

Juha Närhi

CLASSIFICATION OF WHISKIES AND DRINKING GLASSES BASED ON ION MOBILITY SPECTROMETRY MEASUREMENTS

Bachelor's Thesis
Faculty of Information Technology and Communication Sciences
Examiner: Dr. Philipp Müller
May 2025

ABSTRACT

Juha Närhi:
Classification of whiskies and drinking glasses
based on ion mobility spectrometry measurements
Bachelor's Thesis
Tampere University
Bachelor's Programme in Computing and Electrical Engineering
May 2025

Alongside other alcoholic beverages, whiskies require reliable authentication methods to prevent fraud, such as mislabeling and adulteration. To support such authentication efforts, this thesis explored whether data collected from an ion mobility spectrometry (IMS)-based electronic nose (e-nose) could be used to identify different whiskies and determine from which glass the measurement had been taken.

In the measurement setup, the IMS device was placed at the rim of the glass to mimic a real-world sniffing situation, where volatile compounds from the whisky evaporate and are typically perceived by the human nose. Three different whiskies and three different types of glasses were used, resulting in a total of nine whisky-glass combinations. The use of different glasses was motivated by the aim of investigating whether the shape of the glass affects the IMS response and thereby the classification accuracy. Each combination was measured for ten minutes under controlled conditions, and the measurements were repeated five times.

After data collection, the data was preprocessed and then multiple machine learning algorithms, including a Long Short-Term Memory (LSTM) neural network, Support Vector Classification (SVC), and Linear Discriminant Analysis (LDA), were applied to perform the classification. The validation was performed using session-based cross-validation.

The highest classification accuracies were obtained using the Support Vector Classifier (SVC): 87% for whisky identification, 69% for glass identification, and 67% for classifying whisky-glass combinations. While the glasses generated somewhat distinctive IMS responses, the classification of whiskies proved to be consistently more reliable and accurate than the identification of the glasses. The results were encouraging and provide a foundation for future research and potential real-world applications, if classification accuracy can be further improved.

Keywords: electronic nose, ion mobility spectrometry, machine learning, whisky classification

The originality of this thesis has been verified using the Turnitin Originality Check service.

TIIVISTELMÄ

Juha Närhi:

Viskien ja juomalasien luokittelu ioniliikkuvuuspektrometria -mittausten perusteella

Kandidaatin tutkielma

Tampereen yliopisto

Tieto- ja sähkötekniikan kandidaattiohjelma

Toukokuu 2025

Kuten muutkin alkoholijuomat, myös viskit tarvitsevat luotettavia aitouden varmistusmenetelmiä väärinkäytösten, kuten tahallisesti väärin merkittyjen tai laimennettujen viskien havaitsemiseksi. Näiden tarpeiden myötä tässä kandidaatintyössä tutkittiin, että voidaanko ioniliikkuvuuspektrometriaan (IMS) perustuvan sähköisen nenän (e-nose) keräämän datan avulla tunnistaa erilaisia viskejä ja määrittää, mistä lasista mittausta on otettu.

Mittausjärjestelyssä IMS-laite asetettiin lasin reunalle jäljittelemään todellista haistamistilannetta, jossa viskistä haihtuvat yhdisteet päätyvät normaalisti ihmisen nenän aistimiksi. Tutkimuksessa käytettiin kolmea eri viskiä ja kolmea erityyppistä lasia, mikä johti yhdeksään eri viski-lasi-yhdistelmään. Eri lasien käyttö perustui tavoitteeseen selvittää, onko lasin muodolla vaikutusta IMS-vasteeseen ja sitä kautta luokittelutarkkuuteen. Kutakin yhdistelmää mitattiin kymmenen minuutin ajan kontrolloiduissa olosuhteissa, ja mittaukset toistettiin viisi kertaa.

Mittauksen jälkeen data esikäsiteltiin, jonka jälkeen käytettiin useita koneoppimisalgoritmeja, kuten Long Short-Term Memory (LSTM) -neuroverkkoa sekä Support Vector Classification (SVC) - ja Linear Discriminant Analysis (LDA) -menetelmiä luokittelun suorittamiseksi. Tulosten validoimiseksi käytettiin 5-kertaista ristiinvalidointia (5-Fold Cross-Validation).

Parhaisiin luokittelutarkkuuksiin päästiin SVC:llä: 87 % viskien tunnistuksessa, 69 % lasien tunnistuksessa ja 67 % viski-lasi-yhdistelmien tunnistuksessa. Vaikka eri lasit tuottivat jossain määrin erottuvia IMS-vasteita, viskien luokittelu osoittautui helpommaksi ja tarkemmaksi kuin lasien tunnistaminen. Tulokset olivat rohkaisevia ja luovat pohjan tulevalle tutkimukselle sekä mahdollisille käytännön sovelluksille, edellyttäen että luokittelun tarkkuutta saadaan edelleen parannettua.

Avainsanat: sähköinen nenä, ioniliikkuvuuspektrometria, koneoppiminen, viskien luokittelu

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin Originality Check -ohjelmalla.

USE OF AI IN THESIS

I have utilized AI tools in my thesis:

- No
 Yes

The AI tools utilized in my thesis and their purposes are described below:

Names and versions of AI tools:

ChatGPT -4o
ChatGPT -4.5
ChatGPT -o1

Purpose of using AI tools:

AI tools were used to improve the clarity, consistency, and correctness of the thesis text. ChatGPT's different models were applied to refine sentence structure, correct grammar, and suggest appropriate terminology for both scientific and technical expressions. Additionally, AI was used to translate individual sections between Finnish and English, ensuring that the language was accurate and suitable for academic purposes. These tools were not used to generate content or perform analysis, but rather to support the writing and editing process. Artificial intelligence was also utilized during coding to detect syntax errors and to improve overall efficiency in the development process.

Sections where AI tools were used:

AI was used in all sections.

I acknowledge that I am fully responsible for the entire content of my thesis, including the parts generated by AI, and accept accountability for any violations of ethical standards in publications.

PREFACE

The thesis project was conducted in collaboration with Professor Veikko Surakka and Dr. Jari Tuominen, under the supervision of Dr. Philipp Müller. Dr. Tuominen provided the whiskies and glasses used in the experiment, and the measurements were carried out at the ScentTech laboratory at Tampere University, managed by Professor Surakka. I would like to thank everyone who contributed to this project.

Tampere, 14 May 2025

Juha Närhi

CONTENTS

1 INTRODUCTION	1
2 RELATED WORK.....	3
2.1 E-nose Systems	3
2.2 Food and drink identification using IMS.....	4
2.3 Whisky authentication.....	5
3 METHODOLOGY	6
3.1 Ion mobility spectrometry	6
3.2 Measurement setup and data collection.....	7
3.3 Data preprocessing.....	9
3.4 Machine learning techniques.....	10
3.4.1 Feed forward neural network.....	10
3.4.2 Recurrent neural networks and LSTM	12
3.4.3 Linear Discriminant Analysis	14
3.4.4 Five-fold cross-validation	16
4 RESULTS AND DISCUSSION.....	17
4.1 LSTM approach.....	17
4.2 Other approaches	18
4.3 Combined results	19
4.4 Analysis and discussion.....	21
4.4.1 Impact of Sample Variation on Performance	22
4.4.2 The Role of Measurement Conditions in Result Accuracy	24
5 CONCLUSION AND OUTLOOK.....	28
REFERENCES	29

LIST OF FIGURES

3.1	IMS signal from ChemPro 100i, 40% ethanol	7
3.2	Measurement setup	8
3.3	Feedforward neural network structure	10
3.4	Single neuron structure	11
3.5	Basic RNN architecture	12
3.6	LSTM cell	13
3.7	LDA example with three classes	15
4.1	Confusion matrix for LSTM	18
4.2	IMS signal: Jim Beam Rye, SAVU glass	22
4.3	IMS signal: Kyrö Malt Rye, Glencairn	23
4.4	Absolute humidity, all sessions	25
4.5	Relative humidity, all sessions	25
4.6	Device's internal temperature, all sessions	26
4.7	Air pressure, all sessions	26

LIST OF TABLES

4.1	Final classification accuracies	20
-----	---------------------------------	----

LIST OF ABBREVIATIONS

IMS	Ion Mobility Spectrometry
E-nose	Electronic Nose
VOC	Volatile Organic Compound
GC	Gas Chromatography
MS	Mass Spectrometry
KNN	k-Nearest Neighbor
QDA	Quadratic Discriminant Analysis
SVC	Support Vector Classification
CCS	Collision Cross Section
PCA	Principal Component Analysis
LDA	Linear Discriminant Analysis
LSTM	Long Short-Term Memory
RNN	Recurrent Neural Network

Whisky-Glass Combination Abbreviations

JG	Jim Beam Rye + Glencairn glass
JS	Jim Beam Rye + SAVU glass
JT	Jim Beam Rye + Tumbler glass
KG	Kyrö Malt Rye + Glencairn glass
KS	Kyrö Malt Rye + SAVU glass
KT	Kyrö Malt Rye + Tumbler glass
SG	Kyrö Wood Smoke Rye + Glencairn glass
SS	Kyrö Wood Smoke Rye + SAVU glass
ST	Kyrö Wood Smoke Rye + Tumbler glass

1 INTRODUCTION

Among various distilled spirits, the whisky industry is experiencing rapid growth, with the global market expected to reach USD 71.66 billion by 2025, according to a recent market analysis [1]. At the same time, with the prices of the most expensive whisky bottles reaching millions, it's clear that the industry attracts criminal activity. Fraudulent practices, such as counterfeit bottles, adulteration, and age falsification, cause significant harm to both producers and consumers. For maintaining consumer trust and securing the long-term success of the industry, ensuring the authenticity of whisky is crucial.

