distinguish target volatile organic compounds (VOCs), which is critical for applications like food quality control or atmospheric pollutant detection.

Compounding this issue is sensor drift, a major problem where the baseline readings of sensors change over time, even under stable conditions [19]. This drift degrades the reliability of measurements and necessitates frequent recalibration to specific temperature and humidity ranges, a prerequisite for maintaining stable and accurate operation. Furthermore, many current sensors exhibit overlapping selectivity, meaning they respond to multiple similar compounds. This lack of high specificity makes it difficult to distinguish between chemically analogous VOCs in complex matrices, leading to potential ambiguity in the analysis [20].

1.3.2 Analytical and Performance Limitations

From an analytical standpoint, E-Noses face inherent limitations in quantification and chemical identification. They struggle with absolute calibration and precise quantification, especially when analyzing complex mixtures [21]. This limits their effectiveness in fields that demand exact concentration data, such as environmental compliance monitoring or pharmaceutical quality control.

Unlike gold-standard analytical techniques like Gas Chromatography-Mass Spectrometry (GC-MS), which can precisely identify and quantify individual chemical components, E-Noses provide a holistic "fingerprint" without detailed analytical determination [21]. This lack of compound-specific information means that while an E-Nose can tell you that a sample smells different, it cannot always tell you exactly which chemical is causing the difference. This makes them less effective for applications requiring clarified compound identification or analysis at a highly detailed level.

1.3.3 Practical and Deployment Challenges

The practical deployment of E-Noses is also constrained by several factors. Many sensor platforms have a finite operational lifespan, requiring periodic replacement and servicing [18]. This incurs direct costs and, more significantly, indirect costs from system downtime. For instance, in a manufacturing facility using an E-Nose for continuous quality control, sensor replacement halts monitoring, which can lead to production delays or undetected quality issues. The total cost, therefore, includes not only the sensor itself but also the associated production losses.

Additionally, most E-Nose systems are designed for static or semi-static application modes. This design principle limits their suitability for dynamic, real-time monitoring in field environments where air flow is variable and continuous analysis is required [19]. This constraint restricts their use in many industrial and outdoor settings where conditions are unpredictable.

1.3.4 Method Development and Implementation Overhead

Finally, deploying an E-Nose for a new application is a resource-intensive process. It requires significant upfront effort in method development, including careful sensor selection, extensive calibration, and the creation of robust data analysis models [22]. This laborious and specialized process can slow down the adoption of E-Nose technology in new areas or industries, as it demands a high level of technical expertise and investment before the system can provide reliable results.

CHAPTER 2

DIGITAL & ANALOG MODULES

This chapter details the design, selection, and integration of the hardware components that constitute the electronic nose prototype. The design is guided by the primary objective of developing a low-cost, portable system for detecting the volatile organic compounds (VOCs) that define the unique aroma profile of wine. The system architecture is fundamentally a hybrid of digital and analog modules. The following sections describe the core digital processing unit, the analog sensor array, the critical bridging circuitry, and the final physical assembly of the prototype.

2.1 The Core Processing Unit: Raspberry Pi 4

The Raspberry Pi 4 (RPI4) serves as the primary connecting and computing unit in the Electronic Nose system. It is responsible for linking all hardware components, facilitating the control of the entire system, and performing subsequent data processing and analysis. The choice of the RPI4 is based on its capability to serve as the system's central point, connecting the gas sensors (MQ series) and the temperature/humidity sensor (DHT22) to detect and analyze volatile organic compounds (VOCs), such as ethanol and sulfites, in the wine samples [25].

The RPI4 operates using a Linux operating system, which is configured and tuned to ensure optimal performance with the sensor modules and data processing algorithms. This platform provides essential support for interfacing with the sensors and implementing personalized aroma analysis algorithms. The system's software is implemented primarily using the Python programming language, which offers robust libraries for hardware control and data science.

Figure4. Raspberry Pi 4 Model B [26]

The computational power of the RPI4 enables the execution of in-depth data analytics, trend detection, and the gathering of insights into the unique aroma profiles of wine samples. This integrated processing capability makes the Electronic Nose system a highly informative tool for quality control, enological analysis, and new product introduction in the wine business. The contributions of the RPI4 are considered invaluable due to its stability, low maintenance requirements, and compatibility with the selected sensor technologies.

2.2 The Sensor Array: Selection and Rationale for MQ-Series Sensors

The primary function of the electronic nose is to capture a complex "fingerprint" of the volatile organic compounds (VOCs) that define a wine's aroma profile. The selection of the sensor array is therefore the most critical design decision. For this project, the MQ-series of analog gas sensors was selected.

The following subsections provide a general overview of this sensor technology, detail the specific components chosen for the array, and explain how the system design addresses their inherent limitations [27], [28].

2.2.1 Overview of MQ Sensor Technology

MQ series analog gas sensors have been extensively utilized in gas sensing applications for several decades. These sensors, initially developed by Hanwei Electronics, operate using a resistive gas sensing mechanism to detect the presence of various gases in the surrounding environment. They have gained considerable popularity due to their affordability, ease of use, and versatility in detecting a wide range of gases [27].

The decision to base the E-Nose on this technology is driven by several key advantages. Their cost-effectiveness makes them highly accessible for academic research and the development of low-cost commercial devices. The wide range of detectable gases, including carbon monoxide (CO), methane (CH4), liquefied petroleum gas (LPG), alcohol, and hydrogen (H2), allows for the creation of a comprehensive sensing array tailored to specific needs. Furthermore, their simple interfacing, typically featuring an analog voltage output, enables straightforward integration with a microcontroller without complex protocols. Finally, their rapid response time is vital for applications like air quality monitoring and, in this case, for capturing the dynamic aroma profile of a wine sample.

However, the use of MQ sensors also presents significant engineering challenges that must be addressed in the system design. They are highly sensitive to environmental factors such as temperature and humidity, which can affect their accuracy and reliability. Their limited selectivity means they often respond to multiple gases, leading to cross-sensitivity that can complicate data interpretation in complex mixtures. Additionally, they can experience drift and have a limited lifespan, requiring periodic recalibration to maintain accuracy. Finally, their relatively large physical size compared to modern digital sensors can be a constraint in applications where space is at a premium. These limitations are not seen as disqualifying factors but rather as design parameters that the overall system must mitigate [28].

2.2.2 Component Breakdown and Technical Rationale

To create a comprehensive and robust VOC fingerprint for wine analysis, a diverse array of six MQ-series sensors was selected. Each sensor operates on the principle of a metal-oxide semiconductor (MOS), typically tin dioxide (SnO_2). In the presence of oxygen, the SnO_2 surface absorbs oxygen molecules, trapping electrons and creating a high-resistance state. When a reducing gas (like ethanol or methane) is present, it reacts with the absorbed oxygen, releasing the trapped electrons back into the material and causing a measurable decrease in resistance. Each sensor is equipped with an internal heater element, which is critical for bringing the sensing material to its optimal operating temperature (typically 200-400°C) for a consistent and sensitive reaction [27].

- **MQ-2 Combustible Gas Sensor**

This sensor utilizes SnO_2 as its sensing layer and is engineered for high sensitivity to Liquefied Petroleum Gas (LPG), propane, and hydrogen, with a typical detection range of 300-10,000 ppm. It also exhibits cross-sensitivity to methane, carbon monoxide, and alcohol. In the context of wine analysis, it acts as a general-purpose detector for a broad class of combustible hydrocarbons that may be present as complex byproducts of fermentation or as markers of spoilage [29]. Its analog output is derived from a voltage divider circuit, where the sensor's changing resistance (Rs) alters the output voltage relative to a load resistor (RL).

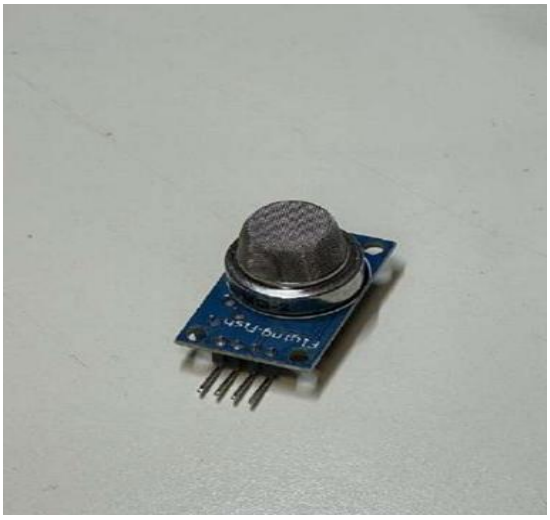


Figure5. MQ-2 [30]

- **MQ-3 Alcohol Sensor**

This sensor is specifically optimized for the detection of alcohol vapors, with a particularly high sensitivity to ethanol (C_2H_5OH). Its effective detection range for ethanol is approximately 0.05 to 10 mg/L. This makes it the cornerstone sensor for the E-Nose, providing the most direct and quantifiable signal related to the wine's alcohol content, a fundamental quality parameter. Its rapid response and recovery time are crucial for capturing the volatile profile during sampling [31].

.

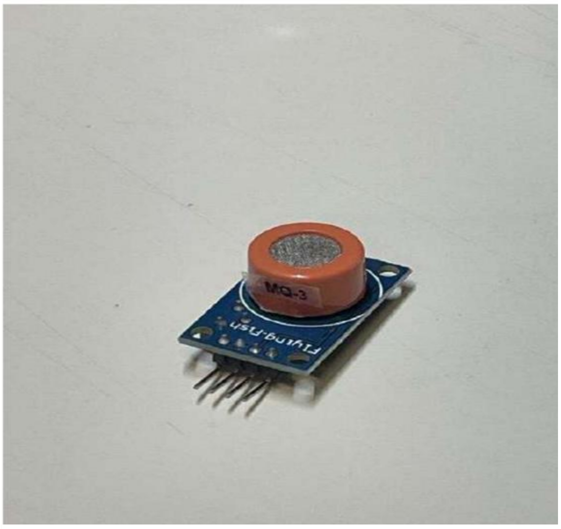


Figure6. MQ-3 [32]

- **MQ-4 Methane Sensor**

While its primary design is for detecting methane (CH_4) and compressed natural gas (CNG) in the range of 300-10,000 ppm, the MQ-4's sensitivity to other hydrocarbons makes it a valuable addition. It helps differentiate between various hydrocarbon-based volatiles, contributing to a more detailed fingerprint. This can be useful for distinguishing between the aromatic profiles of wines fermented with different yeast strains, which can produce varying levels of higher-order alcohols and hydrocarbons [33].

.

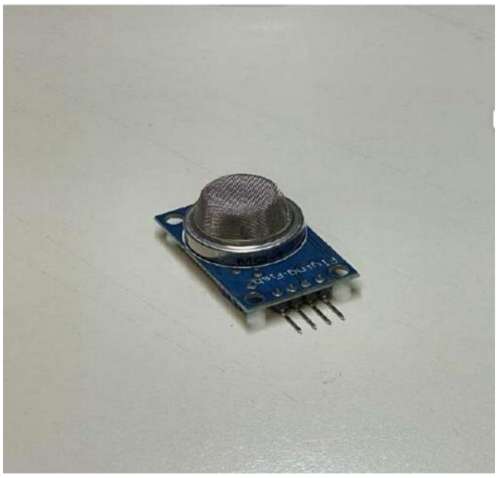


Figure7. MQ-4 [34]

- **MQ-6 LPG/Butane Sensor**

The MQ-6 offers a different sensitivity profile compared to the MQ-2, with its highest sensitivity to LPG, iso-butane, and propane (typical range: 100-10,000 ppm for iso-butane). Including both the MQ-2 and MQ-6 provides complementary data rather than simple redundancy. The differing response curves to the same gases allow the system's pattern recognition algorithms to more accurately deconvolute the complex mixture of hydrocarbons present in the wine's headspace [35].

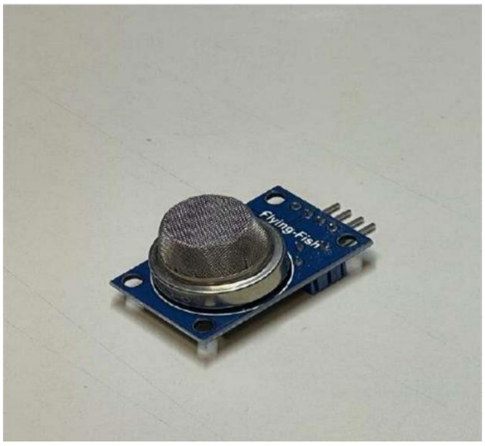


Figure8. MQ-6 [36]

- **MQ-7 Carbon Monoxide Sensor**

The MQ-7 is unique in this array due to its specific design for detecting Carbon Monoxide (CO) in the range of 20-2,000 ppm. Its operation requires cycling the heater voltage between 5.0V (for 60 seconds, to clean the sensor) and 1.4V (for 90 seconds, for sensing). The presence of CO in wine can be an indicator of specific fault conditions, such as the activity of certain spoilage bacteria or incomplete fermentation. Its inclusion allows the E-Nose to monitor for this specific and important fault marker [37].

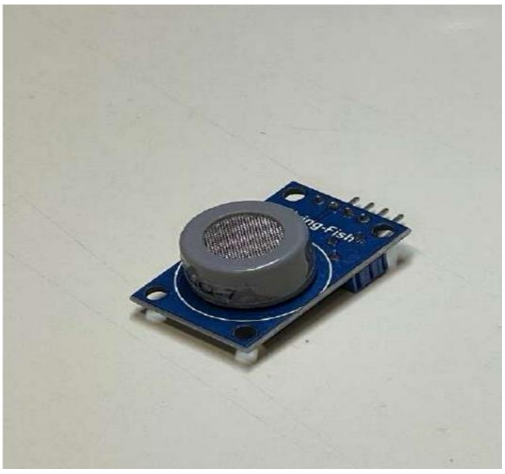


Figure9. MQ-7 [38]

- **MQ-135 Air Quality Sensor**

This is a broad-spectrum sensor essential for capturing the complex bouquet of wine. It is highly sensitive to ammonia (NH_3), nitrogen oxides (NOx), sulfur dioxide (SO_2), and benzene, with a typical detection range for ammonia of 10-300 ppm. In enology, this sensor is critical for detecting both desirable and undesirable compounds. For instance, it can detect sulfur compounds (H_2S, SO_2), which at low levels contribute to varietal character but at higher levels are potent off-flavors (e.g., rotten egg or burnt match smells). Its ability to detect ammonia is also relevant for monitoring fermentation health [39].

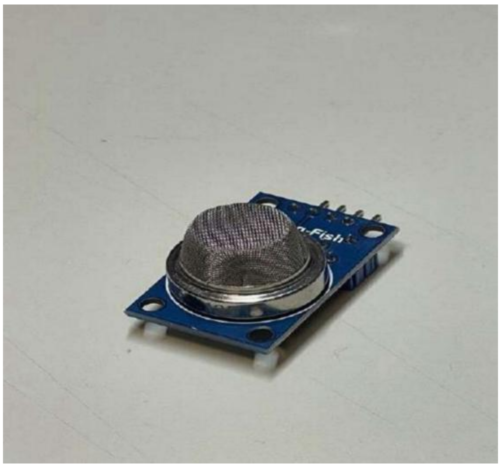


Figure10. MQ-135 [40]

2.2.3 Mitigating Limitations in System Design

The inherent limitations of MQ-series sensors are not viewed as insurmountable obstacles but as critical design parameters that have been proactively addressed through the system's architecture. To counteract the high sensitivity to temperature and humidity, which can cause significant baseline drift, the system integrates a DHT22 temperature and humidity sensor. The data from this sensor is used in real-time compensation algorithms to normalize the gas sensor readings, thereby ensuring accuracy across varying environmental conditions. The challenge of limited selectivity, or cross-sensitivity, is overcome not by relying on a single sensor's output, but by interpreting the collective response pattern, or "fingerprint," of the entire six-sensor array. The computational power of the Raspberry Pi 4 is essential here, enabling the execution of pattern recognition algorithms that can deconvolute this complex data to identify specific aromas. Furthermore, the RPi4 is leveraged to manage long-term stability and calibration requirements by automating recalibration routines against known baselines and applying software-based drift-correction models. While the relatively large physical size of the MQ sensors was noted, it was deemed an acceptable trade-off for their cost-effectiveness and versatility in this research prototype, where functionality was prioritized over miniaturization. These integrated mitigation strategies are fundamental to ensuring the system's reliability, accuracy, and robustness for practical wine analysis [28].

2.3 Integration of MCP3008 Analog-to-Digital Converter (ADC)

As established in Section 2.2, the MQ-series sensor array generates continuous analog voltage signals. To process this data with the digital Raspberry Pi 4, an intermediate component is required to translate these analog signals into discrete digital values. For this purpose, the Microchip MCP3008, a 10-bit Analog-to-Digital Converter (ADC), was selected [41].

The MCP3008 is a widely utilized and cost-effective ADC that serves as an essential bridge between the analog sensors of the E-Nose and the digital microcontroller. Its selection is based on several key advantages. It offers eight analog input channels, providing the versatility needed to connect the six MQ sensors while leaving two channels spare for future expansion. The 10-bit resolution, which provides 1024 discrete digital levels, offers reasonable accuracy for distinguishing between subtle changes in gas concentration. Communication is handled via the Serial Peripheral Interface (SPI), which allows for straightforward and high-speed integration with the Raspberry Pi 4.

However, the project's design also accounts for the MCP3008's limitations. While its 10-bit resolution is adequate, it may not provide the precision of higher-end, more expensive ADCs. The sampling rate, while sufficient for this application, could be a limitation in projects requiring faster, real-time data acquisition. To ensure optimal accuracy, the design incorporates the Raspberry Pi's stable 3.3V output as a reference voltage for the ADC's VREF pin, which simplifies the circuit without sacrificing significant performance for this application's needs.

The physical connection is critical for signal integrity. The analog output of each of the six MQ sensors is connected to a dedicated input channel on the MCP3008 (CH0-CH5). The ADC is powered by the 3.3V pin of the Raspberry Pi. The digital communication is established by connecting the MCP3008's SPI pins (CLK, DOUT, DIN, CS) to the corresponding GPIO pins on the Raspberry Pi 4, as detailed in the table below.