Traditionally, the identification of whisky has been based on expert sensory judgment, which by nature is subjective, time-consuming, and expensive. As a result, there is a growing need for a scientific approach to authenticate whiskies more consistently and efficiently. The most successful methods to date have utilized spectroscopic, chromatographic and novel technologies, yielding promising results [2]. However, these techniques involve equipment-related challenges, mainly because the analyses are time-consuming and costly, and must be conducted in laboratories. In addition, the devices are often complex to operate [2]. Therefore, there is potential for exploring ion mobility spectrometry (IMS) as a new tool for whisky authentication.

In this study, IMS is used as part of an electronic nose (e-nose), a technology designed to simulate olfactory sensing by detecting and analyzing chemical compounds in gas or vapor samples [3]. IMS separates ions based on their mobility in an electric field and produces a characteristic mobility spectrum, which enables the indirect identification of compounds [4, p. 5]. This capability makes IMS particularly suitable for e-nose applications such as food and beverage quality control, environmental monitoring, and medical diagnostics [3].

The aim of this work is to explore whether whiskies can be classified based on IMS data, and similarly to investigate whether it is possible to determine from which glass the sample was taken. Three different types of glasses are used in this study: a flat-bottomed Tumbler glass, a Glencairn glass, which is traditionally used in whisky tastings due to its tulip shape, and a SAVU glass, known for its unique structure that removes the ethanol odor and enhances the aroma of whisky. The whiskies used in this study are Jim Beam Rye, Kyrö Malt Rye, and Kyrö Wood Smoke Malt Rye.

This thesis is divided into six chapters. *Related work* reviews e-nose systems, the use of IMS technology in food and drink identification, and whisky authentication. *Methodology*

presents the theory of IMS and its use in this research, the measurement setup, data collection process, preprocessing steps, and the machine learning techniques applied. *Results and discussion* reports the findings of the study and analyses their implications. *Conclusion and outlook* summarizes the main conclusions and suggests future research directions.

2 RELATED WORK

The origins of IMS trace back to the mid-1960s, when early developments in the United States focused on chemical monitoring applications, particularly for military use [4, p. 22]. The technology saw significant advancements in the 1970s and 1980s, especially for security and military applications, as IMS proved effective in the detection of chemical warfare agents and explosives in the field [5]. From the 1990s onward, its use expanded into civilian sectors, including food and pharmaceutical analysis [4, p. 26], [6]. Today, IMS is a widely respected and utilized technique for conducting rapid and sensitive chemical analyses across various research fields.

In addition to exploring IMS applications for beverage and food identification, this section discusses various types of e-noses and their applications, as well as different methods for whisky authentication. In this study, an IMS device is used alone, but in many other cases, IMS devices are used alongside other instruments like gas chromatography (GC) and mass spectrometry (MS) devices [7]. When IMS data is combined with machine learning techniques such as k-nearest neighbor (KNN) classifiers or neural networks, it has shown impressive accuracy in identifying foods and beverages. For instance, in a study by Minaev et al. classifying food scents from either an open plate or a sealed jar, IMS data yielded optimal results when paired with algorithms such as Quadratic Discriminant Analysis (QDA), C-Support Vector Classification (SVC), and MLPClassifier. Misclassification rates were very low: only 1.004 % for QDA, 1.168 % for MLPClassifier with a single hidden layer of 58 neurons, and 2.434 % for SVC [8]. These findings, among others, illustrate that IMS data, when processed with appropriate machine learning methods, is a valuable tool for authenticating foods and beverages.

2.1 E-nose Systems

All the following, eNose, odor sensor, aroma sensor, mechanical nose, and artificial nose, refer to the same type of technology: systems designed to mimic the human sense of smell [9]. While the term e-nose encompasses various technologies, common implementations include surface acoustic wave (SAW) sensors, quartz crystal microbalance (QCM) sensors, metaloxide semiconductor (MOS) sensors, GCs and MSs [3,9,10]. Additionally, there are few bio-electronic and IMS-based e-noses available, which have been utilized in specific applications, for example in [11] and [12].

In the food and beverage industry, e-noses are used for quality control, authenticity verification, and spoilage detection [3,9]. Healthcare applications include the use of e-noses for early disease diagnosis, such as detecting diseases like lung cancer and tuberculosis. Environmental applications focus on air quality monitoring and pollutant detection, while in security, e-noses are applied to detect explosives and narcotics [9]. As e-nose technology continues to evolve, improvements in sensitivity and accuracy enhance its applications across various fields.

2.2 Food and drink identification using IMS

In food and beverage analysis using IMS, the focus is primarily on safety-related aspects such as food composition, process control, authentication, adulteration, and overall food safety. In this field, IMS is rarely used as a stand-alone technique but is commonly coupled with other analytical instruments [13].

In the context of food composition analysis, IMS is used to detect volatile organic compounds (VOCs), such as lipids, peptides, phenolic compounds, or terpenes present in foods and beverages. The aim may, for example, be to detect allergens [13]. Although IMS is often paired with other analytical instruments, in at least one well-known study in food composition context it was applied as a stand-alone technique without auxiliary methods such as liquid chromatography (LC) or GC: The research by Browne et al. [14], where IMS was used to independently detect and identify volatile sugar alcohol sweeteners in commercial chewing gum. In process control, IMS can be utilized as a tool to monitor fermentation in beer production or the roasting process of coffee [15]. IMS is particularly well suited for this purpose because it enables rapid compound identification, making it suitable for online use as part of automated quality control systems [13].

In food authentication, the focus is often on IMS-generated chemical fingerprints, which reflect the VOC profile of a product and can be used to classify different food products. GC-IMS-based fingerprints, when paired with chemometric analysis, have shown promising results in the authentication of oils, meat products, wines, and honey [13].

One noteworthy IMS-based tool for improving food safety is the collision cross section (CCS), a single descriptor that reflects the size of an ion in the gas phase and depends on its charge [7]. Since each molecule typically exhibits a characteristic and reproducible CCS value, it can serve as a reliable parameter for compound identification. To make use of CCS in practical applications, large CCS databases have been compiled, enabling the identification of molecules in food and beverage samples.

2.3 Whisky authentication

A key focus in whisky authentication is the identification of chemical markers that indicate origin, brand, and quality [16]. Compounds such as phenols, esters, haloanisoles, and γ -lactones play significant roles in establishing chemical fingerprints unique to each whisky type. For instance, Scottish whiskies tend to have higher concentrations of phenolic compounds, particularly o-, m-, and p-cresols, compared to other whisky types [16]. A variety of approaches have been developed to detect the distinguishing features of whiskies, with spectroscopic, chromatographic, and novel technologies being among the most prominent [2].

Several studies have demonstrated that whisky classification can be performed with very high accuracy. For example, Macias et al. [17] developed a bimetallic nano-plasmonic tongue and applied linear discriminant analysis (LDA) to classify seven commercial whiskies, deionized water, 40% ethanol in water, and Absolut® vodka, achieving a classification accuracy of 100% and 99.7% with two different tongues. In another study, Liu et al. [18] proposed a multiscale wavelet kernel regularization-based feature extraction method for an e-nose system. Using four commercial gas sensors, they identified four different whiskies with a classification accuracy of 92.00%. Furthermore, Martins et al. [19] combined UV–VIS spectroscopy with PLS-DA analysis to distinguish authentic and counterfeit whisky samples, achieving classification accuracies of 98.6% and 93.1%.

The study that most closely aligns with this thesis is the work by Zhang et al. [20], *“The Use of Electronic Nose for the Classification of Blended and Single Malt Scotch Whisky”*, in which the team used their custom-developed e-nose prototype, NOS.E, to first classify six whiskies in a laboratory setting and later two whiskies in a field test. The samples were first measured using NOS.E, after which the data was analyzed using several machine learning methods, including LDA, SVM, KNN, Bagged Trees, and Subspace Discriminant. The best-performing classifiers from the lab tests were combined into an ensemble classifier, with decisions made via weighted voting. This ensemble-equipped NOS.E system was then used in the field test involving two whiskies, achieving classification accuracies of 96.15%, 100%, and 92.31% for brand, region, and whisky style, respectively. It is worth noting that earlier, when more whiskies were included in the lab experiments, individual classifiers showed lower accuracy. For example, in brand identification, accuracies were 82.05% for SVM, 74.36% for Bagged Trees, 70.51% for Subspace Discriminant, 69.23% for LDA, and 61.54% for KNN. Region and style classification accuracies were also lower, though not as much: they ranged from 100% to 92.31% for region and from 93.59% to 82.05% for style, depending on the classifier used.

3 METHODOLOGY

In this chapter, the main stages of the study are outlined. In Section 3.1, a general introduction is provided to the IMS device and the data it produces. The implementation of the study is then discussed in Section 3.2, followed by the data processing steps. After preprocessing, machine learning techniques are applied for classification, as described in Section 3.4. Finally, the five-fold cross-validation method employed to validate the study is presented.

3.1 Ion mobility spectrometry

Ion Mobility Spectrometry (IMS) is an analytical method used to separate and detect ions based on their mobility in an electric field [4, p. 3]. After ionization, the ions are injected into the drift region and move through the electric field toward the detector at velocities that depend on their mobility [4, p. 4]. This mobility is influenced by factors such as the ion's size, shape, charge, and interactions with the neutral gas flowing in the tube [4, p. 8-9]. The time required for the ions to move from one end of the tube to the other is known as drift time and it is characteristic of each compound [4, p. 4], as also described in [21].

Among other IMS devices, Environics' ChemPro100i used in this study features a multi-channel setup. The device records data from 16 electrodes referred to as channels, but since two of them are used as control channels (which should remain at zero), the remaining 14 are used for sensor measurements. The measurement values are the raw ion currents detected by each electrode, where each current reflects the ions having a specific range of drift times and charge polarity [22]. In other words, each channel's numerical output directly represents the signal intensity of ions whose flight times lie within that channel's window [22]. In addition to these IMS values, the device monitors environmental parameters such as temperature, humidity and air pressure, some of which can significantly affect ion mobility [4, p. 29]. In the present study, only the absolute signal intensities (i.e., the raw current readings from the sensor channels) and the measured environmental factors are utilized. The final classification is performed based on these measurements even though ChemPro100i also offers relative IMS readings. Figure 3.1 illustrates an example of data from the ChemPro100i's third absolute-value channel, showing the ion signal intensity as a function of time during the measurement.

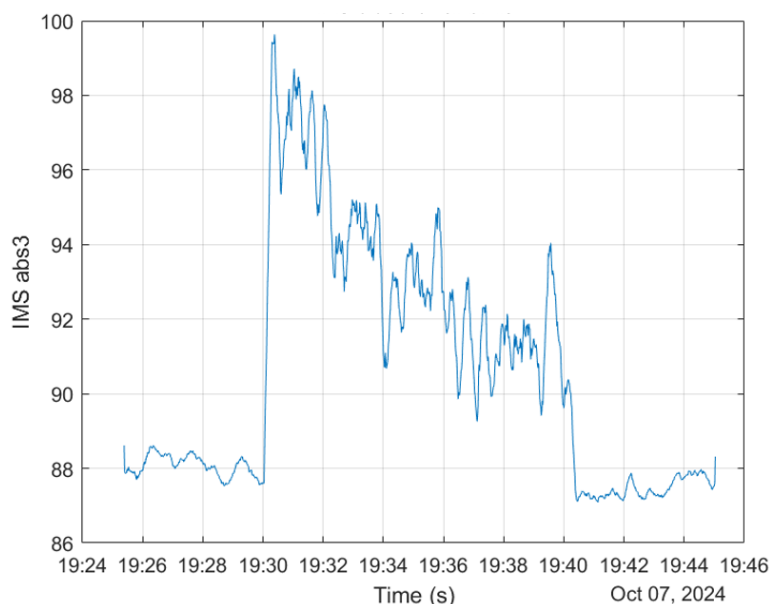


Figure 3.1. IMS abs channel 3 signal from ChemPro 100i during exposure to 40% ethanol between 19:30 and 19:40.

In IMS devices, it is characteristic that when the device is exposed to a certain compound, the signal intensity rapidly increases to a new level. This is referred to as the transient phase, as seen in the image at 19:30 minutes. After the transient phase, the system transitions to the stable phase. However, in some cases the so-called “stable” phase can be highly unstable, as shown in Figure 3.1 between approximately 19:31 and 19:40.

3.2 Measurement setup and data collection

In the first phase of the study, measurement data was collected. The measurements were conducted using a 90-degree tube attached to the IMS device, positioned approximately 1–2 cm above the rim of the glass and about 1 cm inward towards the center of the glass, as shown in Figure 3.2. The aim was to closely mimic an authentic whisky-sniffing scenario, replicating how a person would naturally smell the aroma. The material of the tube, copper, was chosen based on which option would cause the least difference in response and ensure the most stable response during measurements. Each glass was consistently filled with 1 cl of whisky to ensure uniform conditions across all measurements.

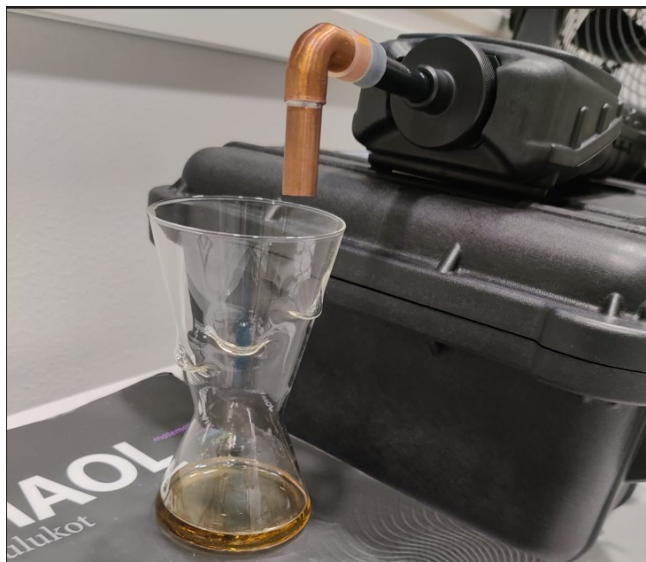


Figure 3.2. The measurement setup.

Each session began with a 5-minute baseline measurement of ambient air to establish a reference point for sensor readings. Following this, a single whisky-glass combination was measured for 10 minutes, allowing sufficient time to capture the chemical profile of the sample. After each measurement, the glass was removed, and the sensor was exposed to ambient air for 5 minutes to allow the response to stabilize and return to baseline levels before proceeding to the next sample. The whisky was poured into the glass immediately before the start of each measurement, ensuring minimal delay between pouring and data collection. This was essential, as the IMS device's response changes over time—making it crucial that each measurement began consistently, right after pouring. This change over time is caused by VOCs in the whisky, which evaporate at different rates depending on their boiling points [23].

This cycle was repeated systematically for nine whisky-glass combinations, including three types of whisky: Jim Bean Rye, Kyrö Malt Rye, and Kyrö Wood Smoke Malt Rye. They were paired with three different glass types: SAVU glass, Glencairn glass, and Tumbler glass. Each combination underwent the same measurement and reset protocol to ensure consistent conditions and minimize cross-contamination between samples. The session concluded with a final 5-minute ambient air measurement. This systematic approach ensured that each whisky-glass combination was analyzed under comparable conditions, providing reliable and reproducible data for further analysis.

This measurement protocol was repeated a total of five times on five different days to account for day-to-day variations, resulting in 45 samples overall. This means that for each whisky-glass combination, five samples were obtained.

3.3 Data preprocessing

After collecting the raw data, it was preprocessed session-wise before applying machine learning techniques. By session-wise it is meant that the following preprocessing steps were applied once to the entire measurement session, which contained nine glass–whiskey combination samples. Since there were five sessions in total, the preprocessing was performed five times. The primary goal of preprocessing the data was to make the results more reliable and to enhance the classification accuracy.

First, in all experiments, the data was detrended to eliminate long-term drift or baseline shifts that could otherwise interfere with pattern recognition. Detrending was performed using SciPy library's detrend function [24]: both the linear (type='linear') and constant (type='constant') modes were tested, but the constant option, which removes the baseline offset by subtracting the sample mean, yielded higher accuracies and was consequently selected for use throughout the experiments.