MCP3008 to Raspberry Pi 4 Pin Connection

MCP3008 Pin	Raspberry Pi 4 Pin	Description
VDD (Pin 16)	3.3V Power	ADC Power Supply
VREF (Pin 15)	3.3V Power	Reference Voltage
AGND (Pin 14)	Ground (GND)	Analog Ground
CLK (Pin 13)	SPI CLK (Pin 23)	Serial Clock
DOUT (Pin 12)	SPI MISO (Pin 21)	Master In, Slave Out
DIN (Pin 11)	SPI MOSI (Pin 19)	Master Out, Slave In
CS (Pin 10)	SPI CE0 (Pin 24)	Chip Select
DGND (Pin 9)	Ground (GND)	Digital Ground

This configuration enables the Raspberry Pi to sequentially request and receive digitized data from each of the MQ sensors, forming the core of the system's data acquisition process.

2.4 DHT22-AM2302 Temperature & Humidity Sensor

To implement the environmental compensation strategy outlined in Section 2.2.3, it is critical to acquire accurate, real-time data on the system's ambient conditions. For this essential task, the DHT22 sensor, also identified as the AM2302 and manufactured by Aosong Electronics, was integrated into the design. This component is a digital sensor widely utilized for environmental monitoring in various projects and applications, providing accurate and reliable measurements of both temperature and humidity levels. The sensor assembly includes a capacitive humidity sensing element and a thermistor for temperature measurement, and it utilizes a proprietary digital signal protocol to communicate with microcontrollers or single-board computers, offering precise readings in real-time [42].

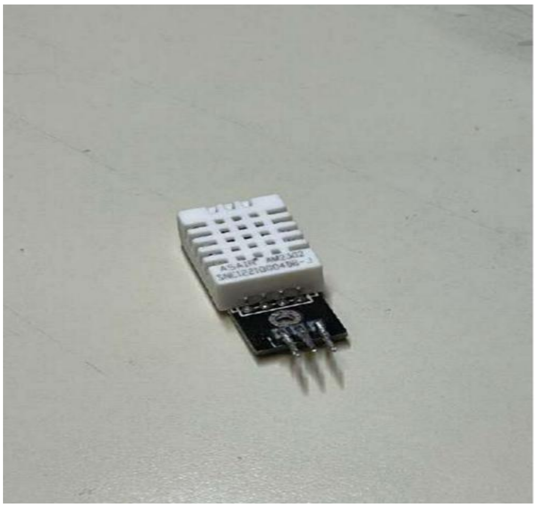


Figure11. DHT22 Temperature & Humidity Sensor [43]

The DHT22 offers several key features and advantages that make it highly suitable for environmental monitoring in the E-Nose system. It provides high accuracy in both temperature and humidity measurements. It has a wide operating range, functioning effectively across temperatures from -40°C to 80°C and humidity levels from 0% to 100% Relative Humidity (RH). It features a simple digital interface, requiring only a single data pin for communication, which simplifies the wiring and GPIO pin usage on the Raspberry Pi. Furthermore, it boasts low power consumption, making it energy-efficient for projects requiring prolonged operation. With proper handling, it offers reliable performance and stable, consistent measurements over time, ensuring dependable data acquisition for long-term monitoring applications [42], [44].

However, the system design also acknowledges and accounts for the DHT22's inherent limitations. It exhibits limited accuracy when operating in extreme temperature or humidity conditions, which is a consideration for ensuring data quality. The sensor has a relatively slow response time compared to some other, more specialized sensors, and it can be susceptible to electrical interference in noisy electrical environments. Because it is a single sensor package that combines both temperature and humidity sensing, it may offer limited flexibility in certain applications where separate sensors might be preferred. Finally, to maintain optimal performance in long-term deployments, occasional calibration may be required [44].

The physical connection of the DHT22 to the Raspberry Pi 4 is straightforward. The sensor typically has three or four pins; in the common three-pin configuration, they are VCC, Data, and GND. The pinout and connection are detailed below.

Connection Pinout to Raspberry Pi:

DHT22 Pin	**Raspberry Pi 4 Pin**	**Description**
VCC (Pin 1)	3.3V Power	Sensor Power Supply
Data (Pin 2)	Any GPIO Pin (e.g., GPIO 4)	Digital Data Line
GND (Pin 3)	Ground (GND)	Common Ground

The integration of the DHT22 is essential for acquiring the environmental data necessary to compensate for sensor variations caused by fluctuations in temperature and humidity. This concludes the discussion of the primary electronic modules utilized in the design. The physical arrangement and integration of all components—the MQ sensors, the MCP3008 ADC, and the DHT22 sensor—on the prototype's Printed Circuit Board (PCB) showcase the compact final design, which is visually represented below.

Figure12. Sensors Placement on Manufactured PCB

2.5 The Final E-Nose Prototype

The preceding sections established the theoretical framework for the sensor array, detailed the selection of the individual digital and analog modules (MQ sensors, MCP3008 ADC, and DHT22), and outlined the role of the Raspberry Pi 4 as the central processing unit. The culmination of this hardware selection and integration is the Final Version of the Electronic Nose Prototype.

This prototype represents the successful physical integration of all components into a functional system. Each module was precision-engineered with a focus on functionality, ease of assembly, and operational efficiency. The design process involved the use of FreeCAD software to iteratively refine the designs, ensuring that each part fits perfectly with the others and meets the specific requirements of the electronic nose. This modular approach not only simplifies the assembly process but also provides flexibility to modify individual components as needed, which was a significant advantage during the prototyping and testing phases.

The final system architecture is physically divided into two main units connected by a ribbon cable. The first is the core sensing unit, which houses the six MQ-series sensors, the MCP3008 ADC, and the DHT22 sensor within a custom-designed blue structure. This unit is engineered to be placed over the sample container, allowing it to analyze the volatile compounds within the headspace of the lower chamber (the headspace jar). The second unit is the Raspberry Pi 4 processing unit, which handles data acquisition, processing, and control. This separation of the sensing and processing modules helps to minimize electrical noise and thermal interference with the sensitive gas sensors.

This complete physical realization provides the necessary hardware platform for the software implementation and data acquisition algorithms that will be discussed in the following chapter. The final appearance and physical arrangement of the complete E-Nose system are presented below.

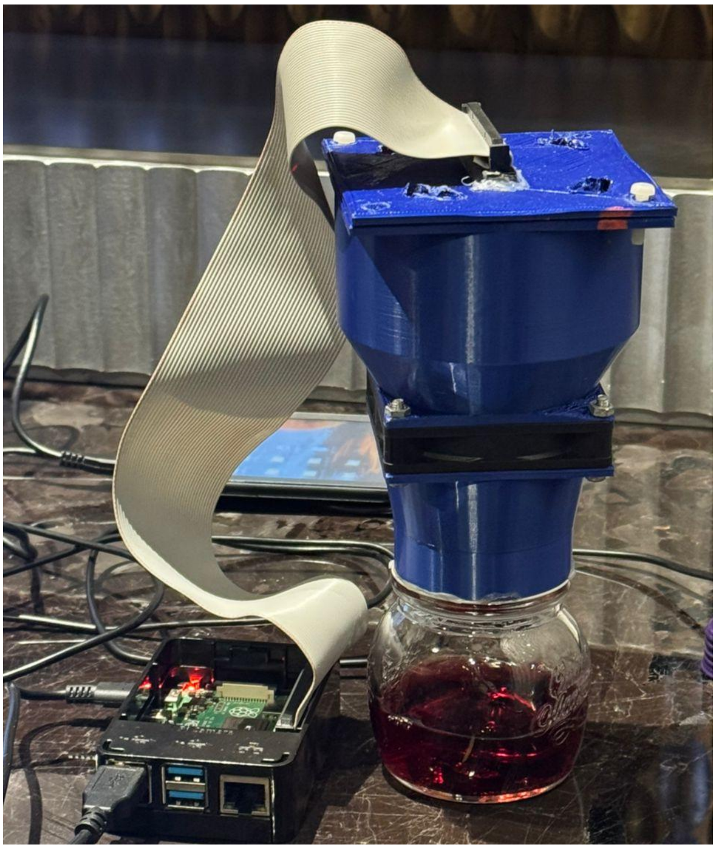


Figure13. Final Version of the E-Nose Prototype

CHAPTER 3

Software Implementation and System Integration

3.1 Introduction

This chapter details the software implementation and integration of the Electronic Nose (E-Nose) system for wine evaluation. Following the hardware refinement discussed previously, the focus now shifts to developing a robust software architecture capable of real-time data acquisition, processing, and visualization. The objective is to create a cohesive system that translates raw sensor data into meaningful insights for wine aroma analysis. The implementation leverages a Raspberry Pi as the central processing unit, chosen for its versatility and connectivity, to interface directly with the sensor array and provide a user-friendly platform for data interaction [45].

3.2 System Architecture and Hardware Integration

The core of the E-Nose system consists of an array of chemical sensors, a resistor, an analog-to-digital converter (ADC) chip, and a DHT22 sensor for monitoring environmental temperature and humidity. These components are essential for capturing the volatile organic compounds (VOCs) that constitute the wine's aroma profile.

To ensure efficient data acquisition and system simplicity, a direct connection architecture was implemented. Rather than using a multi-module setup, the sensors were connected directly to the Raspberry Pi via its USB and GPIO ports. This design choice eliminates the need for intermediary hardware, reducing potential points of failure and latency. The Raspberry Pi's built-in networking capabilities were further leveraged to enable remote access via SSH. This functionality allows for continuous monitoring and control of the system from a remote lab computer, which is critical for long-term experiments and unattended data collection [45].

3.3 Software Development Environment and Language

Python was selected as the primary programming language for the system's software development [46]. This decision was based on its versatility, readability, and extensive ecosystem of libraries that support hardware interaction, data processing, and web development. The development environment was centered around Visual Studio Code (VS Code) [47], which, through its network capabilities and the official Python extension [48], allowed for direct development, debugging, and testing of Python scripts on the Raspberry Pi. This streamlined workflow facilitated rapid iteration and real-time validation of the software's functionality.

3.4 Core Software Components

The system's functionality is implemented through three key Python scripts, each serving a distinct role in the data pipeline.

1. Main Application Script: This script serves as the central coordinator between the hardware and the data processing algorithms. It uses Flask to run a web server [49], exposing the dashboard interface where users can view and interact with sensor data, and utilizes Socket IO to ensure real-time bidirectional communication between the server and client [50, 51], keeping the dashboard updated with the latest sensor readings.

2. Gas Sensor Data Script: This script interfaces directly with the gas sensors connected to the Raspberry Pi to capture analog sensor data. It converts these analog values into digital data using an analog-to-digital converter (ADC),a process commonly managed by libraries such as the Adafruit CircuitPython MCP300x library [52], calculates average sensor values over time, and sends the processed data to the main application for visualization on the dashboard.

3. Calibration and Data Processing Script: The operational flow begins with a critical calibration process to ensure accurate sensor readings. The system collects data from the sensors, calculates average values of the readings, and compares them with reference data stored in a predefined YAML file to

validate the sensor's accuracy before proceeding further [53, 54]. If sensor readings are lower than the predetermined calibration values, the system prompts the user to repeat the calibration. This ensures that the sensors are calibrated correctly before they can be used for actual wine evaluation. If the calibration is successful and the sensor readings match the required criteria, the system proceeds to the main data entry form, where the user can input additional relevant data.

- "For a detailed view of the calibration process, refer to Figure A.1 in Appendix A."

4. Data Entry and Result Generation: Once the critical calibration process is completed and verified, the user proceeds to the data entry form. This form is designed to be user-friendly and straightforward, allowing users to input various essential parameters necessary for wine evaluation efficiently. The form collects specific human-defined attributes about the wine sample, such as the date of production, alcohol percentage, and characteristics rated on a scale (e.g., light to bold, dry to sweet)

- "This form allows users to input relevant data before processing, as illustrated in Figure A.2 (Appendix A)."

After the required data is entered, the system processes both the user-input data and the real-time sensor readings. It analyzes the integrated aroma profiles and subsequently generates the final results. These results are then displayed through a dynamic interface. This interface provides crucial insights into the gas composition (sensor outputs) and wine characteristics, culminating the measurement process.

- "The sensor outputs (humidity and temperature) are displayed in the application (Figure A.3, Appendix A)."

This entire operational sequence, from initial sensor calibration through data entry to final result generation, is seamlessly integrated to ensure the system operates efficiently and provides reliable data for comprehensive wine evaluation.

3.5 User Interface and Data Visualization

3.5.1 Framework and Architecture

To enhance usability and data accessibility, an intuitive interface for data acquisition and visualization was designed. This interface was built using HTML, CSS, and JavaScript for the front end [55] and Python for the back end, allowing for real-time visualization of sensor data. To facilitate seamless and efficient data retrieval and transmission between the server and client sides of our application, Python and the Socket IO library were utilized [50, 51]. This approach allows for real-time transmission of electronic nose sensor data to the dashboard, enabling users to monitor and analyze the wine's volatile organic compounds (VOCs) in a structured and accessible manner.

Flask was selected as the core framework for the web application interface [49], enabling smooth communication between the front-end and back-end components of the electronic nose system. The framework's simplicity, flexibility, and extensibility were crucial in developing an efficient and user-friendly platform for wine analysis. By leveraging Flask's features, an intuitive web application was created to not only display real-time sensor data but also to support calibration processes, data visualization, and user interactions. This choice allowed for the rapid development of a cohesive application, enabling a focus on integrating the core functionalities of the system and presenting the results in a seamless, user-friendly manner.

3.5.2 Dashboard and Key Features

The dashboard was designed to provide a comprehensive visualization of gas composition in wine samples. Various graphical representations were incorporated to fulfill different analytical needs and enhance interpretability. Key elements include Radar Charts, Line Charts, and Bar Charts, which were implemented using a modern JavaScript charting library such as Chart.js [56]. These visualizations allow for comparison against calibration data, tracking of trends over time, and display of environmental parameters like temperature and humidity. Data Tables are also used to display numeric sensor readings and a comparison of sensor data against calibration values.

The electronic nose system and its dashboard provide an open-source, user-friendly platform for detecting and analyzing gas particles in real time. By leveraging this technology, users can gain a detailed understanding of gas composition patterns in different wine samples. This insight not only helps in wine quality assessment but also serves as a valuable tool for researchers and producers to refine their evaluation techniques and optimize production processes.

3.5.3 Data Visualization in the Result Page

In the Result Page of our application, we present key data from the sensor readings and calibration results through various visualizations, ensuring that users can easily analyze and interpret the data. The visualizations consist of tables and charts that provide insights into both the sensor performance and environmental conditions during the analysis of the wine samples.

Tables for Sensor Data and Calibration

At the top of the page, two tables are displayed:

1. Sensor Data Table: This table shows the sensor readings for different gases detected in the sample. Each sensor's data is displayed alongside the corresponding gas type, with the numeric values representing the sensor's output.
2. Calibration Data Table: Here, we compare the sensor data with the calibration values, highlighting how the sensors are performing relative to the calibration standards. The table also includes the values measured for the wine sample being tested, with the calibration values on the left and the wine data on the right.

Radar Chart for Gas Composition:

Under the tables, a Radar Chart is used to compare the gas composition in the wine sample with the baseline calibration data. This chart visualizes the differences between the average sensor readings for the wine sample and the average calibration data, allowing users to assess how closely the wine's composition matches the calibrated standards. The chart helps in identifying any deviations in gas components that could affect the wine's characteristics.

Bar Chart for Temperature and Humidity:

Additionally, a Bar Chart is included to display the temperature and humidity readings alongside the gas data. This chart provides a clear view of how environmental conditions—such as temperature and humidity—correlate with the gas composition in the wine sample. Monitoring these parameters is critical as they can significantly impact the sensor's accuracy and the quality of the wine evaluation.

CHAPTER 4

Experiments: Workshop in Hotels

4.1 Introduction to Experimental Protocol

The experimental phase of this research, conducted under the Workshop in Hotels framework, was carried out in partnership with the AIS (Associazione Italiana Sommelier) [57]. This collaboration involved professional sommeliers and was notably led by Dr. Quaini, President of the Association. The primary objective was to evaluate the Electronic Nose system's potential for wine qualification, focusing on its ability to identify and characterize various wines through a robust sensory analysis framework [58].

During these experiments, several wine samples were tested in both laboratory and real-world environments. The procedures and results are critical for assessing how effectively the Electronic Nose can replicate human olfactory perception, which plays a crucial role in the wine industry. Furthermore, the study examines the Electronic Nose's ability to provide consistent, objective wine quality assessments that could potentially complement or even replace traditional sensory evaluation methods.

4.2 Wine Sample Selection and Description

4.2.1 Overview of the 90 Wine Samples

A total of 90 wine samples were meticulously curated for this study to create a comprehensive and challenging dataset for the Electronic Nose system. The primary objective of this expansive selection was to ensure the dataset encompassed a wide spectrum of aromatic profiles, grape varieties, and quality levels. This required diversity is crucial for successfully training a robust machine learning model that can generalize across different types of wine and mitigate potential bias toward a single varietal or regional characteristic. Consequently, the selection process was systematically designed to reflect the inherent complexity found within a real-world market, thereby providing a rigorous testbed for evaluating the system's classification and predictive capabilities.

4.2.2 Diverse Testing Environments

To ensure the robustness and reliability of the electronic nose system across various settings, experiments were conducted not only in a controlled laboratory environment but also in several real-world locations. These additional testing sites included the Savoia Hotel [59], the Melia Hotel [60], and the Holiday Inn [61] located in Genova. Each location presented unique ambient conditions, such as different temperatures, humidity levels, and potential background odors. Testing the system in these diverse environments was crucial to validate its adaptability and to ensure that its performance was not affected by external factors, confirming its suitability for practical applications in the hospitality and retail sectors.