The second step was to normalize the data to ensure that the measurements were on a comparable scale. In this study, several normalization methods were tested to evaluate their impact on the outcome, but min-max normalization ultimately proved to be the best and most suitable technique, and was therefore chosen for use in all experiments. These two preprocessing steps were performed in all experiments, and in most cases, the data was also smoothed using Savitzky–Golay smoothing. In the Long Short-Term Memory (LSTM) neural network approach, however, smoothing was omitted, as it may alter the temporal characteristics of the signal in a way that could affect sequence-based learning.

After preprocessing, the data was split differently and then compared to each other to get the best results. First, attempts were made to test entire 10-minute samples. Since the IMS device operated at a measurement frequency of 1 hertz, each 10-minute sample contained 600 data points. With 14 different non-zero channels, this resulted in a 600×14 data matrix per sample. That means, 45 samples were used (5 sessions \times 9 combinations of glass and whiskey). The second way was to split these 10 min samples into 60 second samples, resulting in 450 of size 60×14 matrices. However, the best results were achieved using feature vectors. They were formed as follows: From each channel, features such as mean value, standard deviation or max value were extracted. This resulted in feature vectors of length 42 (3 features \times 14 channels). Additionally, Principal Component Analysis (PCA)-transformed data that explained 95% of the total variance were employed in some of the experiments. With PCA, the original 600×14 matrix was reduced to a length of 1200.

Due to the experimental nature of the study, many different approaches were explored to improve the results. These included, for example, using only one or a few channels as input, removing detrending and normalization either entirely or partially, testing various types of features, and extracting features from smaller input segments. Specific time intervals were also tested individually, such as using only the 4–8-minute interval.

3.4 Machine learning techniques

After data preprocessing, the machine learning algorithms were employed to perform the classification. There are many different methods available for analyzing time-series data produced by an IMS device including traditional tools and more developed algorithms such as neural networks. The algorithms used in this study were LSTM neural network, Linear Discriminant Analysis (LDA), KNN, Support Vector Machine (SVM), and Quadratic Discriminant Analysis (QDA). This section focuses on LSTM and LDA, as they have been shown to be effective in similar types of classification tasks, for example in [25] and [26]. The concept of LSTM is introduced by first providing background information on neural networks and recurrent neural networks (RNNs). The theoretical background of the other methods is not discussed in detail in this thesis; however, further information can be found in the following sources: [27] KNN, [28] SVM, [29] QDA.

3.4.1 Feed forward neural network

A neural network is a type of artificial intelligence that imitates the way the human brain works by forming connections between processing units analogous to biological neurons [30]. It consists of layers of interconnected artificial neurons that process input data, learn from it, and produce an output.

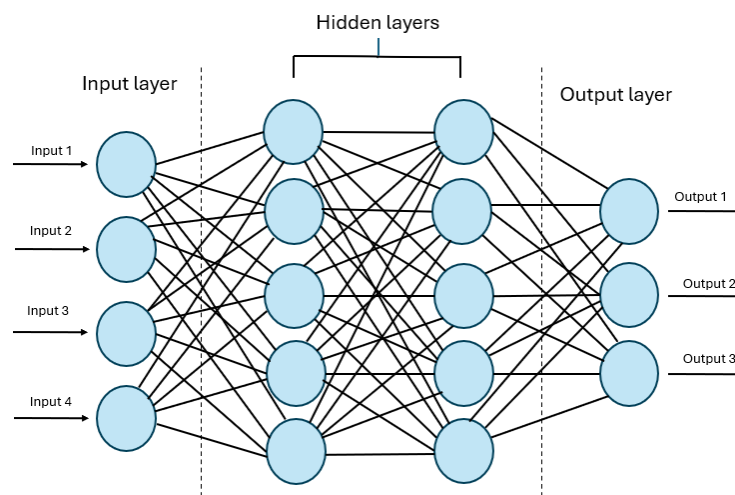


Figure 3.3. A simplified presentation of a feedforward neural network structure.

This image represents a simplified neural network with an input layer, hidden layers, and an output layer. The input layer receives data, which is processed through the hidden layers via weighted connections. Finally, the output layer produces the results or predictions based on the learned patterns. When examining neural networks from the perspective of a single neuron, the situation can be outlined as follows:

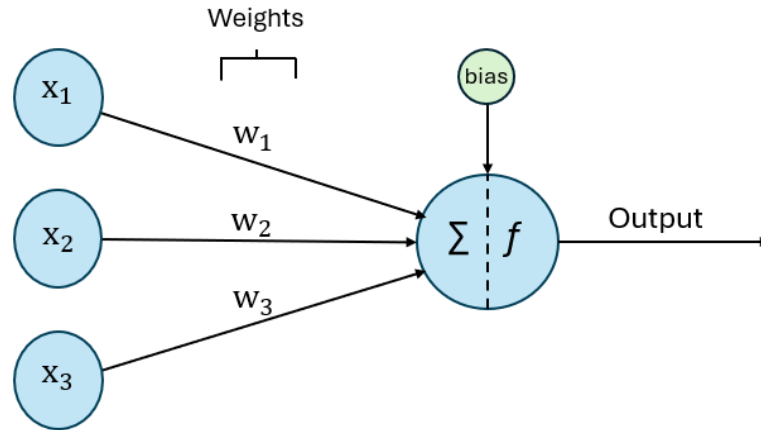


Figure 3.4. A single neuron structure.

A single neuron in a neural network receives multiple inputs (x_1, x_2, \dots, x_n) each associated with a weight (w_1, w_2, \dots, w_n) and adds a bias term to adjust the output [30]. The weighted sum of these inputs is passed through an activation function f , which determines the neuron's final output signal. Regarding the activation functions used in this study, rectified linear unit (ReLU)

$$f(x) = \max(0, x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0, \end{cases} \quad (3.1)$$

is applied to the hidden layers, while the softmax function

$$f(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}, \quad (3.2)$$

is used in the output layer. The softmax function converts raw output scores z_i into class probabilities by normalizing over all classes.

Within this research, categorical cross-entropy was used as the loss function during model training; it measures the difference between the predicted outputs and the true labels. The resulting error is then passed to the Adam optimizer, a variant of gradient descent, which updates each weight and bias via backpropagation. After training, the neural network is ready for classifying tasks.

3.4.2 Recurrent neural networks and LSTM

In this study, classification is performed using a variant of RNN, LSTM, which is a form of neural network well suited to modeling time-series data. In a feedforward network just described, each input is passed strictly forward from layer to layer without any memory of past inputs [30]. In contrast, RNN includes a feedback loop in which the hidden state carries information across time steps, allowing the network to leverage previous inputs for each prediction [31].

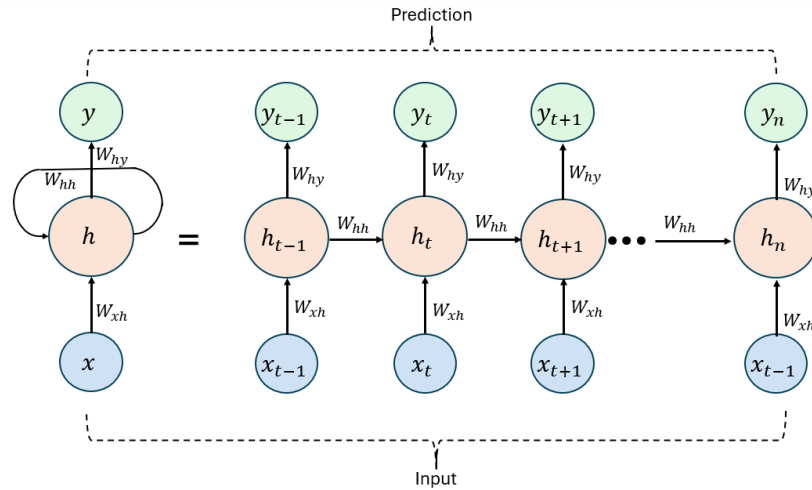


Figure 3.5. Basic RNN architecture, adapted from [31].

The loop can be seen on the left side of Figure 3.5, where the hidden state is fed back into itself via the weight matrix W_{hh} . On the right side, the network is “unrolled” in time, and the same loop appears as arrows carrying the previous state h_{t-1} directly into the next state h_t . In the figure, W_{xh} is the weight matrix between input and hidden layer and W_{hy} is the weight matrix between hidden and output layer. At each time index t , the RNN processes the input vector x_t and computes its new hidden state h_t as follows:

$$h_t = f_h(W_{xh}x_t + W_{hh}h_{t-1} + b_h), \quad (3.3)$$

where f_h is the activation function and b_h is the bias vector [31]. The output at each time step t is given by the following equation:

$$y_t = f_y(W_{hy}h_t + b_y), \quad (3.4)$$

where f_y is the activation function for the output layer and b_y is the bias vector [31].

In an LSTM layer, each hidden-state node of a standard RNN is replaced by a full LSTM cell, as illustrated in Figure 3.6.

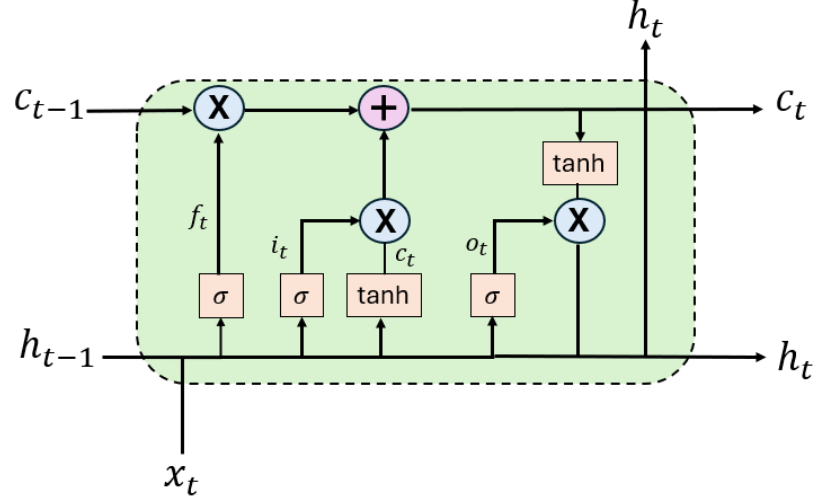


Figure 3.6. LSTM cell, redrawn based on [31].

The principal innovation lies in three gating mechanisms—input, forget, and output gates—which together manage the cell state c_t and hidden state h_t [31]. The input gate

$$i_t = \sigma(\mathbf{W}_{xi}x_t + \mathbf{W}_{hi}h_{t-1} + \mathbf{b}_i) \quad (3.5)$$

controls how much of the new input is considered, the forget gate

$$f_t = \sigma(\mathbf{W}_{xf}x_t + \mathbf{W}_{hf}h_{t-1} + \mathbf{b}_f), \quad (3.6)$$

decides how much of the previous state to forget, and the output gate

$$o_t = \sigma(\mathbf{W}_{xo}x_t + \mathbf{W}_{ho}h_{t-1} + \mathbf{b}_o), \quad (3.7)$$

regulates how much of the cell state c_t is used to compute the hidden state h_t [31]. In these update equations, each \mathbf{W} represents a weight matrix in at the position indicated by the subscript, σ is the sigmoid activation function, and each \mathbf{b} is the bias term for its respective gate. Additionally, LSTM needs the following update equations:

$$g_t = \tanh(\mathbf{W}_{xg}x_t + \mathbf{W}_{hg}h_{t-1} + \mathbf{b}_g), \quad (3.8)$$

compute the candidate cell input g_t and

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t, \quad (3.9)$$

for updating the cell state by first forgetting part of the old state ($f_t \odot c_{t-1}$) and then writing in the new information ($i_t \odot g_t$). Finally, LSTM needs

$$h_t = o_t \odot \tanh(c_t), \quad (3.10)$$

to produce the new hidden state by applying a hyperbolic tangent activation function, \tanh , to the updated cell state c_t , and then gating it element-wise with the output gate o_t [31].

Thanks to these gating mechanisms, LSTM cells can maintain and adapt their internal state across long sequences, excelling at capturing distant temporal patterns. When applied to IMS time-series data, this enables the network to retain crucial context over time and deliver markedly improved classification accuracy.

3.4.3 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a supervised dimensionality-reduction and classification technique that finds a linear combination of features to transform the data into a lower-dimensional space by maximizing the ratio of between-class to within-class variance, thereby ensuring maximum separability of two or more classes [32]. The implementation of LDA can be divided into three stages, as outlined by Tharwat et al. [32]. The first step is to compute the between-class variance (also known as the between-class scatter matrix) for each class. That means calculating the distance between the means of each class. The second step is calculating the within-class variance (within-class scatter matrix), which means calculating the distance between mean and the samples of each class. The third step involves constructing a lower dimensional space which minimizes the within-class variance and maximizes the between class variance. This setting corresponds to the class-independent variant of LDA, in which a single shared projection space is used for all classes. However, LDA relies on specific assumptions, particularly that all classes share a common covariance matrix and that the data within each class follows approximately a Gaussian distribution [33, p.108].

The between-class variance is given by the formula,

$$\mathbf{S}_B = \sum_{i=1}^c n_i (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T, \quad (3.11)$$

where c denotes the number of classes, n_i indicates the number of samples in class i , $\boldsymbol{\mu}_i$ is the mean vector of class i and $\boldsymbol{\mu}$ represents the overall mean vector, calculated from all samples across all classes [32]. The within-class variance is obtained by summing the individual within-class scatter matrices of all classes. For class i , the within-class variance matrix \mathbf{S}_{Wi} is given by:

$$\mathbf{S}_{Wi} = \sum_{x \in \omega_i} (x_i - \boldsymbol{\mu})(x_i - \boldsymbol{\mu})^T, \quad (3.12)$$

and the total within-class variance is then:

$$\mathbf{S}_W = \sum_{i=1}^c \mathbf{S}_{Wi}, \quad (3.13)$$

where ω_i is the set of all samples in class i and \mathbf{x} is an individual sample [32]. The final step is to determine a projection that maximizes the separability between classes by optimizing the ratio of the between-class scatter to the within-class scatter, as defined in Fisher's criterion:

$$\arg \max_{\mathbf{W}} \frac{\mathbf{W}^T \mathbf{S}_B \mathbf{W}}{\mathbf{W}^T \mathbf{S}_W \mathbf{W}}. \quad (3.14)$$

By solving this optimization problem, the axes of the new space are selected such that they best separate the classes while minimizing overlap. The resulting matrix \mathbf{W} contains the directions along which the classes are most distinguishable [32]. To simplify, LDA aims to find a new coordinate space where the class means are as far apart as possible, while the samples within each class are clustered as closely together as possible, as illustrated in Figure 3.7.

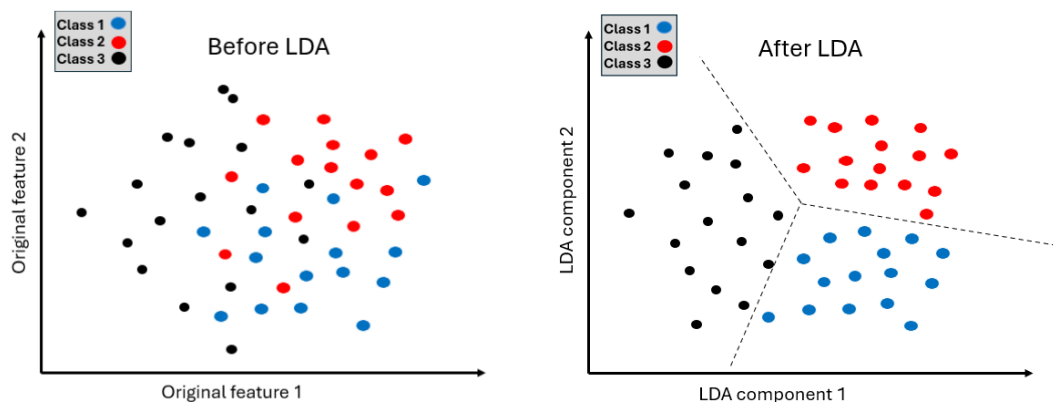


Figure 3.7. Illustrative example of how LDA works with three classes.

In the right-hand plot, the classes are clearly more separable compared to the original feature space, making classification easier and more accurate. However, although both plots are shown in two dimensions, the right-hand plot represents a transformed coordinate space of lower dimensionality constructed by LDA.

The actual classification of each sample is performed based on its proximity to the class centers in the lower-dimensional space. This can, for example, be based either on the Euclidean distance to the projected class means or on a discriminant score, where the decision is made according to the highest estimated class probability [33, p.110].

3.4.4 Five-fold cross-validation

In this study, five-fold cross-validation was used to evaluate the classification performance of the models across different measurement sessions. In each iteration, the samples from one session were used as test set while the samples from the remaining four sessions were used for training. This process was repeated five times so that each session was used once for testing. This validation method was applied consistently throughout all experiments. To implement this, group k-fold cross-validation was used, where the data was partitioned so that each measurement session formed a distinct group and appeared in only one fold. This ensured that session-specific information did not leak between training and testing sets, reducing the risk of overfitting to group-related patterns [34].

Using cross-validation in this context is particularly beneficial, as it helps to average out variability caused by differing measurement conditions across sessions, such as changes in air pressure, humidity, and temperature. It also reduces the influence of session-specific systematic errors that might otherwise distort the model's generalization ability. This makes the evaluation more robust and representative of real-world performance. This group k-fold method was used in all experiments presented in the following chapter.