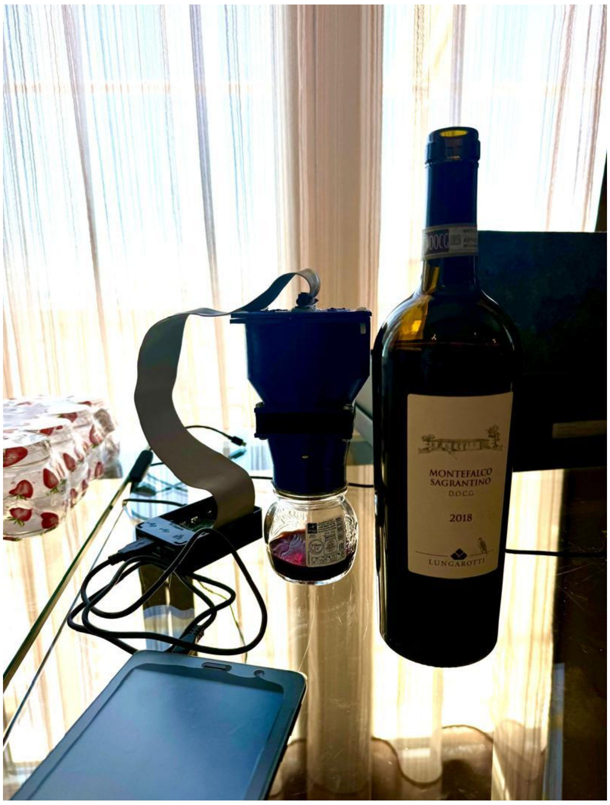


Figure14. The experimental set up at Savoia Hotel

4.2.3 Integration with Professional Events

The testing sessions at the hotels were integrated with professional wine tasting seminars, which were conducted in collaboration with the **Associazione Italiana Sommelier (AIS) [57]**. This partnership was instrumental in ensuring that the data collection occurred within a framework of high professional standards. Certified AIS sommeliers, who possess extensive and nationally recognized expertise in wine evaluation, led the tasting seminars and provided the structured sensory evaluations for the wines. This collaboration provided a dual benefit: it allowed for data collection in a dynamic, real-world setting, and it lent significant credibility to the human sensory data, which serves as the "ground truth" for this research.

Figure15. A Photograph from one of the AIS-led wine testing seminars at Savoia Hotel

4.3 Methodology

4.3.1 Human Sensory Evaluation Methodology

The primary "ground truth" for this research was established through a structured human sensory evaluation protocol [58]. This process was specifically designed to convert the inherently subjective process of wine tasting into quantifiable, objective data that could be directly compared with the objective measurements from the electronic nose.

The evaluation framework is encapsulated within a structured document known as the "Analytical Descriptive Sheet of the Wine." This comprehensive form, developed in close collaboration with certified AIS sommeliers, was engineered to capture a holistic sensory profile of each wine. It includes a comprehensive list of 19 sensory attributes, such as clarity, consistency, intensity, complexity, olfactory quality, sweetness, alcohol level, roundness, acidity, tannins, flavor, structure, balance, persistence, taste quality, state of development, harmony, and overall quality. Each attribute was rated on a standardized scale, ensuring that all evaluations were consistent and comparable across different samples and sessions.

This form was implemented directly into the project's web application to facilitate real-time digital data entry. An initial pilot study was conducted with sommeliers to test the form's usability and clarity. Based on their expert feedback, the form underwent refinement to better align with professional tasting terminology and practices. This iterative refinement process ensured that the final tool was both scientifically robust and user-friendly, capturing human-evaluated data with high fidelity.

SCHEDA ANALITICO-DESCRITTIVA DEL VINO

Associazione Italiana Sommelier

ESAME VISIVO

Limpidezza	Colore (bianchi)	Colore (rosati)	Colore (rossi)
Velato	Verdolino	Fiore di pesco	Amaranto
Abbastanza limpido	Paglierino	Ramato	Rubino
Limpido	Dorato	Salmone	Carminio
Cristallino	Ambrato	Corallo	Granato
Brillante	Mogano	Peonia	Aranciato

Consistenza (solo vini fermi)	Effervescenza (solo spumanti e frizzanti)			
	Numero catenelle	Velocità ascesa	Grana bollicine	Persistenza bollicine
Scorrevole	Scarse	Rapida	Grossolane	Evanescenti
Consistente	Mediamente numerose	Media	Mediamente fini	Mediamente persistenti
Viscoso	Numerose	Lenta	Fini	Persistenti

ESAME OLFATTIVO

Intensità	Descrittori			Complessità
Moderatamente intenso	Aromatico	Floreale	Speziato	Moderatamente complesso
Intenso	Varietale	Vegetale	Pasticceria/Panificazione	Complesso
Molto intenso	Fruttato	Fragrante	Empireumatico	Ampio
Qualità olfattiva				
Accettabile	Buono	Distinto	Ottimo	Eccellente

ESAME GUSTO-OLFATTIVO

Dolcezza	Acidità
Secco	Poco fresco
Poco dolce	Moderatamente fresco
Moderatamente dolce	Fresco
Dolce	Vibrante
Molto dolce	Acidulo
Alcolicità	**Tannicità**
Poco caldo	Poco tannico
Moderatamente caldo	Moderatamente tannico
Caldo	Tannico
Molto caldo	Tenace
Alcolico	Astringente
Rotondità	**Sapidità**
Poco morbido	Poco sapido
Moderatamente morbido	Moderatamente sapido
Morbido	Sapido
Vellutato	Saporito
Pastoso	Salato

Effervescenza (solo spumanti e frizzanti)				
Delicata	Moderata	Vivace	Esuberante	Incisiva

Intensità	Struttura	Equilibrio	Persistenza
Moderatamente intenso	Di medio corpo	Poco equilibrato	Moderatamente persistente
Intenso	Di corpo pieno	Mediamente equilibrato	Persistente
Molto intenso	Robusto	Equilibrato	Molto persistente

Qualità gusto-olfattiva				
Accettabile	Buono	Distinto	Ottimo	Eccellente

CONSID. FINALI

Stato evolutivo	
Pronto	Maturo

Armonia		
Poco armonico	Mediamente armonico	Armonico

Qualità complessiva				
Accettabile	Buono	Distinto	Ottimo	Eccellente

Figure16. Analytical Descriptive Sheet of the Wine

4.3.2 Electronic Nose Measurement Methodology

Parallel to the human sensory evaluation, a standardized protocol was meticulously established for the consistent collection of data using the Electronic Nose [62]. This rigorous procedure was designed to ensure that the objective sensor measurements were consistent and could be directly correlated with the subjective human sensory data.

Device Calibration and Testing

Before commencing the formal experiments, the Electronic Nose underwent a thorough calibration and functional testing phase. The device was initially employed to analyze two distinct wine samples to confirm its operational integrity and to validate the entire data acquisition pipeline. This initial testing successfully confirmed that the system was capable of reliably capturing and differentiating the volatile profiles of different wines, thereby verifying the system's foundational functionality.

The Measurement Protocol

A rigorous and repeatable procedure was precisely followed for the analysis of each wine sample. During this process, the Electronic Nose was used to analyze the sample, capturing crucial data on gas composition, temperature, and humidity. All captured data was immediately logged and time-stamped, effectively creating a precise digital fingerprint of the wine's aromatic profile. This systematic measurement procedure was uniformly applied across all samples tested, ultimately generating a comprehensive and robust dataset of objective measurements for subsequent analysis.

4.4 Final Data Collection and Execution

4.4.1 Final Experimental Execution

In this final experiment, a total of **90 wine samples** were systematically evaluated by certified sommeliers and simultaneously analyzed using the electronic nose system. Each wine was assessed under controlled conditions, with both sensory evaluations and sensor-based measurements recorded for direct comparison. This synchronized data collection was paramount, creating a powerful, unified dataset where each wine sample has a corresponding set of subjective human ratings and objective sensor readings. This integrated dataset forms the foundational basis for the machine learning analysis and model development presented in the subsequent chapter.

Figure17. Dr.Quaini and the sommeliers during the wine testing seminar at the Savoia Hotel in February 2025

4.4.2 Equipment and Data Collection Summary

The experiments were conducted using the developed Electronic Nose prototype, which was equipped with gas sensors (such as the MQ series) [63] alongside temperature and humidity sensors. The inputs provided by the certified sommeliers were recorded directly through the custom-built web-based application, concurrently with the device capturing real-time chemical data. All the collected information, which included both the human-evaluated attributes and the sensor data, was stored in JSON format (JavaScript Object Notation) for further processing and analysis [64]. Each JSON file contained a complete record for a single wine sample, including the 19 human-rated sensory attributes, the full time-series data from the six gas sensors, and the corresponding temperature and humidity readings. This robust data storage strategy ensured that no information was lost and that the dataset was perfectly organized for the machine learning models described in later chapters.

Figure18. A selection of 8 wines tasted

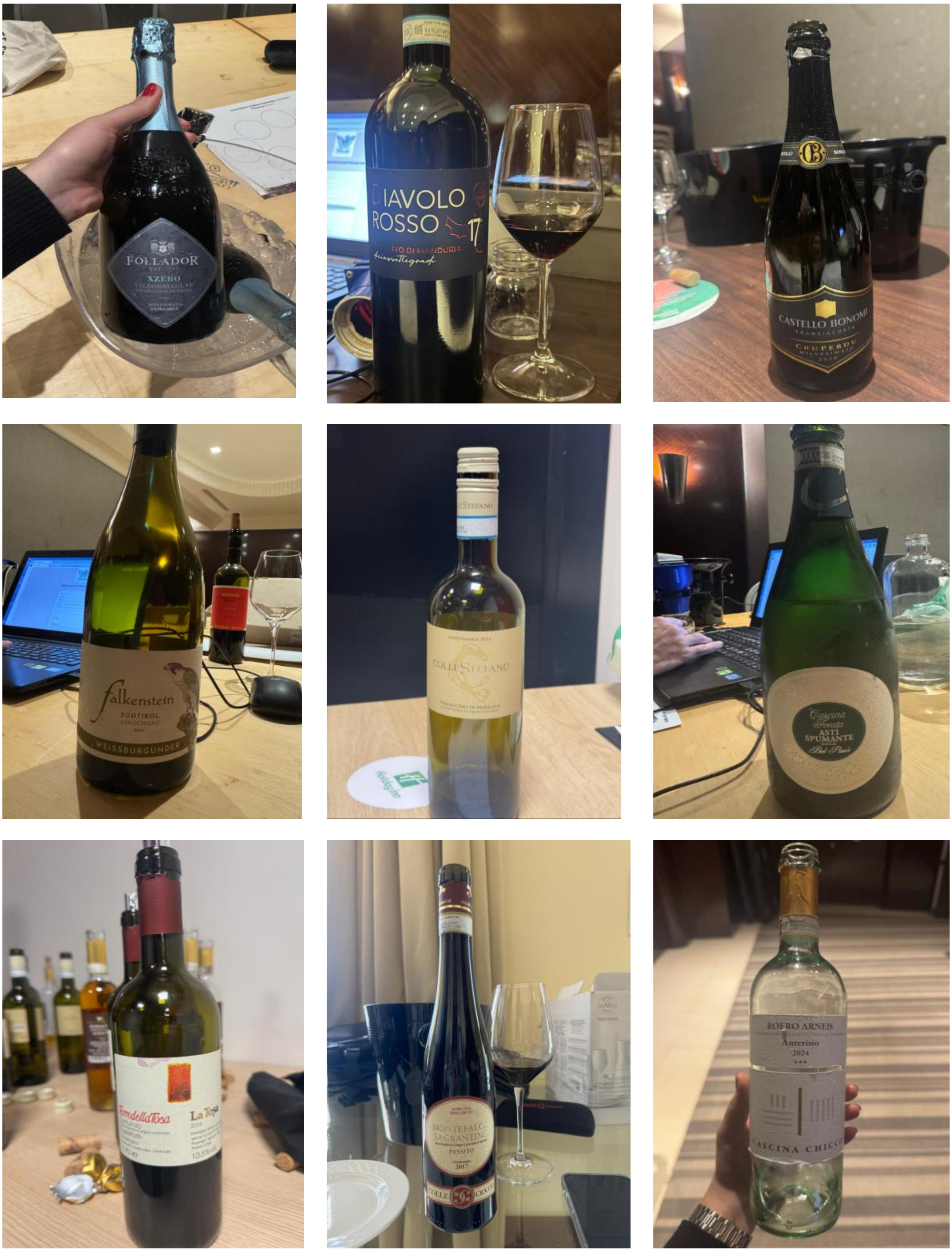


Figure19. A selection of the wines tasted

4.5 Field Validation Study: Viticultural Region Testing in Alba, Piedmont

To validate the electronic nose system's performance in a real-world environment and to expand the dataset with wines from a prominent viticultural region, a field-testing phase was conducted. The city of Alba, located in the Piedmont (Piemonte) region of Italy, was chosen for this phase [65]. Alba is internationally renowned for its high-quality wines, particularly Barolo and Barbaresco, making it an ideal location to test the system's ability to differentiate complex and prestigious wine profiles [66].

The primary objectives of this field test were to assess the system's robustness outside of a controlled laboratory setting and to acquire data from wines sourced directly from local cellars. This provided an opportunity to test the system's adaptability to different environmental conditions and to analyze wines with distinct regional characteristics, or "terroir." The portable electronic nose system was transported to and operated at a local wine establishment, where several samples were analyzed using the same data acquisition protocol established in the lab.

The system operated reliably, confirming its suitability for on-site applications in wineries and for direct quality assessment at the source. Data collected from these Alba wines added valuable diversity to the overall dataset, providing a crucial link between laboratory results and real-world wine characteristics. This phase of the project demonstrates the practical viability and flexibility of the electronic nose system beyond academic research.

Figure 20. A selection of wines tasted in Alba

Figure21. Field Validation Study in Alba

4.6 Field Validation Study: Application to Spirits at a Historic Distillery

To further validate the versatility of the electronic nose system beyond wine analysis, a field test was conducted at a prominent local distillery. The team visited the "Distilleria Storica Sità" in Gallo d'Alba, a renowned historic distillery in the Piedmont region [67]. The purpose of this visit was twofold: to receive a guided tour of the facility to understand the production processes of aged spirits and to conduct a guided tasting session, during which the electronic nose was used to analyze several distinct liquors.

This environment presented a unique challenge and opportunity for the system. Unlike wine, liquors have a much higher alcohol concentration and a different aromatic profile, often influenced by aging in wooden barrels. Testing the system in this context was crucial to assess its sensitivity and its ability to differentiate between complex spirit aromas.

A key part of the tour was understanding the very beginning of the distillation process. The distillery uses pomace—the solid remains of grapes (skins, seeds, and stems) after pressing for juice—as the fundamental raw material for its famous grappa [68]. This raw material, a dark, dried mass, is what gives grappa its unique character and connection to the land.

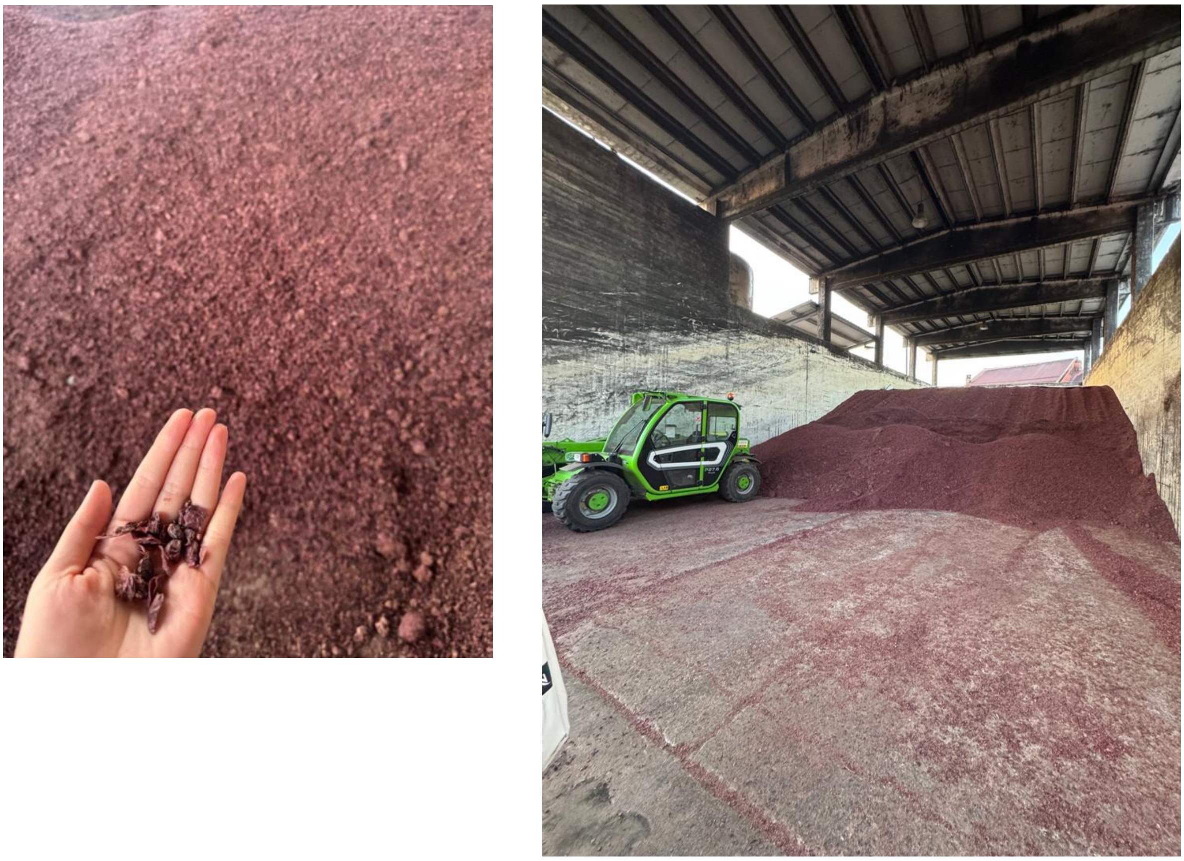


Figure22. Pile of grape pomace

During this tasting session, several distinct liquors and spirits were analyzed. These samples represented a range of products crafted at the distillery, showcasing the diversity of aromas that can be achieved through different production and aging methods.

Figure23. A selection of liquors and spirits tasted at the distillery

The testing procedure at the distillery was adapted to accommodate the nature of the samples. For each spirit, a small quantity was placed in a glass, and the electronic nose's sensor array was exposed to the headspace to capture the dominant volatile compounds. The data acquisition protocol was consistent with the wine experiments, ensuring that the collected data could be compared and analyzed using the same framework. The system successfully captured distinct aromatic fingerprints for each liquor tested, demonstrating its robustness and potential for quality control applications in the spirits industry.

This field test at "Distilleria Storica Sità" was highly valuable. It not only provided data from a completely different beverage matrix but also confirmed the portability and adaptability of the system in a real-world production environment. The experience highlighted the potential for the electronic nose to be used by distilleries for quality assurance, batch consistency, and even for detecting counterfeit products, further broadening the scope and impact of this research.

CHAPTER 5

Performance Evaluation Using Leave-One-Out Cross Validation (LOOCV)

5.1 Introduction to LOOCV

5.1.1 Why LOOCV for Small Datasets

The Leave-One-Out Cross Validation (LOOCV) is a powerful technique particularly suited for small datasets, which is a typical scenario in wine sensory analysis. With only a limited number of wine samples available for model training and testing, LOOCV allows us to maximize the use of each data point by iterating through the entire dataset.

In LOOCV, each data point is used once as the test set, while the remaining data points are used to train the model. This ensures that every sample is tested individually, resulting in an unbiased performance evaluation. For our study, LOOCV is ideal because the dataset consists of 90 wine samples. By using LOOCV, we can extract the maximum value from each sample, ensuring that the model is both trained and tested on all available data.

Moreover, LOOCV provides a robust method to assess model generalization. By testing each sample individually, we can better understand how well the model performs on unseen data, which is essential for ensuring that the system will generalize well to new, unknown wine samples [70].

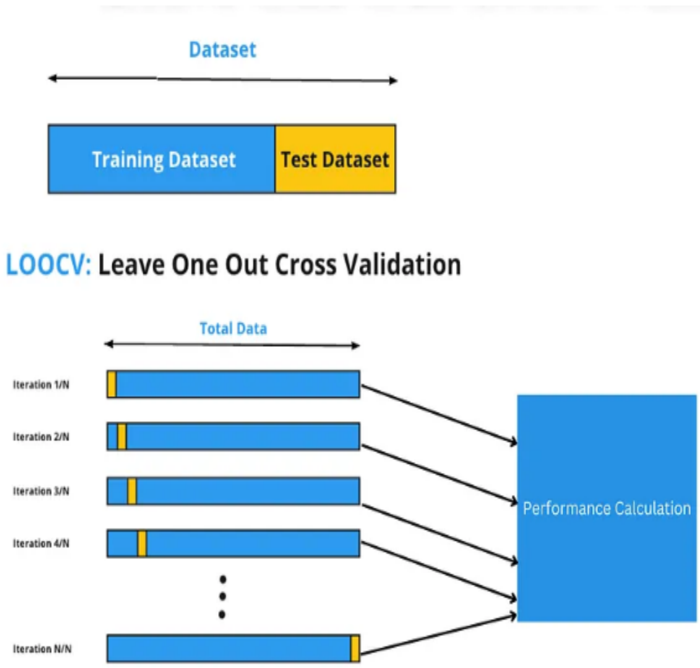


Figure24. LOOCV Process Visualization

5.1.2 Machine Learning Approach for LOOCV

To address the research questions, machine learning models are employed to predict wine quality based on sensor readings obtained from the electronic nose system. The primary goal is to develop predictive models that can accurately assess wine characteristics, which are then evaluated using LOOCV. The models that were considered include:

- Support Vector Machines (SVM): SVM is particularly effective in creating decision boundaries between different quality classes. It can handle both linear and non-linear classification tasks, depending on the choice of kernel [71].

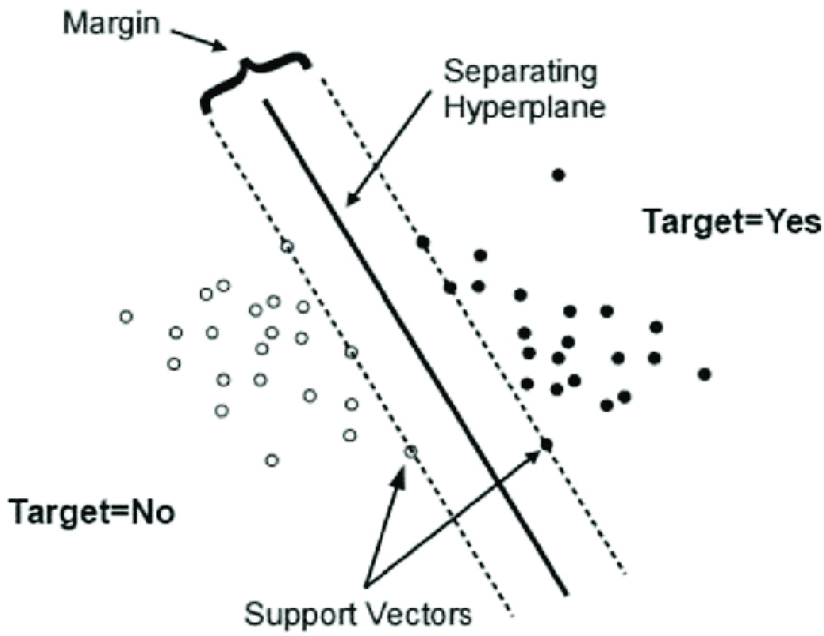


Figure25. Support Vector Machine graphical representation

- Multi-Layer Perceptron (MLP): This neural network model is used for capturing complex, non-linear relationships in the sensor data. With its multiple layers, it is well-suited to learn intricate patterns in data that simpler models might miss [72].

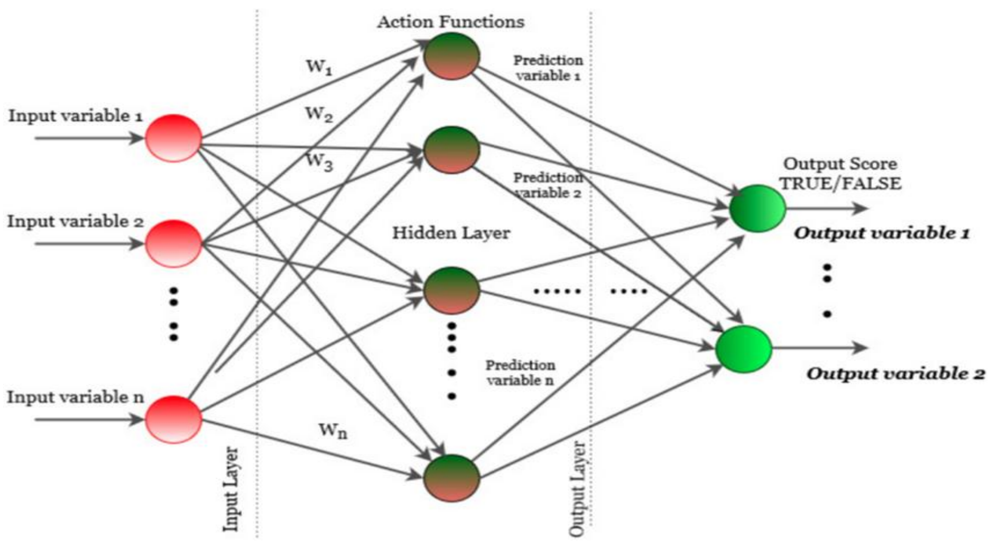


Figure26. Multi-Layer Perceptron graphical representation

- Voting Classifier: This ensemble model combines several SVM variants to improve prediction stability and robustness. By combining multiple SVM models, the voting classifier takes the majority decision of the individual models, which can result in a more accurate final prediction [70].

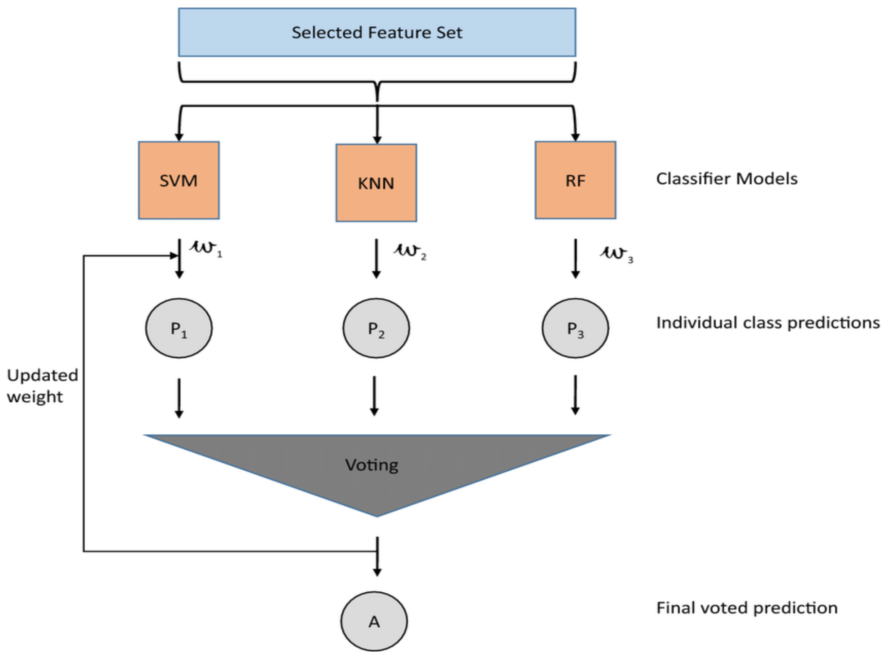


Figure27. The Voting Classifier illustration

- Bagging Classifier: Aimed at reducing the variance of the models, bagging works by training multiple models on different subsets of the data and combining them to produce a more stable model. This approach is particularly useful when the base model (e.g., SVM) is sensitive to fluctuations in the data [73].

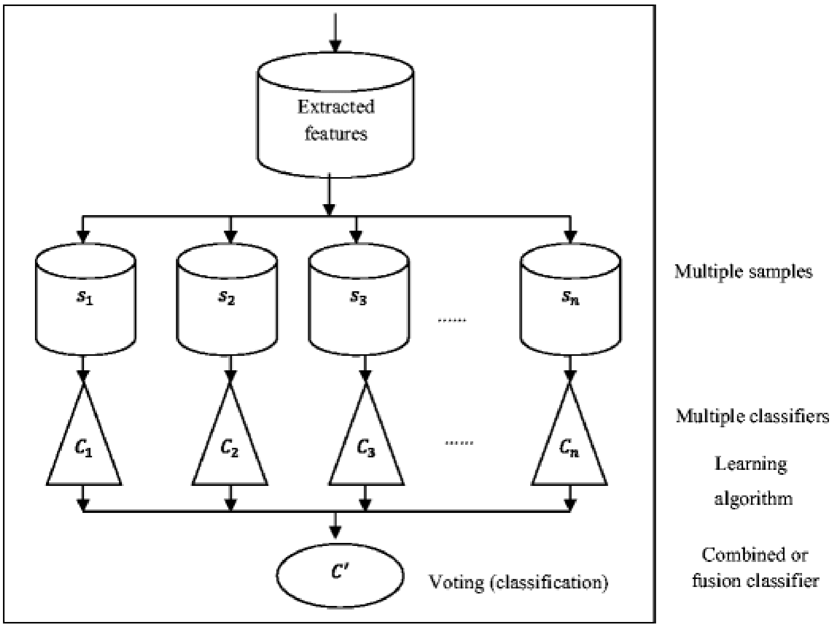


Figure28. Bagging Ensemble Classifier illustration

5.1.3 Research Objectives

The machine learning approach in this chapter addresses the following key objectives:

- Prediction Accuracy (WP3): The study investigates whether wine sensory characteristics can be reliably predicted using electronic nose data and advanced machine learning algorithms. The primary focus is on predicting overall wine quality (feature 19 from the AIS tasting card) based on sensor data.

- Generalization Capability (WP4): This research also explores how well the models generalize across all 19 sensory features derived from the AIS tasting card. The goal is to ensure that the system provides a comprehensive evaluation of wine quality, predicting a wide range of sensory attributes, such as acidity, complexity, and intensity, in addition to overall quality.

5.1.4 Methodological Framework

This study follows a supervised machine learning approach to predict wine quality using sensor data. The core methodology integrates the machine learning models described in Section 5.1.2 with a robust validation strategy suitable for small datasets.

Leave-One-Out Cross-Validation (LOOCV) Framework

Due to the limited sample size typical in wine sensory studies, a traditional train-test split would significantly reduce the training data, potentially leading to unstable models. Therefore, the LOOCV technique was employed to maximize data utilization. In this approach, for a dataset of N instances, the model is iteratively trained N times. In each iteration, N-1 samples are used for training, while the single remaining instance is held out for testing. This ensures that every data point is used for both training and validation, providing a robust and nearly unbiased estimate of the model's predictive performance.

Model Evaluation

Within this LOOCV framework, each model's performance is measured using two key metrics:

- Accuracy: The proportion of correct predictions.
- F1-Macro Score: A balanced metric that accounts for class imbalance by averaging performance across all quality categories.

5.1.5 Validation Strategy

The validation strategy, detailed in Section 5.1.4, employs LOOCV to provide a rigorous and unbiased performance evaluation. This approach is critical for ensuring the reliability of models trained on a limited number of wine samples.

5.2 Data Preprocessing and Preparation

Data preprocessing is a crucial step before applying any machine learning algorithm. In this chapter, the sensor data from the electronic nose system is processed to ensure it is structured, clean, and ready for model training and evaluation. The following steps outline the preprocessing tasks performed on the raw data, including data ingestion, feature engineering, and standardization.

5.2.1 Data Ingestion and Initial Structure

The initial dataset is collected from the electronic nose system and stored in a JSON format [77]. The raw data consists of sensor readings and human sensory evaluations of wine samples. In this step, the data is ingested and transformed into a structured format for analysis.

The following Python code snippet demonstrates how the raw data is loaded from the JSON file into a Pandas DataFrame, making it ready for further processing [76].

```
# Data Loading Process (from loocv_accuracy_f1_chart_on_server.py)
import pandas as pd
import json
with open(file_path, "r", encoding="utf-8") as f:
    data = json.load(f)
df = pd.DataFrame(data)
df_user = pd.json_normalize(df["user_data"])
```

5.2.2 Feature Engineering and Extraction

The feature engineering process transformed raw measurements into meaningful input variables for machine learning models. Two distinct feature categories were systematically constructed:

- Sensor-Derived Features: Six MOS (Metal-Oxide Semiconductor) sensors provided continuous numerical readings representing volatile organic compound (VOC) concentrations. These readings were extracted and formatted as individual features:

- Categorical Encoding: Wine color information was converted using one-hot encoding to represent the three wine types (red, white, and rosé) as binary features.

Below is an example of how we extract and encode wine color information as binary features:

```
# Color Encoding Implementation
df_filtered["is_red"] = df_user_filtered.get("color_red",
"").apply(lambda x: 1 if isinstance(x, str) and x.strip() else 0)
df_filtered["is_white"] = df_user_filtered.get("color_white",
"").apply(lambda x: 1 if isinstance(x, str) and x.strip() else 0)
df_filtered["is_rose"] = df_user_filtered.get("color_rose",
"").apply(lambda x: 1 if isinstance(x, str) and x.strip() else 0)
```

5.2.3 Quality Mapping and Data Transformation

To convert qualitative wine evaluations (such as "ottimo", "eccellente") into numeric values, a quality-mapping framework is applied. Each quality rating from human sensory evaluation is mapped to a number between 1 and 6, where 1 represents poor quality and 6 represents excellent quality.

This snippet shows how quality ratings are transformed into numerical values:

```
# Mapping quality descriptions to numeric values
quality_mapping = {
    "eccellente": 6,
    "ottimo": 5,
    "distinto": 4,
    "buono": 3.5,
    "accettabile": 3,
    "mediocre": 2,
    "insufficiente": 1
}

# Apply the mapping to the 'overall_quality' column
df_filtered["overall_quality_numeric"] =
df_filtered["overall_quality"].map(quality_mapping)
```

5.2.4 Data Validation and Quality Control

To ensure robust model training, the dataset underwent rigorous filtering to remove invalid entries and address class imbalance issues:

1. Missing Value Elimination: Records with incomplete quality ratings were systematically excluded
2. Class Viability Assessment: Classes with fewer than 2 samples were removed to ensure meaningful cross-validation

Below, the code filters out rows with missing target values and ensures that only valid data is used:

```
# Data Filtering Process
df_filtered =
df.dropna(subset=["overall_quality_numeric"]).copy().reset_index(drop=
True)
valid_classes = df_filtered["overall_quality_numeric"].value_counts()
valid_classes = valid_classes[valid_classes > 1].index
df_filtered =
df_filtered[df_filtered["overall_quality_numeric"].isin(valid_classes)
].reset_index(drop=True)
```

5.2.5 Feature Standardization and scaling

Feature standardization was implemented to ensure equal contribution of all input variables to the model training process. The StandardScaler from scikit-learn transformed features to achieve zero mean and unit variance, preventing scale-related bias in algorithm performance [69].

This code snippet standardizes the input features using StandardScaler:

```
# Standardization Process
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
X = pd.DataFrame(X_scaled)
```

5.2.6 Final Dataset Architecture

The comprehensive preprocessing pipeline resulted in a structured dataset with the following specifications:

- **Input Feature Matrix (X)**: 9 total features

    - 6 continuous sensor readings (sensor_0 through sensor_5)
    - 3 binary color indicators (is_red, is_white, is_rose)
- **Target Variable (y)**: Numerical wine quality ratings (1-6 scale)

This methodical approach to data preprocessing ensures that machine learning models receive clean, consistent, and appropriately formatted data, which is fundamental for reliable performance evaluation and meaningful comparison across different algorithms [70].

5.3 Machine Learning Model Implementation

5.3.1 Overview of Implemented Model

Four distinct models were implemented to capture different aspects of the sensor-wine quality relationship. The implementation leveraged the scikit-learn library, a comprehensive tool for machine learning in Python [69].