4 RESULTS AND DISCUSSION

In this section, all the algorithms used, and their corresponding results are discussed. First, the use of LSTM is explored, as its structure differs from the other models used. After that, the performance of the other algorithms is analyzed. The compiled results are presented in Table 4.1. Finally, factors influencing the results are discussed.

All the classification algorithms were implemented using Python programming language. Various open-source machine learning libraries were used for assistance, with the most important tools coming from scikit learn, pandas, and NumPy.

4.1 LSTM approach

As a first approach, classification was performed using an LSTM network. The data for the LSTM neural network classifier was detrended and normalized as described in Section 3.3, and then utilized in three ways: First, the full 600×14 samples were used as input, resulting in a total of 45 samples—36 for training and 9 for testing. This led to poor results, with accuracy reaching only around 30%.

In the second test, the sample size was reduced by dividing each 10-minute sample into smaller segments, thereby increasing the total number of samples. When the sample duration was set to one minute, the number of samples increased tenfold, resulting in a total of 450 samples, of which 360 were used for training and 90 for testing. Using one-minute samples, the classification accuracy reached 48% at best. As classification accuracy improved when sample size was reduced and the number of samples increased, shorter samples of 30 seconds in duration were also evaluated. This resulted in a total of 900 training samples and 180 testing samples. This led to a few percentage points of improvement in accuracy, with the best accuracies reaching approximately 53%, depending on the run. However, further reducing the sample size and increasing the number of samples did not lead to improved results.

In the third test, ten subtests were used. In each subtest, the test data consisted of a 60-second segment from one session, while the corresponding time segments from the other four sessions were used as training data. This resulted in 9 test samples and 36 training samples (4×9) per subtest. Each subtest covered a different time interval within the 0–600 second range (e.g., 1–60 s, 61–120 s, ..., 541–600 s). This approach led to similar results as in the first test, with accuracy remaining around 30%.

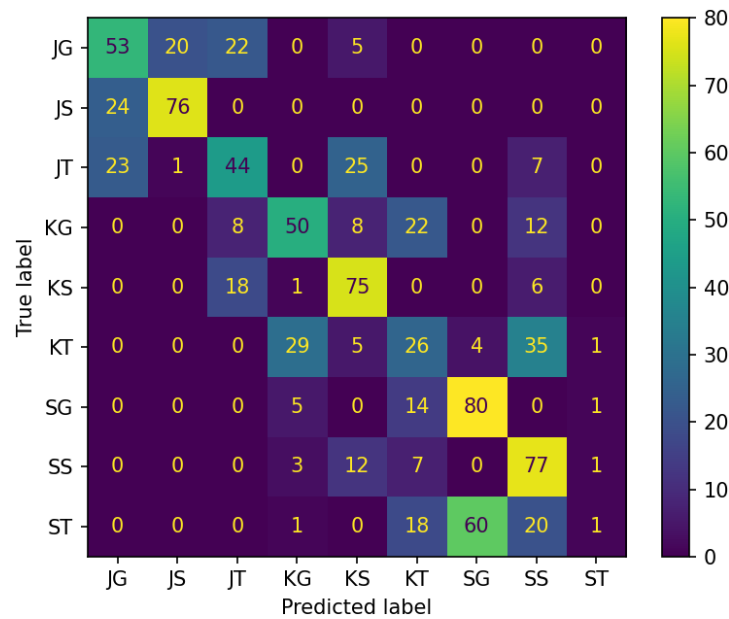


Figure 4.1 Confusion matrix across all five folds for LSTM with 30 second samples as an input.

Figure 4.1 shows the confusion matrix for the test samples from all five folds when using 30-second samples. Thus, the matrix includes a total of $180 \times 5 = 900$ test samples, 300 samples per glass type. The rows represent the true labels (actual whisky-glass combinations), and the columns represent the predicted labels given by the classifier (see the combination abbreviations listed above). The matrix shows that with SAVU glass the classification accuracy was highest. For the Savu glass, identified by the letter 'S' as the second character in each combination, 228 out of 300 samples were correctly classified, corresponding to an accuracy of 76%. The next best-performing glass was the Glencairn ($181/300 = 61\%$), while the poorest performance was observed with the Tumbler glass, with only 71 out of 300 samples correctly classified (24%). However, no strong conclusions should be drawn from these results alone, as other algorithms yield markedly different outcomes.

Although there is no universal threshold for the required training data size, LSTM models generally benefit from larger datasets due to their high model complexity, as demonstrated for example in [35]. In the present work, the limited size of the dataset likely contributed to the relatively poor performance of the LSTM model.

4.2 Other approaches

In addition to the LSTM, other methods were also tested to validate the results and to improve the classification accuracy. Among these slightly simpler approaches, SVC and LDA performed the best. In all experiments, various preprocessing steps were tested,

but most often the best results were achieved by applying smoothing alongside detrending and normalization (Section 3.3).

Several input formats were evaluated, including full 10-minute samples from all channels (600×14), single-channel data (600), and feature vectors derived from all channels, as described in Section 3.3. PCA-transformed data explaining 95% of the total variance was also tested, resulting in a single feature vector of length 1200. Feature vectors proved to be the most effective input format, as they resulted in the highest classification accuracies. Different combinations of features—such as standard deviation, mean, and others—were tested to find the optimal configuration for classification. What is remarkable is that not every time the same combination of features yielded the best results; for different classifiers, different feature combinations worked best.

To further improve accuracy, also the hyperparameters of the algorithms were tuned. For example, in the case of SVC, the best results were achieved using the linear kernel and setting the regularization parameter C to 30. In the whisky classification, several methods outperformed LSTM (79% accuracy): SVC achieved 86.7%, LDA and KNN both reached 82.2%, while QDA performed significantly worse at 46.7%. Further examination of the classification results is provided in the next section.

4.3 Combined results

This section presents a combined analysis of the most significant results obtained across all experiments. The results are presented in Table 4.1, which is sorted by overall accuracy, from highest to lowest. The first column of the table shows the classifier used along with its corresponding hyperparameters. The next columns list the preprocessing methods applied. In the 'Features extracted' column, it is indicated whether features were extracted from the original samples and which features were used as input for the classifier. The 'Input, size' column describes the type of input given to the algorithm and its size. The following two columns show the classification accuracy for identifying the whisky contained in the sample, as well as the accuracy for identifying the glass used. Next, it is shown which whiskies were identified with the highest accuracy. The 'Most accurate glass' column indicates from which glass the samples were the easiest to classify. Finally, the overall accuracy is presented, which refers to the accuracy of the model in correctly identifying both the whisky and the glass.

Classifier	Detrending + Normalizing	Smoothing	Features extracted	Input, size	Whiskey accuracy (%)	Glass accuracy (%)	The most accurate whiskey (%)	The most accurate glass (%)	Overall accuracy (%)
SVC (linear, C=30)	Yes	Yes	mean, sd, max	Feature vector, 42	86.7	68.9	Jim Bean Rye, 100	Tumbler, 80.0	66.7
LDA	Yes	No	sd, mean, max, min median, kurtosis	Feature vector, 84	82.2	62.2	Kyrö Wood Smoke Malt Rye, 93.3	Glencairn, 80.0	57.8
KNN (n=3)	Yes	Yes	mean, var	Feature vector, 28	82.2	57.8	Jim Bean Rye, 93.3	Glencairn, 73.3	57.8
LDA (smoothing)	Yes	Yes	sd, mean, max, min median, kurtosis	Feature vector, 84	77.7	60.0	Jim Bean Rye, 86.7	Glencairn, 73.3	53.3
LSTM	Yes	No	-	30 s segment, 30 x 14 matrix	79.1	57.2	Jim Bean Rye, 89.7	Savu, 70.7	53.0
KNN (n=1)	Yes	Yes	mean, var	Feature vector, 28	86.7	53.3	Kyrö Wood Smoke Malt Rye 93.3	Savu, 73.3	51.1
SVC (linear, C=30)	Yes	Yes	-	60 s segment 60x14 matrix	76.7	56.7	Jim Bean Rye, 87.3	Savu, 76.0	50.7
KNN (n=1) (95% PCA)	Yes	No	-	Vector, 1200	75.6	57.8	Jim Bean Rye, 86.7	Savu 80.0	48.9
KNN (n=1)	Yes	Yes	-	60 s segment 60x14 matrix	77.3	51.3	Jim Bean Rye, 84.0	Savu, 69.3	47.7
LSTM	Yes	No	-	60 s segment 60 x 14 matrix	77.6	50.7	Jim Bean Rye, 85.3	Savu, 62.2	46.4
SVC (poly, C=30)	Yes	Yes	mean, sd, var	Feature vector, 42	71.1	53.3	Jim Bean Rye, 93.3	Glencairn, 60.0	44.4
LDA	Yes	No	-	60 s segment, 60x14 matrix	58.4	59.1	Jim Bean Rye, 83.3	Savu, 71.3	41.8
SVM (95% PCA)	Yes	No	-	Vector, 1200	53.3	53.3	Jim Bean Rye, 86.7	Glencairn 73.3	40.0
QDA	Yes	Yes	mean, sd, max	Feature vector, 42	46.7	40.0	Jim Bean Rye, 53.3	Tumbler, 46.7	26.7
SVC (sigmoid, C=30)	Yes	Yes	mean, sd, var	Feature vector, 42	44.4	42.2	Jim Bean Rye, 66.7	Tumbler, 53.3	22.2
LSTM	Yes	No	-	10 min sample, 600x14 matrix	71.1	28.9	Kyrö Wood Smoke Malt Rye, 86.7	Glencairn, 46.7	22.2

Table 4.1 Final classification accuracies, including information about preprocessing methods, extracted features, and input data used.