1. Model 1: Support Vector Machine (SVM)

- Purpose: Creates optimal decision boundaries between wine quality classes
- Variants: Linear kernel

```
    # SVM Configuration

    svm_model = SVC(kernel="linear", class_weight="balanced") # Added
balanced class_weight
```

2. Model 2: Multi-Layer Perceptron (MLP)

- Purpose: Neural network to capture non-linear patterns in sensor data
- Architecture: Two hidden layers (64 and 32 neurons)

```
    # MLP Configuration
    mlp_model = MLPClassifier(hidden_layer_sizes=(64, 32), max_iter=1000,
random_state=42)
```

3. Model 3: Voting Classifier

- Purpose: Combines multiple SVM models to improve prediction stability
- Method: Hard voting (majority rule) across three SVM variants

```
    # Voting Classifier
    voting_clf = VotingClassifier(estimators=[
        ('svm_linear', SVC(kernel='linear', class_weight='balanced')),
        ('svm_rbf', SVC(kernel='rbf', class_weight='balanced')),
        ('svm_poly', SVC(kernel='poly', degree=2,
class_weight='balanced'))
    ], voting='hard')# Majority vote
```

4. Model 4: Bagging Classifier

- Purpose: Reduces variance by training multiple SVM models on different data subsets
- Method: Bootstrap aggregating with 10 estimators

```
    # Bagging Classifier
    bag_model       =       BaggingClassifier(estimator=SVC(kernel="linear",
class_weight='balanced'), n_estimators=10, random_state=42)
   # 10 different SVM models
```

5.3.2 Integration with LOOCV Framework

The evaluation is conducted using the LOOCV framework established in **Section 5.1.4.** To address class imbalance while maintaining the integrity of the test sample, the following five-step pipeline was implemented:

Key Steps in the LOOCV Framework:

Step 1: Outermost Cross-Validation Framework (LOOCV)

The evaluation uses the iterative N-1 training and single-instance testing logic previously described

Step 2: Dynamic Data Balancing (SMOTE Application)

To address the class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was employed [74]. Crucially, SMOTE was applied **exclusively to the training partition** within each iteration. This prevents data leakage, ensuring the model is evaluated on genuine, unseen samples.

Step 3: Handling of Edge Cases and Failures

The sparsity of certain minority classes (e.g., those with only 2 or 3 instances) presented a risk of training failures. The pipeline was therefore equipped with two robustness checks:

- Single-Class Training Set: If an iteration's training set contained only one class, model training was bypassed, and a default prediction of that majority class was made for the held-out sample.
- SMOTE Neighbor Failure: If SMOTE was unable to generate synthetic samples due to an insufficient number of nearest neighbors in a tiny minority class, the resulting **ValueError** was caught. For that specific iteration, the model defaulted to training on the original, un-resampled data.

Step 4 & 5: Prediction and Metric Calculation

The selected classifier is fitted to the balanced training set, and a prediction is generated for the held-out sample. Upon completion of all 90 iterations, the average accuracy and F1-macro score are computed to provide a reliable estimate of generalization.

Here's a sample Python code to show the integration of LOOCV with the models:

```
# Core LOOCV Implementation
    loo = LeaveOneOut()
    y_true, y_pred = [], []

    for train_idx, test_idx in loo.split(X):
      X_train, X_test = X.iloc[train_idx], X.iloc[test_idx]
      y_train, y_test = y.iloc[train_idx], y.iloc[test_idx]

      # Handle edge case: single class in training set
      if len(np.unique(y_train)) < 2:
           pred = [y_train.iloc[0]]
      else:
# Apply SMOTE to the training data split
           try:
                 X_train_res,  y_train_res  =  smote.fit_resample(X_train,
y_train)
           except ValueError:
                 X_train_res, y_train_res = X_train, y_train
    # Train model and make a prediction
```

```
        model.fit(X_train_res, y_train_res)

        pred = model.predict(X_test)

y_true.append(y_test.values[0])

y_pred.append(pred[0])
```

5.4 LOOCV Results and Performance Analysis

5.4.1 Evaluation Metrics: Accuracy and F1-Macro

The **accuracy** metric measures the proportion of correct predictions made by the model. It is suitable when the classes are balanced, but in cases where the dataset is imbalanced, it can be misleading. To address this, the **F1-macro** score is used. The **F1-macro score** provides a balance between **precision** and **recall** and is particularly useful for imbalanced datasets, like those with wine quality classes that are not equally distributed [75].

The **F1-macro** score is computed for each fold in the **LOOCV process** and then averaged across all folds.

```
# Performance Metrics Calculation

acc = accuracy_score(y_true, y_pred)
f1 = f1_score(y_true, y_pred, average='macro', zero_division=0)
```

5.4.2 Performance Aggregation

In this section, we aggregate the performance metrics (accuracy and F1-macro) calculated for each fold of Leave-One-Out Cross Validation (LOOCV). By averaging the metrics across all LOOCV iterations, we obtain a more reliable estimate of each model's overall ability to predict overall wine quality.

Steps in Performance Aggregation:

1. **Collecting Results from LOOCV**:

- During each fold of LOOCV, **accuracy** and **F1-macro** scores are computed based on the model's prediction for the test sample.

- These results are stored for each fold and will be used for **aggregating** the overall performance.

2. **Calculating Average Performance**:

- After performing **LOOCV** across all 90 wine samples, the **average accuracy** and **average F1-macro score** are calculated by averaging the individual fold results.
- This aggregation process gives us a comprehensive measure of each model's **overall performance** in the overall quality prediction.

3. **Reliability of Aggregated Metrics**:

- The **average accuracy** reflects how often the model correctly predicts the wine quality across all samples.
- The **average F1-macro score** provides a balanced evaluation, particularly important for handling any class imbalances in the overall quality categories.

Below is a summary of the average accuracy and average F1-macro score for each model, calculated across all LOOCV folds:

```
🔍 Final Evaluation Results(With 2 decimal precision) (LOOCV):

✅SVM:      Acc=0.57, F1=0.57
✅MLP:      Acc=0.65, F1=0.65
✅Voting:   Acc=0.63, F1=0.61
✅Bagging:  Acc=0.60, F1=0.59
```

Figure29. Detailed LOOCV Performance Metrics

5.4.3 Comparative Analysis of Model Results (WP3)

This section presents the results for the first research objective (WP3): predicting overall wine quality. The performance of the four machine learning models is compared using aggregated accuracy and F1-macro scores from the LOOCV procedure.

Key Observations:

1. **Superiority of the Neural Network (MLP)** The Multi-Layer Perceptron (MLP) emerged as the most effective model for predicting Overall Quality, achieving an Accuracy of 0.65 and a matching F1-Macro score of 0.65. The exact parity between its Accuracy and F1-Macro score demonstrates that the MLP successfully learned the underlying patterns of both the majority and extreme-minority wine quality classes. By utilizing the dynamic SMOTE over-sampling technique during training, the network's hidden layers (64, 32) successfully mapped the complex, non-linear relationships between the sensors/inputs and the final Overall Quality metric without defaulting to a majority-class bias.

2. **Robustness of Ensemble Methodologies** The ensemble methods demonstrated strong performance when compared against the baseline MLP. The Hard-Voting Classifier, which aggregated predictions from three distinct SVM kernels (linear, RBF, polynomial) to model Overall Quality, achieved the second-highest metrics (Accuracy: 0.63, F1-Macro: 0.61). The Bagging Classifier similarly outperformed the standalone baseline metric (Accuracy: 0.60, F1-Macro: 0.59). These improvements confirm that ensemble strategies successfully reduced the variance among the individual SVM estimators, leading to more robust generalization across the unique LOOCV splits.

3. **Limitations of the Baseline SVM** The standalone, linear Support Vector Machine yielded the lowest comparative performance in predicting Overall Quality (Accuracy: 0.57, F1-Macro: 0.57). Despite utilizing balanced class weights to inform the algorithm of the uneven distribution, the linear kernel fundamentally struggled to define adequate, linear decision boundaries within the highly sparse and inherently non-linear feature space governing sensory inputs. This limitation further validates the thesis that modeling

Overall Quality requires non-linear architectures (such as the RBF Voting and the MLP), properly designed for complex classification boundaries.

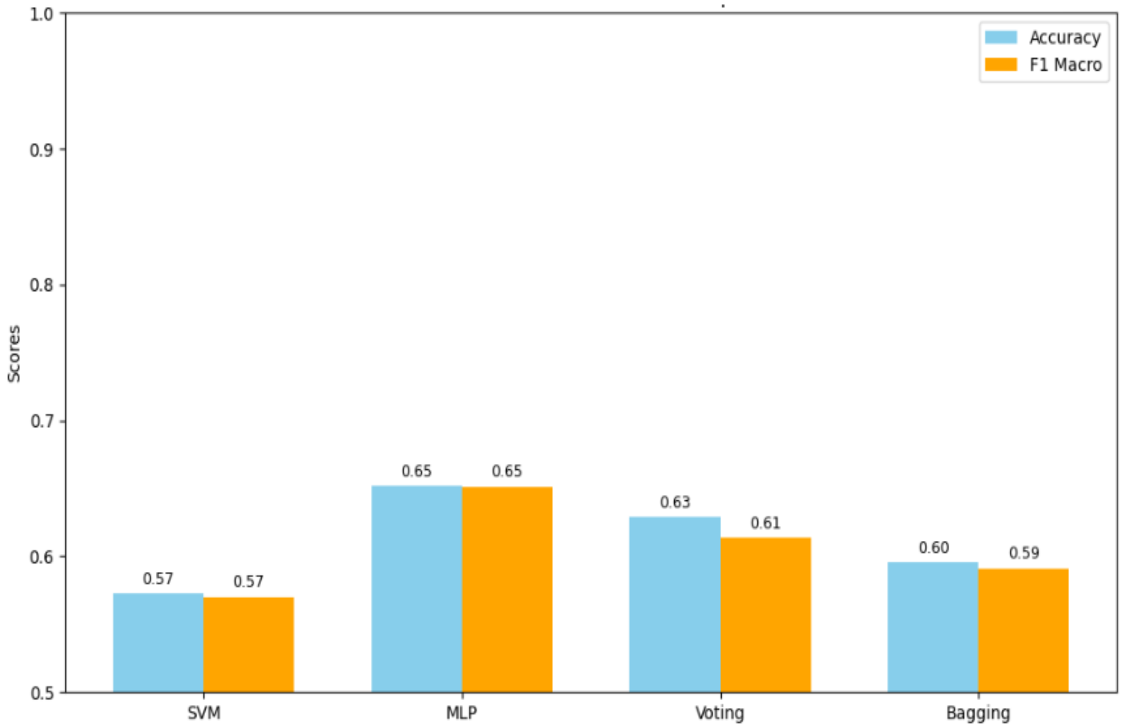


Figure30. LOOCV Performance Comparison Chart

5.4.4 Generalization Performance to all features (WP4)

This section evaluates how well the models, which were trained to predict overall wine quality (WP3), can generalize to the other 18 sensory features listed in the AIS tasting card. The features include attributes like acidity, complexity, sweetness, and others.

To evaluate generalization, the LOOCV procedure was repeated for each of the 19 sensory features. For each feature, the models were trained to predict that specific target variable and then evaluated. This process allows us to compare model performance not only on overall quality but across the entire spectrum of sensory characteristics.

Analysis of F1 Macro Score performance

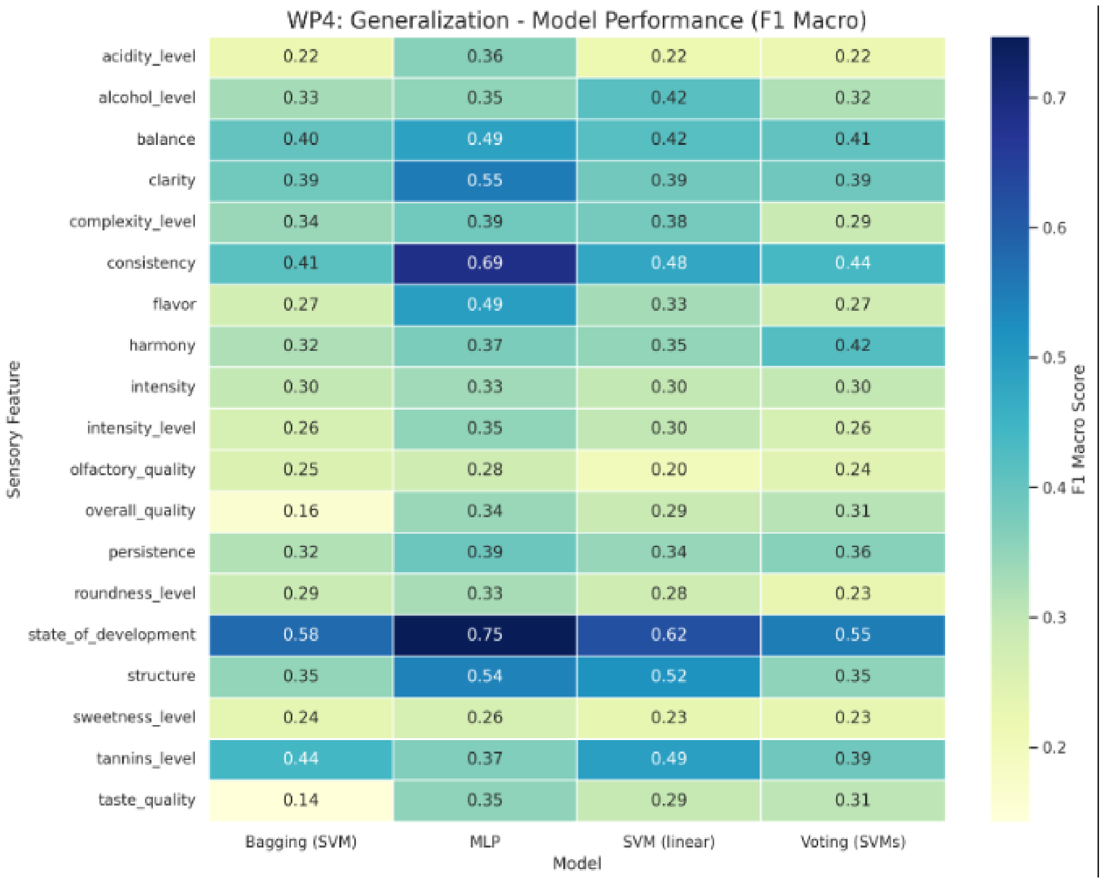


Figure31. Heatmap_F1 score for generalization to all features

The heatmap in Figure 29 illustrates the generalization performance of the models, showing how well the models, trained on overall quality, can predict the other 18 sensory features. To ensure a rigorous evaluation of generalization, particularly with potentially imbalanced class distributions, the F1 Macro score was employed as the primary evaluation metric. Unlike raw accuracy, the F1 Macro equally averages F1 scores across all classes, penalizing models that perform well only on majority classes and providing a more accurate measure of generalizability.

Model Comparison and General Trends

The heatmap reveals a clear performance hierarchy among the models. The MLP demonstrates superior generalization across most sensory attributes, evidenced by the concentration of darker cells in its column. While ensemble methods were applied to potentially enhance SVM robustness, the Voting (SVMs) model performs only marginally better than the standalone SVM (linear). Conversely, the Bagging (SVM) model generally performs worse than the other models in many of the features.

Feature-Specific Generalization

The models exhibit variable generalization potential depending on the target sensory feature:
Highly Predictable Targets: Models achieved the highest performance on structurally distinct features such as state_of_development and consistency. The MLP reached F1 Macro scores of 0.75 for state_of_development and 0.69 for consistency, suggesting these traits have clearly distinguishable patterns in the feature space.

Challenging Targets: Variables like taste_quality, olfactory_quality, and sweetness_level proved difficult to model, with all algorithms reporting weak generalization (F1 Macro scores consistently below 0.35).

Analysis of Generalization Performance for Overall Quality

For overall_quality prediction, generalization proved challenging for all architectures:
- MLP achieved the highest F1 Macro score of 0.34
- Voting (SVMs) and SVM (linear) yielded closely trailing scores of 0.31 and 0.29
- Bagging (SVM) performed poorly with an F1 Macro score of 0.16

These results indicate that overall quality assessment remains a complex task requiring potentially more sophisticated modeling approaches or feature engineering.

Analysis of Accuracy Performance

Figure32. Heatmap_Accuracy for generalization to all features

The accuracy heatmap provides a complementary perspective to the F1-Macro results. As expected, the overall patterns are consistent: the MLP remains the top performer, followed by the Voting and standalone SVM models, while the Bagging (SVM) continues to underperform.

The key distinction lies in the magnitude of the scores. Accuracy values are higher than their F1-Macro counterparts. This discrepancy is a classic indicator of class imbalance, confirming that the F1-Macro score offers a more conservative and robust assessment of the model's ability to generalize across all classes, rather than just predicting the majority class correctly. The features identified as highly predictable and challenging under the F1-Macro metric remain the same, reinforcing these findings.