The best overall classification accuracy (66.7 %) was achieved using a linear SVC (C = 30) with detrended, normalized, and smoothed data. The input was a feature vector consisting of mean, standard deviation, and maximum values extracted from all channels. Notably, all the top four results were obtained using feature vectors, highlighting their clear advantage over raw time-series data and PCA-based representations. In general, smoothing improved classification performance. Tests based on only a single or a few channels consistently resulted in lower accuracy and were therefore excluded from the table. Additionally, it can be noted that PCA did not improve performance, and QDA yielded the poorest overall results, achieving only 26.7% accuracy in distinguishing whisky–glass combinations.

A key finding was that the classifiers were consistently more successful at identifying the whisky than the glass. At best, the whiskies were identified with 86.7% accuracy, whereas the highest accuracy achieved for glass identification was 68.9%. Tumbler’s accuracies varied from 13% to 80%, Glencairn’s from 47% to 80%, and SAVU’s from

27% to 80% across all tests (these ranges are not included in Table 4.1). As shown in the table, the best-performing glass varied depending on the classifier, and all glasses reached similar maximum accuracies. These findings suggest that it is not possible to establish a clear ranking among the glasses or to conclude that any particular glass consistently enables more accurate classification with the e-nose. Thus, within the context of this experimental setup, the results do not support the idea that the shape of the glass has a significant effect on classification accuracy.

4.4 Analysis and discussion

The original goal, that was presented in the introduction section, was to find out if IMS device combined with machine learning techniques can classify whiskies and the glass they are served in. The result is that with three different whiskies and three different glasses, the correct whisky-glass combination was identified with a maximum accuracy of 66.7%. While this clearly shows that the classifier performs better than random selection among nine possible combinations (i.e., 1/9 accuracy), the result is still not sufficient for the model to be considered a reliable classifier for these combinations.

The results may be attributed to multiple factors, including potential inaccuracies in the measurement process. For example, the distance between the sensor and the sample may not have been completely consistent in all measurements, or an uneven amount of whisky may have accumulated in the pits of the Savu glass. Naturally, care was taken to ensure consistent measurements, but small variations are always possible in manual measurements.

It is also possible that the measurement setup used in this study is simply not sufficient for achieving highly accurate results. Similar challenges have been discussed in previous research. For example, in research by Kondratev et al. [36], it was stated that: "The measurements always contain noise. The magnitude of noise depends on where the scent source was sampled. If a scent is measured from a controlled headspace, e.g., sealed flask, the readings are in general rather stable [...] Scents measured in a less controlled setting, e.g., from a plate, in general yield more instable channel responses or cause no valuable responses at all." This plate situation corresponds well to the open-air measurement setup used in this study, where the whisky sample was measured directly at the edge of the glass without enclosure, allowing surrounding air to dilute the signal.

This explanation regarding unstable responses is further supported by the observation that, in this study, samples from the same whisky-glass combination occasionally

appeared visibly different. This will be examined in more detail in Section 4.4.1. In addition, changing environmental conditions—such as temperature, humidity, and air pressure—may have contributed to the variation in the results, as discussed in Section 4.4.2. It is also possible that alternative data preprocessing methods or more suitable machine learning algorithms could lead to improved classification accuracy.

4.4.1 Impact of Sample Variation on Performance

As noted in Section 4.4, data from the same combination occasionally appeared noticeably different. Therefore, it is reasonable to assume that very high classification accuracies cannot be achieved in this case. Figure 4.2 illustrates two measurements of the same whisky-glass combination, taken from different sessions, highlighting the variation between them. The data includes one minute of baseline followed by the entire 10-minute sample on channel 3.

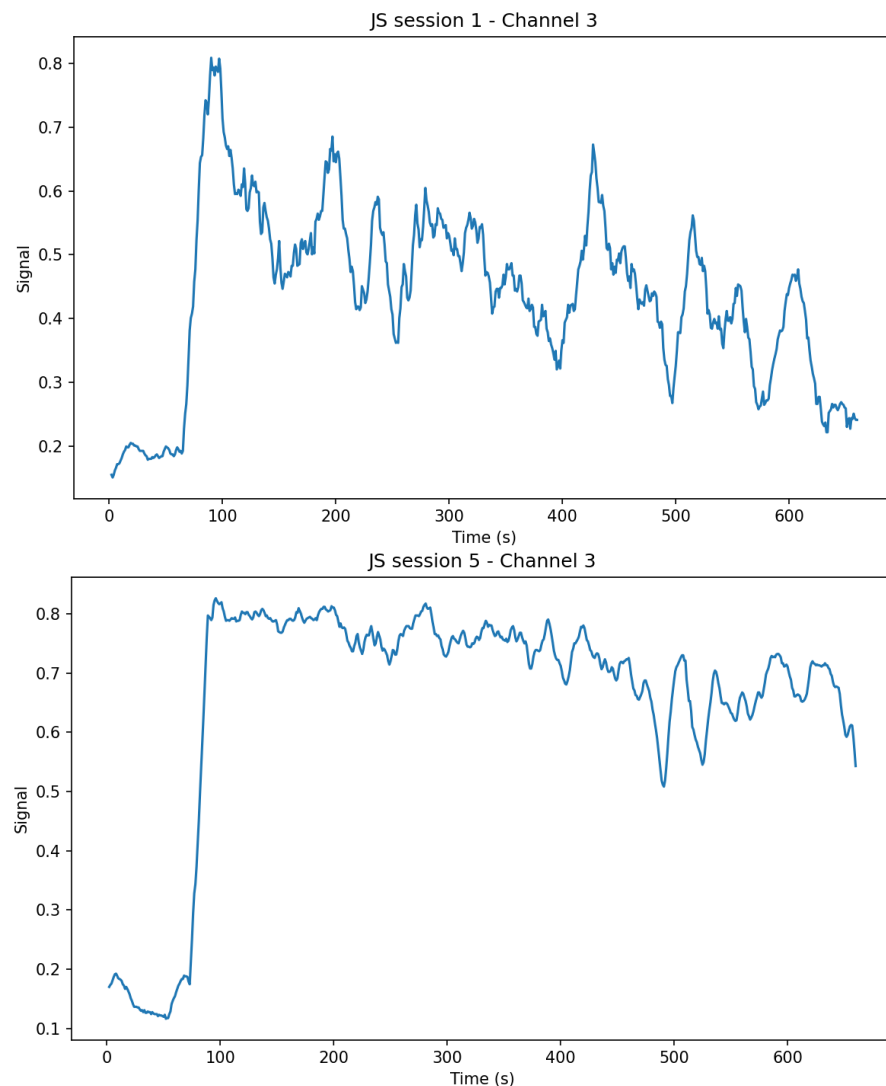


Figure 4.2. Normalized IMS signal (channel 3) of Jim Beam Rye in Savu glass during sessions 1 and 5.

The transient phase looks almost the same in both cases but after that, the signal shows a clear difference. Measurement from session 5 is much more stable whereas measurement from session 1 differs significantly.

However, in most cases, the measurement results from the same sample appeared much more like each other, as illustrated in Figure 4.3.