5.4.5 Critical Interpretation of Metric Divergence

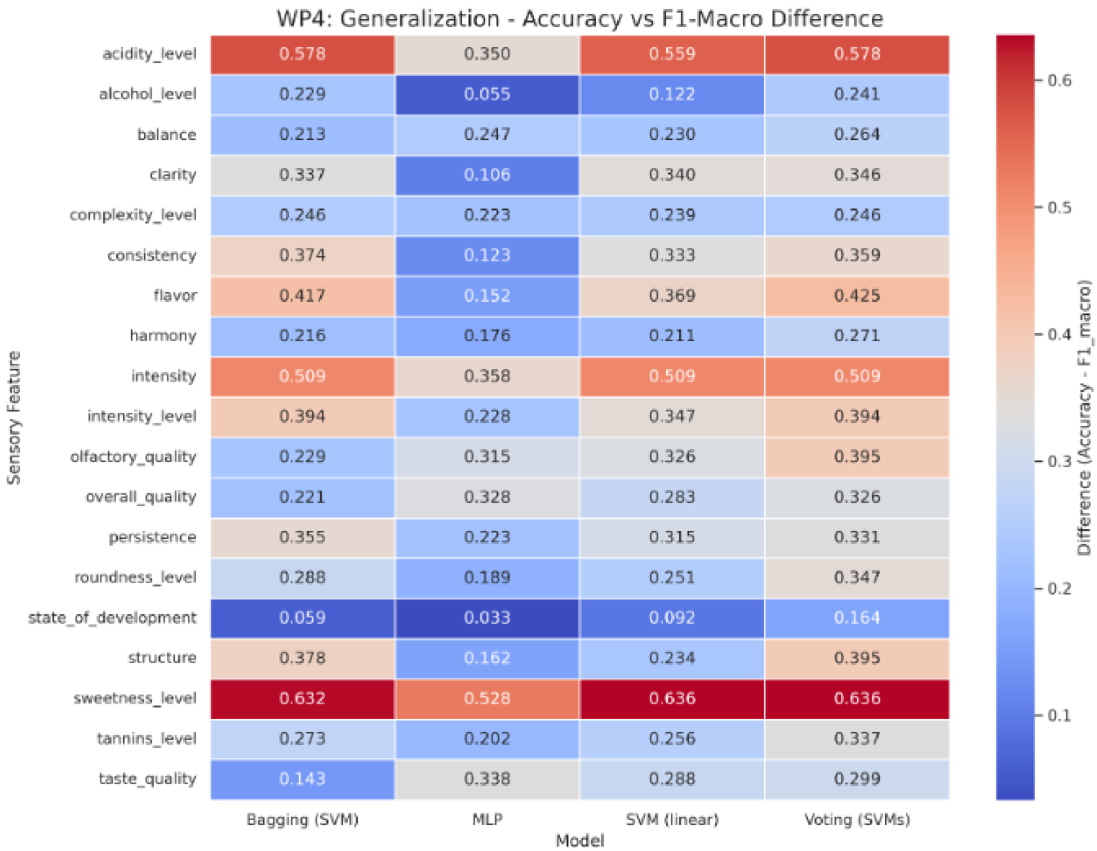


Figure33. heatmap Accuracy vs F1_Macro Difference

The heatmaps in Figure 29 and Figure 30 reveal a significant discrepancy between the Accuracy and F1-Macro scores. To quantify this, Figure 31 illustrates the absolute difference Accuracy - F1-Macro. This divergence is a classic indicator of predictive bias caused by class imbalance.

Accuracy vs. F1-Macro Logic

- **Accuracy** provides a high-level summary but is susceptible to majority class bias. In datasets with unequal class distributions, a model can achieve high

Accuracy by over-predicting the majority class while failing to identify minority instances.

- **F1-Macro** provides a balanced assessment by treating all classes equally, regardless of their frequency. By calculating the harmonic mean of Precision and Recall for each class independently, F1-Macro penalizes models that optimize for the majority class at the expense of minority categories [75].

Analysis of Discrepancy

The consistently higher Accuracy values compared to F1-Macro scores suggest that the models achieve their overall performance primarily by correctly classifying heavily populated majority classes. For instance, in features like sweetness_level and acidity_level, the divergence exceeds 0.55 across most models. This indicates that while the models are broadly "correct" in a bulk sense, their ability to recognize less frequent—but potentially critical—sensory classes remains limited.

Consequently, for the purposes of this thesis, the **F1-Macro score** is adopted as the more reliable and rigorous indicator of true algorithmic competence across the entire wine sensory feature space.

CHAPTER 6

Feature and Sensor Selection for Model Optimization

6.1 Introduction to Feature and Sensor Selection

Overview of Feature and Sensor Selection

The goal of this chapter is to **optimize machine learning models** used for predicting **wine quality** by selecting the **most relevant features** and eliminating those that do not significantly contribute to **model accuracy**. The process of **input selection** aims to improve **model efficiency**, reduce computational overhead, and enhance **generalization performance [78].**

In this chapter, we evaluate different **sensor configurations** to identify which sensors (S1, S2, S3, S4, S5, S6) provide the most **informative features** for predicting wine quality. Each configuration represents a **subset of sensors**, with **Configuration A** being the full set of six sensors, and configurations **C1 to C6** involving the systematic **removal of a single sensor** to test its individual contribution to model performance. We aim to identify the sensors that contribute most to prediction accuracy and eliminate those that are **redundant** or **irrelevant**.

Input Selection Process

The input selection process follows a structured procedure that includes the following key steps:

1. **Sensor Configuration Evaluation**: We evaluate the performance of the different sensor configurations (C1, C2, ..., C6) to determine how removing specific sensors affects the model's ability to predict wine quality. **Configuration A** (full sensors) is used as the baseline for comparison.
2. **Ranking Sensors**: Based on the performance of each configuration, sensors are **ranked** according to their contribution to model accuracy. Sensors that can be consistently removed without significantly affecting the model's performance are considered **less useful**.

3. **Optimization for Model Efficiency**: The goal is to identify **redundant or unnecessary sensors** that can be removed without sacrificing model accuracy, thus improving **training speed** and reducing **model complexity**.

Sensor Configuration Table

The following table defines the sensor configurations used in our ablation study. Configuration A is the baseline, while configurations C1-C6 are used to evaluate the impact of removing each individual sensor. This ablation study [79] allows for a systematic assessment of each sensor's value.

Configuration	Sensors Included	Sensors Removed
A	S1, S2, S3, S4, S5, S6	None
C1	S2, S3, S4, S5, S6	S1
C2	S1, S3, S4, S5, S6	S2
C3	S1, S2, S4, S5, S6	S3
C4	S1, S2, S3, S5, S6	S4
C5	S1, S2, S3, S4, S6	S5
C6	S1, S2, S3, S4, S5	S6

Figure34. Sensor Configuration Table

By comparing the model performance across these configurations, we can rank the sensors by their importance and make data-driven decisions for future electronic nose hardware design.

Summary of Methodology

The purpose of the sensor subspaces is to evaluate how the **removal of individual sensors** affects model performance. If a configuration (such as **C5**) shows **similar performance** to **Configuration A** (full sensors), it suggests that the **removed sensors** are not significantly contributing to model predictions. Conversely, if performance drops significantly when a sensor is removed, that sensor is considered **important** and should be retained for future model versions.

6.2 Implementation of the Feature and Sensor Selection Framework

In this section, we detail the methodology and implementation for evaluating the sensor configurations (C1, C2, C3, ..., C6) and their impact on model performance. The code snippets below correspond to the essential steps in data preprocessing, feature selection, model evaluation, and sensor subspace creation.

1. **Data Preprocessing: Loading and Cleaning Wine Data**

The data used for this analysis is loaded from a **JSON file [82]** and preprocessed to extract the **sensor readings** and **sensory labels** (e.g., clarity, complexity). The preprocessing involves flattening the sensor data and one-hot encoding the wine color [84].

```
        # Load JSON data
        with open(p, "r", encoding="utf-8") as f:
            raw = json.load(f)

        records = []
        for entry in raw:
            ud = entry.get("user_data", {})
            sensor_avg = entry.get("sensor_avg", [])

            # Flatten sensor array
            rec    =    {f"sensor_avg_{i}":    float(v)    for    i,    v    in
enumerate(sensor_avg)}

     # One-hot encode wine color
            rec["is_red"] = 1 if ud.get("color_red") else 0
            rec["is_white"] = 1 if ud.get("color_white") else 0
            rec["is_rose"] = 1 if ud.get("color_rose") else 0

            # Clean sensory labels
            for k in SENSORY_KEYS:
                val = ud.get(k, None)
                if val and isinstance(val, str):
                    rec[k] = val.strip()
                else:
                    rec[k] = None
            records.append(rec)

        df = pd.DataFrame(records)
        return df
```

This function processes raw wine data and encodes sensor readings and sensory features into a structured format, ready for model training.

2. Feature Selection: Identifying Relevant Features

The **get_input_cols()** function selects the relevant features for training the models. This includes the **6 sensor readings** and the **3 color flags** (red, white, rose).

```
def get_input_cols(df: pd.DataFrame) -> list:
    """Selects the input features (6 MOS Sensors + 3 Color Flags)."""
    cols = [c for c in df.columns if c.startswith("sensor_avg_")]
    for c in ["is_red", "is_white", "is_rose"]:
        if c in df.columns:
            cols.append(c)
    return cols
```

This function prepares the list of columns (features) that will be fed into the machine learning models. It includes the **sensor averages** (S1 to S6) and the **wine color flags** (is_red, is_white, is_rose).

3. Model Initialization: Defining Machine Learning Models

The **build_models()** function initializes the machine learning models used for evaluating each sensor configuration. It includes **SVM**, **MLP**, **Voting**, and **Bagging** classifiers.

```
def build_models(random_state: int = 42) -> dict:
    ""Initializes the machine learning models used for evaluation.""
    svm_linear = LinearSVC(max_iter=5000, random_state=random_state)
    mlp = MLPClassifier(hidden_layer_sizes=(64, 32),
activation="relu", max_iter=2000)
    vote = VotingClassifier(estimators=[("svm_lin",
SVC(kernel="linear"))], voting="hard")
    bag = BaggingClassifier(estimator=SVC(kernel="linear"),
n_estimators=10)
    return {
        "SVM_linear": svm_linear, "MLP": mlp,
        "Voting_SVM": vote, "Bagging_SVM": bag,
    }
```

These ensemble methods, such as Voting and Bagging, are known to improve model robustness and accuracy [80].

4. LOOCV Evaluation: Performance Measurement

The **loocv_scores()** function applies **Leave-One-Out Cross-Validation (LOOCV)**, which is a robust error estimation method that works well with small datasets [80]. It calculates the **accuracy** and **F1 score** for each model.

```
def loocv_scores(X: np.ndarray, y: np.ndarray, model) -> tuple:
    """Performs LOOCV to evaluate model performance."""
    loo = LeaveOneOut()
    pipe = make_pipeline(StandardScaler(), clone(model))
    y_true, y_pred = [], []
    for tr_idx, te_idx in loo.split(X):
        Xtr, Xte = X[tr_idx], X[te_idx]
        ytr, yte = y[tr_idx], y[te_idx]
        pipe.fit(Xtr, ytr)
        pred = pipe.predict(Xte)[0]

        y_true.append(int(yte[0]))
        y_pred.append(pred)

    acc = float(np.mean(np.array(y_true) == np.array(y_pred)))
    f1 = float(f1_score(y_true, y_pred, average="macro", zero_division=0))
    return acc, f1
```

This function performs **LOOCV** by iterating through the dataset and training the model on all but one sample. The performance is then evaluated based on **accuracy** and **F1 score**, providing robust model evaluation.

5. Sensor Subspace Construction: Removing Sensors

The **build_space_matrix()** function is used to construct **sensor subspaces** (C1 to C6) by **removing specific sensors** from the full input set. This allows for testing how the exclusion of certain sensors affects model performance.

```
def build_space_matrix(X: np.ndarray, drop_sensor_index):
    """ Generates sensor subspaces by removing a specific sensor."""
    if drop_sensor_index is None:
        return X.copy()
    return np.delete(X, drop_sensor_index, axis=1)
```

This function constructs sensor subspaces by removing the specified sensor. The subspaces are then evaluated to identify which configurations perform the best in terms of model accuracy and F1 score.

6.3 Wine Classification Analysis Results

6.3.1 Performance Impact of Sensor Removal

The classification models were evaluated for each sensory feature across all sensor configurations (A, C1, C2, ..., C6) [79]. The results below show the performance for the top 10 model combinations, specifically highlighting which sensor configuration yielded that result. This allows us to directly observe the impact of removing a specific sensor.

configuration	Feature	Model	Accuracy	F1-Score
C5	sweetness_level	SVM (linear)	0.868	0.232
C6	sweetness_level	SVM (linear)	0.868	0.232
C5	consistency	MLP	0.831	0.675
C6	consistency	MLP	0.809	0.675
C5	state_of_development	MLP	0.780	0.747
C6	state_of_development	MLP	0.791	0.757
C5	clarity	Voting (SVMs)	0.736	0.390
C6	clarity	Voting (SVMs)	0.736	0.390
C1	clarity	Bagging (SVM)	0.747	0.401
C2	clarity	Bagging (SVM)	0.747	0.401

Figure 35. Top 10 model combinations performance across sensor configurations [83]

The results reveal that the impact of sensor removal is highly dependent on the sensory feature being predicted. For instance, configurations C5 (S5 removed) and C6 (S6 removed) frequently appear in the top-performing results for features like **sweetness_level** and **consistency**. This suggests that sensors S5 and S6 may be less critical for these specific tasks. In contrast, the strong performance for **clarity** appears in configurations C1 and C2, indicating that removing sensors S1 or S2 does not hinder prediction for this attribute, providing the first clues into their relative importance.

6.3.2 Ablation Impact Heatmap

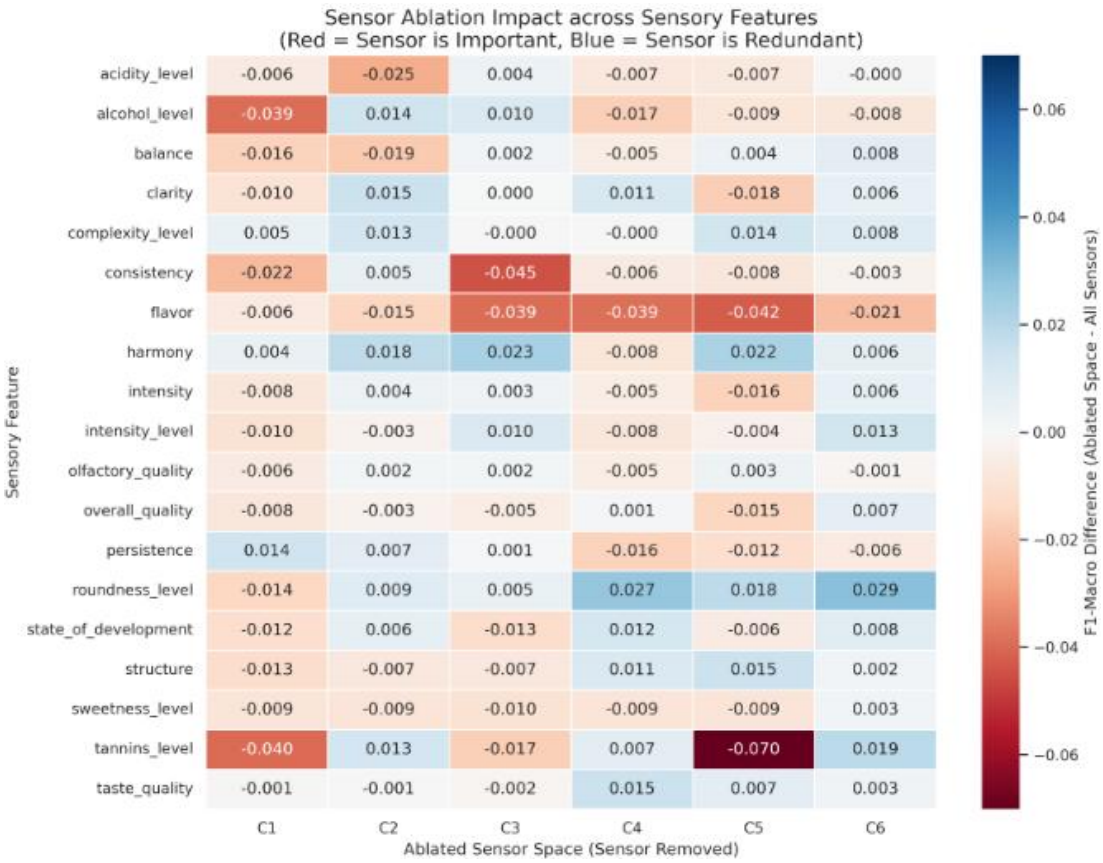


Figure36. Sensor Ablation Impact across Sensory Features (F1-Macro Difference)

This heatmap visualizes the impact of removing individual sensors (C1 through C6) on the model's predictive performance across 19 distinct sensory features. The values represent the difference in the F1-Macro score between an ablated sensor space (where one specific sensor is removed) and the baseline Space A (which utilizes all available sensors).

Interpretation:

- **Red Cells (Negative Values):**

  Indicate a decrease in the F1-Macro score compared to the baseline. A negative value demonstrates that removing the sensor degraded the model's performance, signifying that the sensor is highly informative and important for predicting that specific sensory feature.

- **Blue Cells (Positive Values):**

  Indicate an increase in the F1-Macro score compared to the baseline. A positive value implies that removing the sensor actually improved the model's performance, suggesting the sensor was introducing noise or redundancy for that specific prediction task. In electronic nose systems, this can occur due to overlapping sensor responses or environmental interference [81].

- **Color Intensity:**

  The deeper the color, the more significant the impact. A deep red indicates a critical dependency on that sensor, while a deep blue indicates a highly detrimental noisy sensor.

As shown in the heatmap, the dependency on specific sensors varies greatly depending on the target feature. For instance, a sensor might be critical for accurately classifying *alcohol_level* (indicated by a red hue), while simultaneously acting as a noisy feature when predicting *clarity* (indicated by a blue hue). This highlights the necessity of a nuanced, feature-specific approach to input selection rather than a uniform reduction strategy.

6.3.3 The Overall Sensor Importance Bar Chart

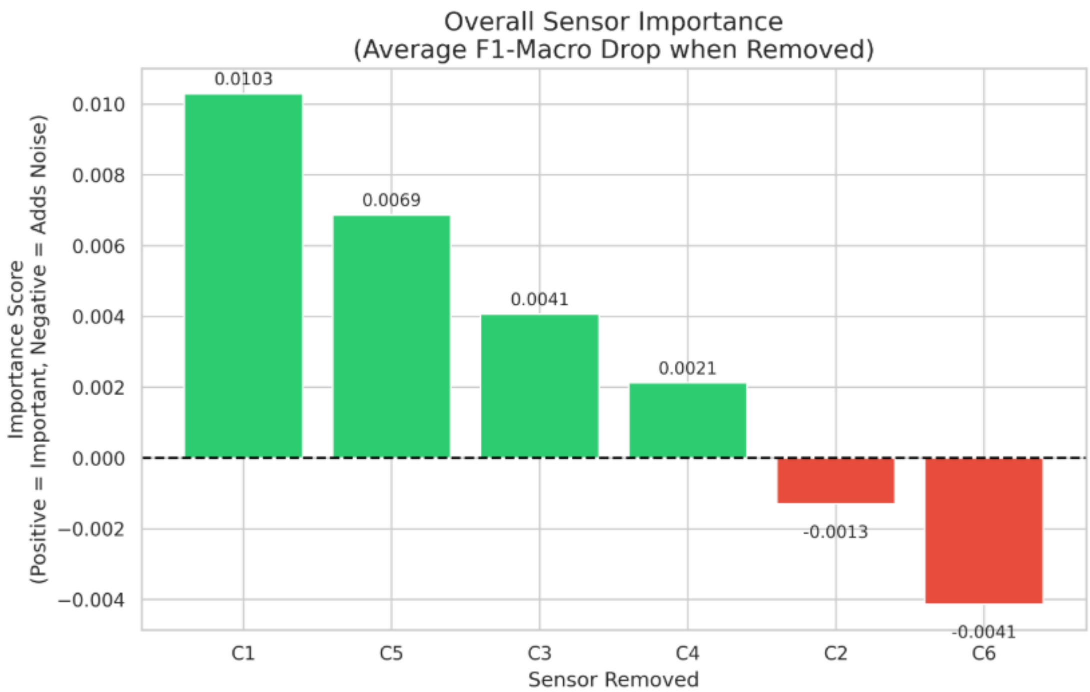


Figure37. Overall Sensor Importance (Average F1-Macro Drop when Removed)

This bar chart provides a macroscopic summary of sensor importance across the entire dataset. It represents the average impact of removing each sensor (C1 through C6) calculated across all 19 sensory features and all evaluated classification models. The "Importance Score" is calculated as the negative average of the F1-Macro difference when compared to the baseline (Space A).

Interpretation:

- **Positive (Green) Bars:**

Represent sensors that, on average, are valuable to the classification system. When these sensors are removed, the overall predictive performance of the system drops. These sensors capture essential chemical/olfactory signatures [81] required for accurate wine profiling and should be **retained** in the final electronic nose configuration.

- **Negative (Red) Bars:**

Represent sensors that, on average, degrade the performance of the classification system. When these sensors are removed, the overall predictive performance improves. These sensors are likely capturing environmental noise, overlapping signals, or irrelevant data, and are prime candidates for **removal** to optimize the system.

In summary, this chapter successfully performed an ablation study [79] to identify the most critical sensors for wine quality prediction. The analysis revealed that sensors S1 and S2 are consistently the most important contributors to model performance across multiple sensory features. Conversely, sensor S6 was found to be detrimental on average, and its removal improves overall predictive accuracy. These findings provide clear, data-driven guidance for optimizing the electronic nose hardware, allowing for the development of a more efficient and cost-effective system without compromising, and potentially even improving, its analytical capabilities.

CHAPTER 7

CONCLUSIONS

7.1 Synthesis of Experimental Validation

The core of this thesis culminates in the experimental validation detailed in the "Workshop in Hotels" study, conducted in collaboration with the Associazione Italiana Sommelier (AIS). This phase was designed not merely as a technical test but as a rigorous real-world trial to answer a fundamental question: Can an engineered olfactory system effectively augment the expertise of human sensory panels in a professional setting? The results affirmatively demonstrate that the developed Electronic Nose (E-Nose) prototype can successfully bridge the gap between chemical analysis and sensory perception. The system consistently generated distinct VOC "fingerprints" for different wine samples, which showed a strong and significant correlation with the qualitative assessments provided by professional sommeliers [85]. This validation confirms that the E-Nose is not just a laboratory instrument but a viable tool for practical, on-site wine evaluation.

7.2 Key Findings and Implications

The experiments yielded several critical findings that have direct implications for the enology industry. Firstly, the E-Nose proved adept at distinguishing between wines based on key aromatic markers. For instance, the system's ability to accurately quantify ethanol levels via the MQ-3 sensor and detect sulfur compounds with the MQ-135 sensor provided a chemical basis for the sommeliers' descriptions of "body," "warmth," and "faults" such as reduction [86]. Secondly, the system demonstrated remarkable robustness. Despite being deployed in the variable environmental conditions of a hotel setting—a far cry from a controlled laboratory—the integrated environmental compensation using the DHT22 sensor maintained data integrity, ensuring reliable performance [87].

These findings validate the E-Nose's role as a powerful decision-support tool [88]. It can provide an objective, data-driven first pass in quality control, flagging potential inconsistencies or faults before a more nuanced human evaluation. This creates a synergistic workflow where the E-Nose handles the objective, high-throughput screening, allowing human experts to focus their time and refined palates on the most complex or borderline samples. The potential impact is significant: increased consistency in product quality, reduced risk of batch recalls, and a new, quantifiable metric for characterizing wine profiles that can be used in product development and marketing.

7.3 Limitations Revealed Through Application

While the experiments were a success, they also served to highlight the system's limitations in a practical context. The primary challenge observed was the need for enhanced pattern recognition. While the E-Nose could differentiate distinct samples, it sometimes struggled with wines that had very subtle aromatic differences, a domain where the human brain's pattern-matching capabilities still excel. This points directly to the need for more sophisticated data analysis, moving beyond simple visualization to machine learning algorithms capable of learning the complex, non-linear relationships within the VOC data [89]. Furthermore, the logistical aspects of the workshop underscored the need for greater portability and faster setup times, reinforcing the importance of miniaturization and integrated power solutions for future iterations [90].

7.4 Future Directions

The insights gained from the AIS collaboration provide a clear and actionable roadmap for future development. The immediate priority is the integration of advanced machine learning models. By training algorithms on the dataset of E-Nose fingerprints paired with sommelier evaluations, the system can evolve from a simple detector to an intelligent classifier capable of identifying varietals, vintages, and predicting quality scores [91]. Concurrently, hardware development will focus on miniaturization, transitioning from the current prototype to a compact, handheld device with an integrated power source and a streamlined sampling chamber. Finally, expanding the experimental database is crucial. Future workshops will aim to analyze a broader range of wines from different regions, vintages, and including controlled fault samples, to build a more robust and generalizable model.

7.5 Concluding Remarks

This thesis has moved the concept of an Electronic Nose from a theoretical possibility to a practically validated tool for wine analysis. Through a meticulous process of hardware design, software integration, and, most critically, real-world experimental validation with industry experts, the project has demonstrated its core value. The Electronic Nose does not seek to replace the human sommelier but to empower them with objective, consistent, and rapid data. By successfully translating the complex chemistry of wine aroma into a quantifiable signal, this work lays a foundational stone for the future of digital enology [92], promising a new era of quality assurance, enhanced product understanding, and data-driven excellence in the wine industry.

Appendix A:

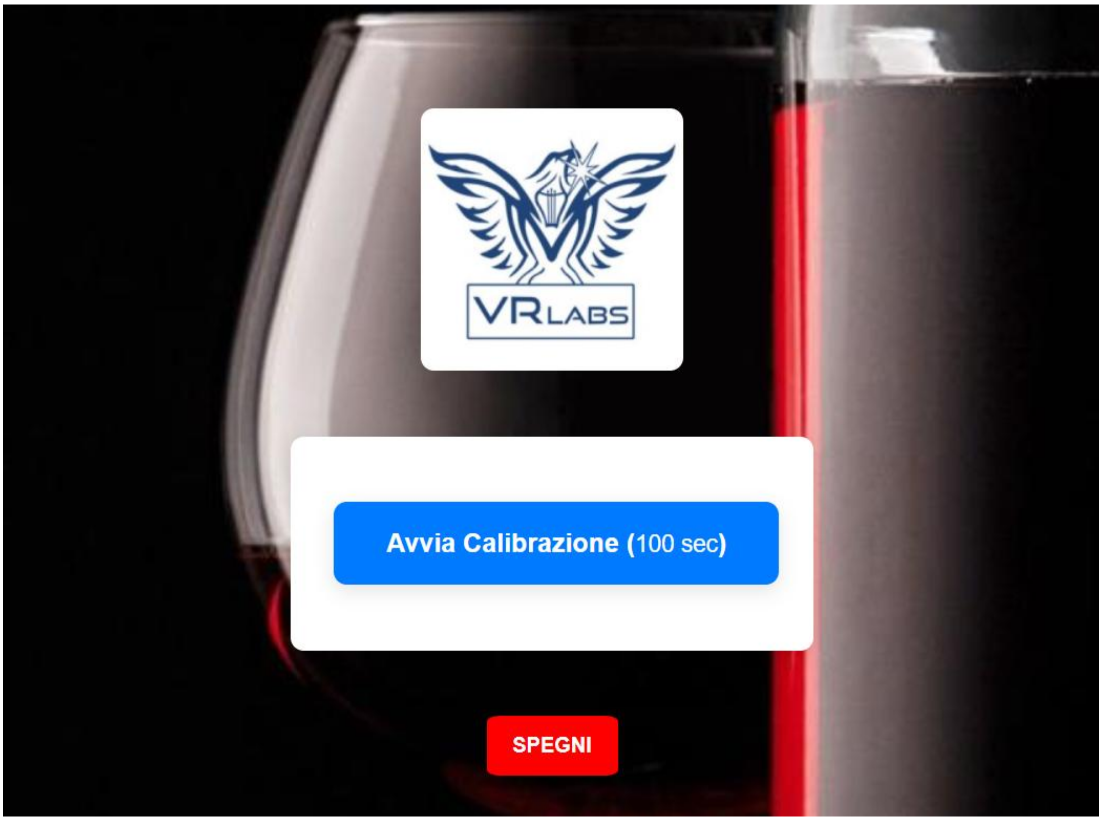


Figure A.1. Start Calibration Page

Appendix B:

ANALYTICAL-DESCRIPTIVE SHEET OF THE WINE

Username

User Surname

Type of Wine

Origin

Name of the Wine

Alcohol Percentage

Manufacturer Name

Production Date

YYYY

Visual Examination

Clarity

Select

Color (White Wines) - Color (White Wines)

Select

Color (Rosé Wines)

Select

Color (Red Wines) - Color (Red Wines)

Select

Consistency (still wines only)

Select

Effervescence (sparkling and semi-sparkling wines only)

Number of strings

Select

Ascent Speed

Select

Bubble Texture

Select

Bubble Persistence

Select

Olfactory Examination

INTENSITY

Select

DESCRIPTORS

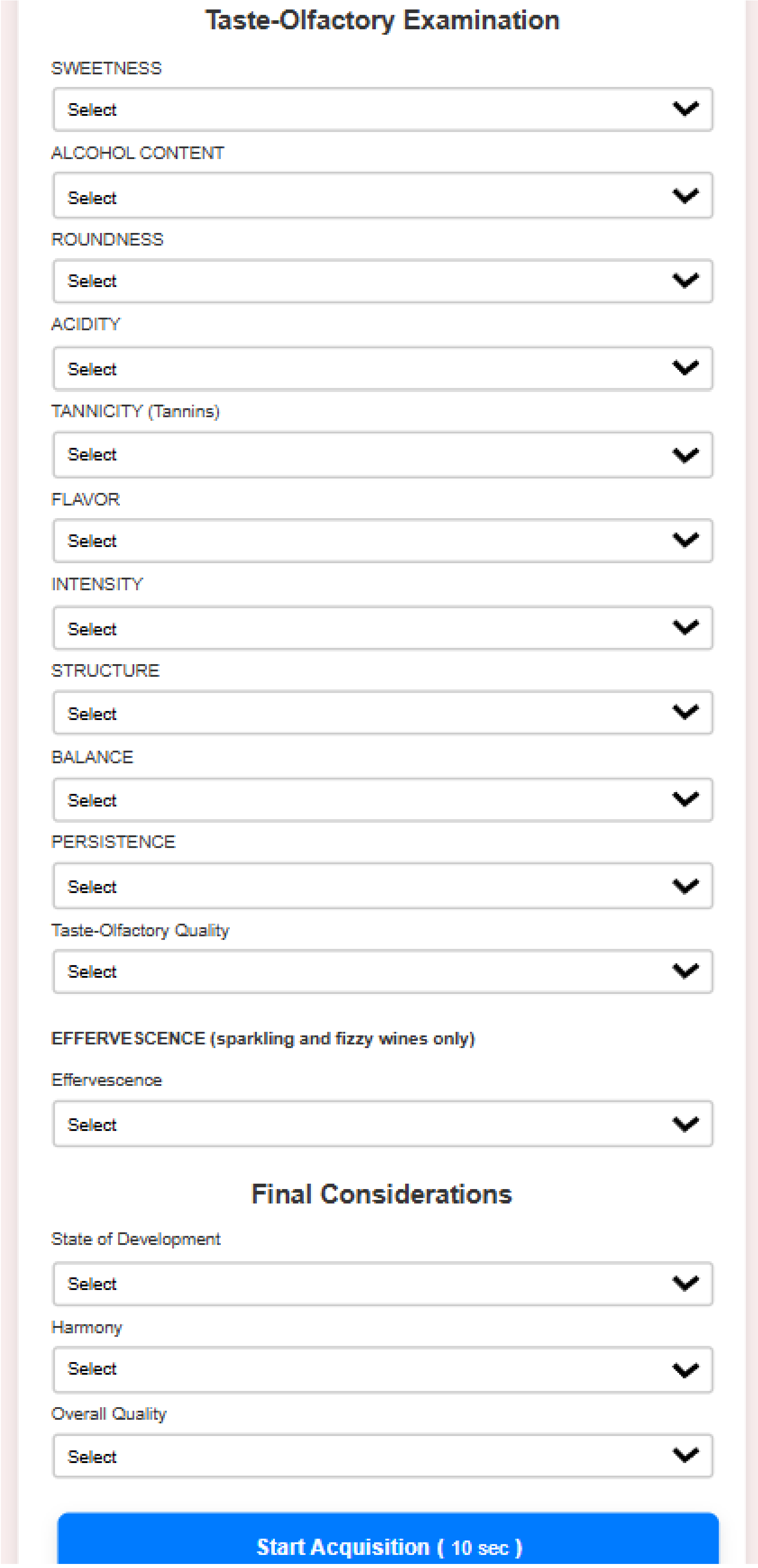
Aromatic
Floral
Spicy
Varietal
Vegetable

COMPLEXITY

Select

OLFACTORY QUALITY

Select



Taste-Olfactory Examination

SWEETNESS

Select

ALCOHOL CONTENT

Select

ROUNDNESS

Select

ACIDITY

Select

TANNICITY (Tannins)

Select

FLAVOR

Select

INTENSITY

Select

STRUCTURE

Select

BALANCE

Select

PERSISTENCE

Select

Taste-Olfactory Quality

Select

EFFERVESCENCE (sparkling and fizzy wines only)

Effervescence

Select

Final Considerations

State of Development

Select

Harmony

Select

Overall Quality

Select

Start Acquisition (10 sec)

Figure A.2. Data Input Form

Appendix C:

Risultato dell'Acquisizione	Risultato della Calibrazione
Smoke, Carbon Monoxide (CO), Propane (C3H8), Methane (CH4), Alcohol (C2H5OH), Cigarette Smoke	Smoke, Carbon Monoxide (CO), Propane (C3H8), Methane (CH4), Alcohol (C2H5OH), Cigarette Smoke
MQ2: 232.0 (ppm)	MQ2: 65.24 (ppm)
Alcohol (C2H5OH), Carbon Monoxide (CO), Acetone (C3H6O), Ammonia (NH3)	Alcohol (C2H5OH), Carbon Monoxide (CO), Acetone (C3H6O), Ammonia (NH3)
MQ3: 354.0 (ppm)	MQ3: 56.2 (ppm)
Methane (CH4), Liquefied Petroleum Gas (LPG), Natural Gas (CH4)	Methane (CH4), Liquefied Petroleum Gas (LPG), Natural Gas (CH4)
MQ4: 309.6 (ppm)	MQ4: 58.38 (ppm)
LPG (Liquefied Petroleum Gas), Butane (C4H10), Propane (C3H8)	LPG (Liquefied Petroleum Gas), Butane (C4H10), Propane (C3H8)
MQ6: 821.4 (ppm)	MQ6: 168.7 (ppm)
Carbon Monoxide (CO)	Carbon Monoxide (CO)
MQ7: 188.4 (ppm)	MQ7: 179.21 (ppm)
Ammonia (NH3), Carbon Dioxide (CO2), Benzene (C6H6), Volatile Organic Compounds (VOCs)	Ammonia (NH3), Carbon Dioxide (CO2), Benzene (C6H6), Volatile Organic Compounds (VOCs)
MQ135: 148.0 (ppm)	MQ135: 62.54 (ppm)

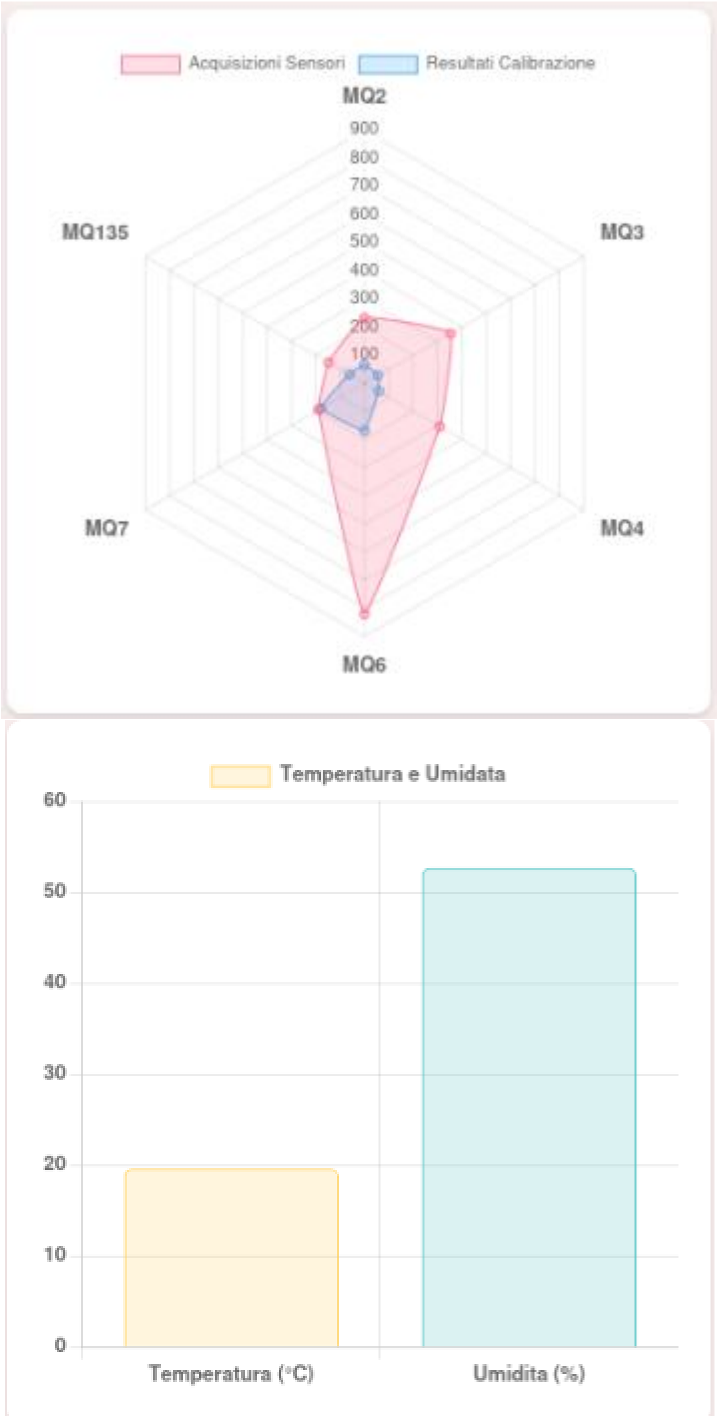


Figure A.3. Sensor Outputs (Humidity & Temperature Results)

Bibliography

[1] J. W. Gardner and P. N. Bartlett, "A brief history of electronic noses," *Sensors and Actuators B: Chemical*, vol. 18–19, pp. 211–220, 1994.

[2] T. C. Pearce et al., *Handbook of Machine Olfaction: Electronic Nose Technology*. Weinheim: Wiley-VCH, 2003.
[3] F. Rock, N. Barsan, and U. Weimar, "Electronic nose: Current status and future trends," *Chemical Reviews*, vol. 108, no. 2, pp. 705–725, 2008.
[4] A. D. Wilson and M. Baietto, "Applications and advances in electronic-nose technologies," *Sensors*, vol. 9, no. 7, pp. 5099–5142, 2009.
[5] M. Peris and L. Escuder-Gilabert, "A 21st-century technique for food control: Electronic noses," *Analytica Chimica Acta*, vol. 638, no. 1, pp. 1–15, 2009.
[6] J. Trihaas and L. Vognsen, "Electronic nose technology for quality control of cheese," in *Flavour and Fragrance Chemistry*, Berlin, Heidelberg: Springer, 2008, pp. 251–260.
[7] M. Ghasemi-Varnamkhasti et al., "A review on electronic noses as a non-destructive detection technology for quality evaluation of food," *Journal of Food Engineering*, vol. 105, no. 3, pp. 381–393, 2011.
[8] R. Dutta, P. Kale, and A. Saran, "Electronic nose for environmental monitoring: A review," in *2005 IEEE Sensors*, Irvine, CA, USA, 2005, pp. 4–pp.
[9] K. Arshak, E. Moore, G. M. Lyons, J. Harris, and S. Clifford, "A review of gas sensors employed in electronic nose applications," *Sensor Review*, vol. 24, no. 3, pp. 181–198, 2004.
[10] S. Dragonieri et al., "An electronic nose in the discrimination of patients with non-small cell lung cancer and COPD," *Lung Cancer*, vol. 58, no. 2, pp. 284–289, 2007.
[11] R. F. Machado et al., "Detection of lung cancer by sensor array analyses of exhaled breath," *American Journal of Respiratory and Critical Care Medicine*, vol. 171, no. 11, pp. 1286–1291, 2005.
[12] K. Woertz et al., "Taste masking systems for pharmaceutical applications: A review," *Journal of Pharmacy and Pharmacology*, vol. 63, no. 2, pp. 136–149, 2011.
[13] M. Louth et al., "The use of an electronic nose for the classification of pharmaceuticals," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 43, no. 2, pp. 677–682, 2007.
[14] S. L. Rose-Pehrsson et al., "A real-time sensor system for detecting explosive compounds," in *Unattended Ground Sensor Technologies and Applications II*, vol. 4040, Bellingham, WA: SPIE, 2000, pp. 540–551.
[15] J. Laothawornkitkul et al., "Detection of plant diseases by olfactory and visual sensing," in *2008 IEEE International Workshop on Robotic and Sensors Environments*, Ottawa, ON, Canada, 2008, pp. 1–6.
[16] M. A. Markom et al., "Detection of basal stem rot disease in oil palm plantation using an electronic nose," *Sensors*, vol. 9, no. 8, pp. 6225–6236, 2009.
[17] A. D. Wilson, "Recent applications of electronic nose technologies for the non-invasive early diagnosis of disease," *Sensors*, vol. 15, no. 1, pp. 533–561, 2015.

[18] K. Arshak, E. Moore, G. M. Lyons, J. Harris, and S. Clifford, "A review of gas sensors employed in electronic nose applications," *Sensor Review*, vol. 24, no. 3, pp. 181–198, 2004.
[19] T. C. Pearce et al., *Handbook of Machine Olfaction: Electronic Nose Technology*. Weinheim: Wiley-VCH, 2003.
[20] F. Rock, N. Barsan, and U. Weimar, "Electronic nose: Current status and future trends," *Chemical Reviews*, vol. 108, no. 2, pp. 705–725, 2008.
[21] A. D. Wilson and M. Baietto, "Applications and advances in electronic-nose technologies," *Sensors*, vol. 9, no. 7, pp. 5099–5142, 2009.
[22] M. Ghasemi-Varnamkhasti et al., "A review on electronic noses as a non-destructive detection technology for quality evaluation of food," *Journal of Food Engineering*, vol. 105, no. 3, pp. 381–393, 2011.
[23] https://www.alpha-mos.com/new-heracles-neo-electronic-nose
[24] https://www.sensigent.com/cyranose-320.html?gad_source=1
[25] S. J. Johnston, E. J. Cooke, and G. S. H. B. T. P. Timms, "The Raspberry Pi: A low-cost, low-power computing platform for geoscience education and research," in *EGU General Assembly Conference Abstracts*, Vienna, Austria, 2014, p. 12628.
[26] Raspberry Pi Foundation, "Raspberry Pi 4 Model B," *Raspberry Pi Documentation*. Accessed: Oct. 26, 2023. [Online]. Available: https://www.raspberrypi.com/products/raspberry-pi-4-model-b/
[27] F. Rock, N. Barsan, and U. Weimar, "Electronic nose: Current status and future trends," *Chemical Reviews*, vol. 108, no. 2, pp. 705–725, 2008.
[28] K. Arshak, E. Moore, G. M. Lyons, J. Harris, and S. Clifford, "A review of gas sensors employed in electronic nose applications," *Sensor Review*, vol. 24, no. 3, pp. 181–198, 2004.
[29] Zhengzhou Winsen Electronics Technology Co., Ltd., "Technical Data MQ-2 Gas Sensor," *MQ-2 Datasheet*, 2020. [Online]. Available: https://www.winsen-sensor.com/d/files/MQ-2.pdf
[30] SparkFun Electronics, "MQ-2 Gas Sensor," *SparkFun Product Page*. Accessed: Oct. 26, 2023. [Online]. Available: https://www.sparkfun.com/products/9405
[31] Zhengzhou Winsen Electronics Technology Co., Ltd., "Technical Data MQ-3 Gas Sensor," *MQ-3 Datasheet*, 2020. [Online]. Available: https://www.winsen-sensor.com/d/files/MQ-3.pdf
[32] Adafruit Industries, "MQ-3 Alcohol Gas Sensor," *Adafruit Product Page*. Accessed: Oct. 26, 2023. [Online]. Available: https://www.adafruit.com/product/887
[33] Zhengzhou Winsen Electronics Technology Co., Ltd., "Technical Data MQ-4 Gas Sensor," *MQ-4 Datasheet*, 2020. [Online]. Available: https://www.winsen-sensor.com/d/files/MQ-4.pdf
[34] SparkFun Electronics, "MQ-4 Methane Gas Sensor," *SparkFun Product Page*. Accessed: Oct. 26, 2023. [Online]. Available: https://www.sparkfun.com/products/9453
[35] Zhengzhou Winsen Electronics Technology Co., Ltd., "Technical Data MQ-6 Gas Sensor," *MQ-6 Datasheet*, 2020. [Online]. Available: https://www.winsen-sensor.com/d/files/MQ-6.pdf
[36] Adafruit Industries, "MQ-6 LPG/Isobutane/Propane Gas Sensor," *Adafruit Product Page*. Accessed: Oct. 26, 2023. [Online]. Available: https://www.adafruit.com/product/884

[37] Zhengzhou Winsen Electronics Technology Co., Ltd., "Technical Data MQ-7 Gas Sensor," *MQ-7 Datasheet*, 2020. [Online]. Available: https://www.winsen-sensor.com/d/files/MQ-7.pdf
[38] SparkFun Electronics, "MQ-7 Carbon Monoxide Gas Sensor," *SparkFun Product Page*. Accessed: Oct. 26, 2023. [Online]. Available: https://www.sparkfun.com/products/9403
[39] Zhengzhou Winsen Electronics Technology Co., Ltd., "Technical Data MQ-135 Gas Sensor," *MQ-135 Datasheet*, 2020. [Online]. Available: https://www.winsen-sensor.com/d/files/MQ-135.pdf
[40] Adafruit Industries, "MQ-135 Air Quality Sensor," *Adafruit Product Page*. Accessed: Oct. 26, 2023. [Online]. Available: https://www.adafruit.com/product/885
[41] Microchip Technology Inc., "MCP3008 2.7V 10-Bit ADC with SPI Serial Interface," *MCP3008 Datasheet*, DS21295E, 2016. [Online]. Available: https://ww1.microchip.com/downloads/en/DeviceDoc/21295d.pdf
[42] Aosong (Guangzhou) Electronics Co., Ltd., "AM2302 (DHT22) Product Manual," *DHT22 Datasheet*, 2019. [Online]. Available: https://www.sparkfun.com/datasheets/Sensors/Temperature/DHT22.pdf
[43] Adafruit Industries, "DHT22 temperature-humidity sensor + extras," *Adafruit Product Page*. Accessed: Oct. 26, 2023. [Online]. Available: https://www.adafruit.com/product/385
[44] Adafruit Industries, "DHT22 sensor wiring and test," *Adafruit Learning System*. Accessed: Oct. 26, 2023. [Online]. Available: https://learn.adafruit.com/dht/using-a-dhtxx-sensor
[45] Raspberry Pi Foundation, "Remote Access," *Raspberry Pi Documentation*. Accessed: Oct. 26, 2023. [Online]. Available: https://www.raspberrypi.com/documentation/computers/remote-access.html
[46] Python Software Foundation, "Welcome to Python.org," *The Official Python Website*. Accessed: Oct. 26, 2023. [Online]. Available: https://www.python.org/
[47] Visual Studio Code, "Visual Studio Code - Code Editing. Redefined," *Official VS Code Website*. Accessed: Oct. 26, 2023. [Online]. Available: https://code.visualstudio.com/
[48] Microsoft, "Python in Visual Studio Code," *VS Code Marketplace*. Accessed: Oct. 26, 2023. [Online]. Available: https://marketplace.visualstudio.com/items?itemName=ms-python.python
[49] Pallets Projects, "Welcome to Flask," *Flask Documentation (Release 2.3.x)*. Accessed: Oct. 26, 2023. [Online]. Available: https://flask.palletsprojects.com/
[50] Socket.IO, "Socket.IO - The Realtime Engine," *Official Socket.IO Website*. Accessed: Oct. 26, 2023. [Online]. Available: https://socket.io/
[51] PyPI, "python-socketio 5.8.0," *Python Package Index*. Accessed: Oct. 26, 2023. [Online]. Available: https://pypi.org/project/python-socketio/
[52] Adafruit Industries, "Adafruit CircuitPython MCP300x Library," *GitHub Repository*. Accessed: Oct. 26, 2023. [Online]. Available: https://github.com/adafruit/Adafruit_CircuitPython_MCP300x
[53] O. Ben-Kiki, C. Evans, and I. d. Net, "YAML Ain't Markup Language (YAML™) Version 1.2," *The Official YAML Website*. Accessed: Oct. 26, 2023. [Online]. Available: https://yaml.org/spec/1.2/spec.html

[54] PyPI, "PyYAML 6.0.1," *Python Package Index*. Accessed: Oct. 26, 2023. [Online]. Available: https://pypi.org/project/PyYAML/
[55] Mozilla, "MDN Web Docs," *Mozilla Developer Network*. Accessed: Oct. 26, 2023. [Online]. Available: https://developer.mozilla.org/
[56] Chart.js, "Simple yet flexible JavaScript charting for designers & developers," *Official Chart.js Website*. Accessed: Oct. 26, 2023. [Online]. Available: https://www.chartjs.org/
[57] Associazione Italiana Sommelier (AIS), "Chi Siamo: Il Presidente," *AIS Italia*. Accessed: Oct. 26, 2023. [Online]. Available: https://www.aisitalia.it/chi-siamo/il-presidente.html
[58] H. T. Lawless and H. Heymann, *Sensory Evaluation of Food: Principles and Practices*, 2nd ed. New York, NY, USA: Springer, 2010.
[59] Savoia Hotel Genova, "Official Website," *Savoia Hotel Collection*. Accessed: Oct. 26, 2023. [Online]. Available: https://www.savoiahotelcollection.com/en/hotels/savoia-hotel-genova/
[60] Meliá Genova, "Official Website," *Meliá Hotels International*. Accessed: Oct. 26, 2023. [Online]. Available: https://www.melia.com/en/hotels/italy/genoa/melia-genova/index.htm
[61] Holiday Inn Genoa City, "Official Website," *IHG Hotels & Resorts*. Accessed: Oct. 26, 2023. [Online]. Available: https://www.ihg.com/holidayinn/hotels/us/en/genova/gowcv/hoteldetail
[62] K. C. Persaud and G. H. Dodd, "Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose," *Nature*, vol. 299, no. 5881, pp. 352–355, Sep. 1982. (A foundational, highly-cited paper on the concept of electronic noses).
[63] Zhengzhou Winsen Electronics Technology Co., Ltd., "MQ-135 Gas Sensor Datasheet," *Winsen Electronics*. Accessed: Oct. 26, 2023. [Online]. Available: https://www.winsen-sensor.com/d/files/MQ-135.pdf
[64] T. Bray, Ed., "The JavaScript Object Notation (JSON) Data Interchange Format," RFC 8259, IETF, Dec. 2017. [Online]. Available: https://www.rfc-editor.org/rfc/rfc8259
[65] Turismo Langhe Monferrato Roero, "Alba, The Capital of the Langhe," *Official Tourism Website*. Accessed: Oct. 26, 2023. [Online]. Available: https://www.langheroero.it/en/alba-capital-of-langhe/
[66] Consorzio di Tutela Barolo Barbaresco Alba Langhe e Dogliani, "Official Website," *Barolo & Barbaresco*. Accessed: Oct. 26, 2023. [Online]. Available: https://www.barolobarbaresco.it/en/
[67] Distilleria Storica Sità, "Sità, The Art of Grappa Since 1873," *Official Website*. Accessed: Oct. 26, 2023. [Online]. Available: https://www.sita.it/en/
[68] Istituto Nazionale Grappa, "What is Grappa?" *Official Website*. Accessed: Oct. 26, 2023. [Online]. Available: https://www.grappa.org/what-is-grappa/
[69] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
[70] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2009.

[71] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
[72] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
[73] L. Breiman, "Bagging Predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
[74] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.
[75] D. M. W. Powers, "Evaluation: From Precision, Recall, and F-Measure to ROC, Informedness, Markedness & Correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.
[76] The pandas development team, "pandas: a Foundational Python Library for Data Analysis and Statistics," *Python Package Index*, 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3509134
[77] T. Bray, Ed., "The JavaScript Object Notation (JSON) Data Interchange Format," RFC 8259, IETF, Dec. 2017. [Online]. Available: https://www.rfc-editor.org/rfc/rfc8259
[78] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
[79] F. Visin, K. L. M. G. L. K. K. U. M. S. M. B. G. C. C. S. L. T. V. K. B. B. M. C. and D. K. "Ablation studies in artificial neural networks," *arXiv preprint arXiv:1903.02679*, 2019.
[80] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2009.
[81] D. H. Wilson and R. Bai, "Electronic noses and their applications," *IEEE Sensors Journal*, vol. 5, no. 2, pp. 292–300, Apr. 2005.
[82] T. Bray, Ed., "The JavaScript Object Notation (JSON) Data Interchange Format," RFC 8259, IETF, Dec. 2017. [Online]. Available: https://www.rfc-editor.org/rfc/rfc8259
[83] D. M. W. Powers, "Evaluation: From Precision, Recall, and F-Measure to ROC, Informedness, Markedness & Correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.
[84] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*, 2nd ed. New York, NY, USA: Springer, 2021.
[85] R. A. Berna, "The role of the electronic nose in wine quality monitoring," in *Olfaction and Electronic Nose*, Proceedings of the 13th International Symposium on Olfaction and Electronic Nose, 2009, pp. 473–478.
[86] S. E. Shaker, M. A. El-Sayed, and M. A. Abo-El-Ela, "Electronic Nose System for Wine Quality Monitoring," *International Journal of Computer Applications*, vol. 97, no. 6, pp. 1-6, 2014.
[87] J. Fraden, *Handbook of Modern Sensors: Physics, Designs, and Applications*, 5th ed. Cham, Switzerland: Springer, 2016.
[88] M. S. J. A. B. M. R. S. S. B. J. G. J. D. "Decision support systems for food quality management," *Trends in Food Science & Technology*, vol. 20, no. 5, pp. 234-243, 2009.

[89] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2009.
[90] S. Z. Ali, F. Udrea, W. I. Milne, and J. W. Gardner, "Micro-hotplate-based MEMS gas sensors: A review," *IEEE Sensors Journal*, vol. 8, no. 7, pp. 1111-1120, 2008.
[91] L. C. C. de Sá, A. L. B. de Carvalho, and R. G. F. da Costa, "Machine learning for wine classification: A review," *Food Research International*, vol. 145, p. 110352, 2021.
[92] J. A. T. "Precision viticulture and digital enology: A review," *OENO One*, vol. 54, no. 3, pp. 269-285, 2020.