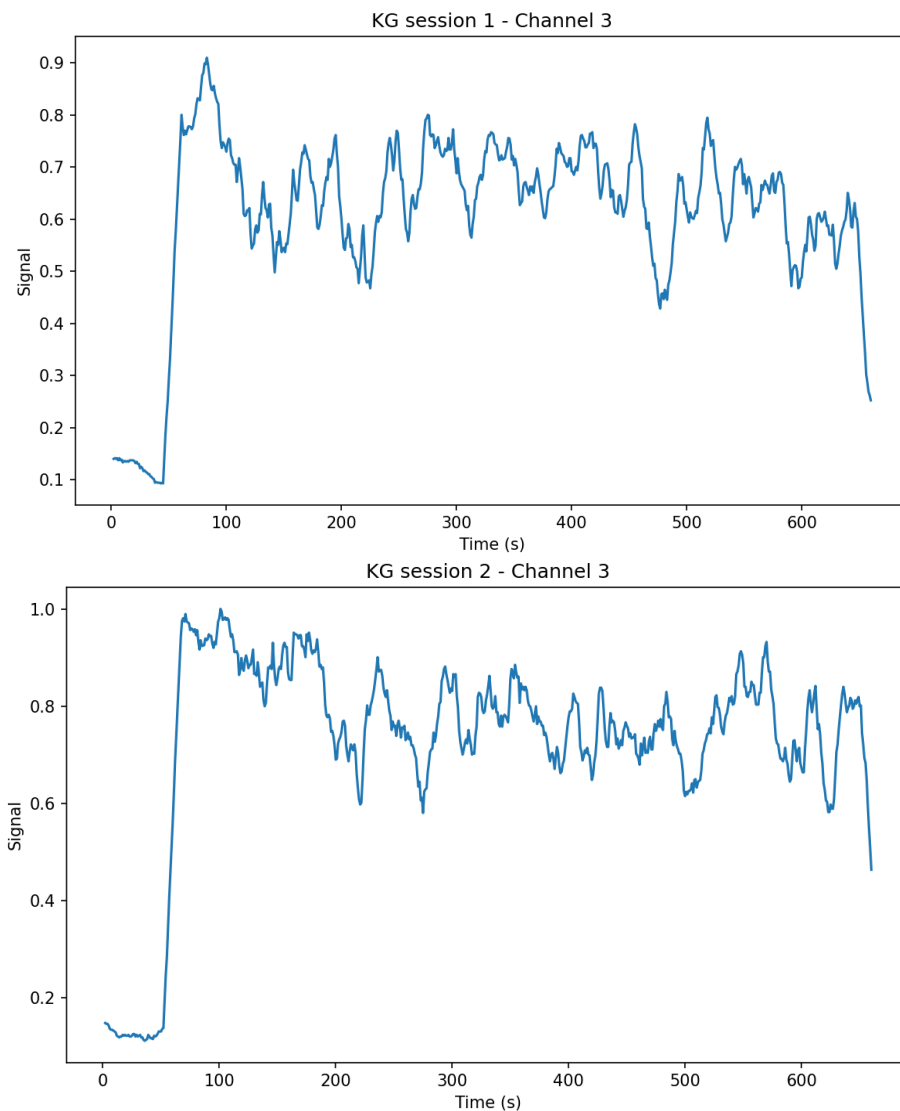


Figure 4.3. Normalized IMS signal (channel 3) of Kyrö Malt Rye in Glencairn glass during sessions 1 and 2.

If all samples taken from the same combination had shown this level of similarity, the classification accuracy would likely have been considerably better.

The IMS response depends strongly on the chemical composition of the substance being measured [4, p. 7]. Chemically, smoky whiskies differ from non-smoky whiskies mainly in their higher concentration of phenolic compounds such as phenol, guaiacol,

and cresols [23]. Since these phenolic compounds such as guaiacol and cresols exhibit relatively low volatility [23], it is reasonable to assume that smoky whiskies would produce a distinct IMS response profile. This, in turn, could facilitate more accurate classification of smoky whiskies compared to non-smoky ones.

However, this was not the case in this study — at least visually, it was not possible to distinguish smoky and non-smoky whiskies from each other based on the IMS signals. This observation is supported by the fact that the smoky whisky, Kyrö Wood Smoke Malt Rye, was not the most accurately classified whisky (see Table 4.1). Instead, the best classification accuracy was achieved with Jim Beam Rye. On the other hand, it should be noted that Kyrö Malt Rye and Kyrö Wood Smoke Malt Rye are produced by the same manufacturer, which may contribute to their similarity in IMS measurements.

4.4.2 The Role of Measurement Conditions in Result Accuracy

The standing point for the experiment was that the conditions would always remain constant. For that reason, the measurements were conducted in the same special scent analyzing laboratory to ensure environmental stability and minimize external interference. It is crucial to maintain constant measurement conditions because IMS devices are sensitive to environmental factors such as temperature, humidity, and air pressure. For example, variations in humidity can lead to significant change in IMS signal intensity for many compounds [21], which may in turn influence the classification accuracy in this study.

Stable conditions ensure that the measurement results are comparable across different samples. The room was equipped with exceptionally good ventilation, and no other activities were conducted in the room during the measurements. However, the conditions changed significantly. The measurements had two major problems: the conditions didn't remain the same during a single measurement session and they also changed between all five sessions.

Figures 4.4 – 4.7 display all five measurement sessions in sequence, separated by red dashed lines. Each session includes nine individual 10-minute measurements corresponding to different sample combinations.

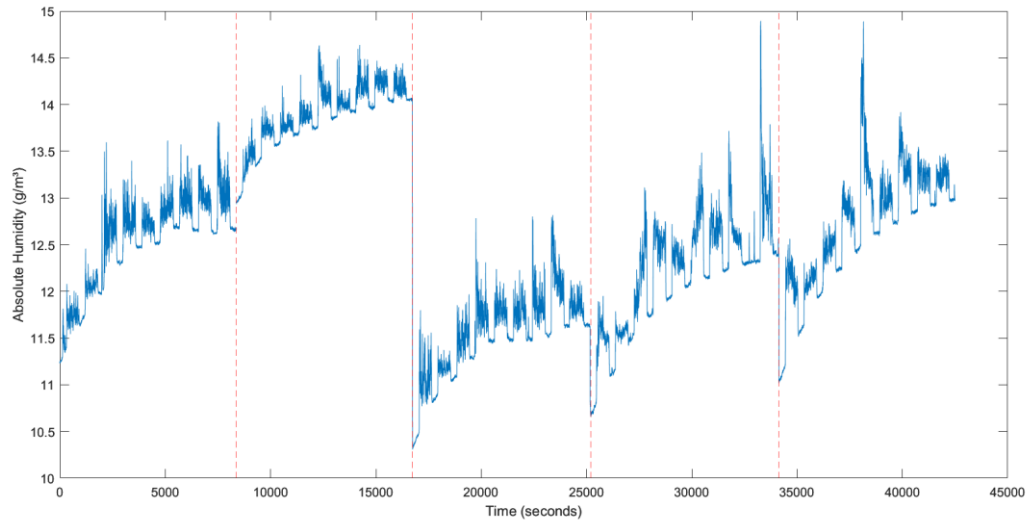


Figure 4.4. Absolute humidity across all five measurement sessions.

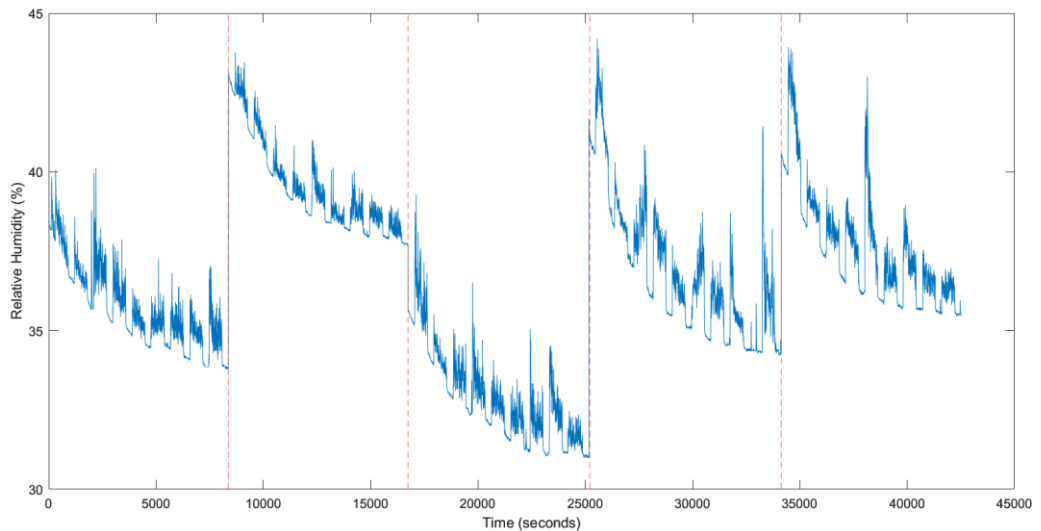


Figure 4.5. Relative humidity across all five measurement sessions.

Both absolute and relative humidity show logarithmic behavior. Only at the end of each measurement session, after approximately 8000 seconds, does the humidity begin to stabilize.

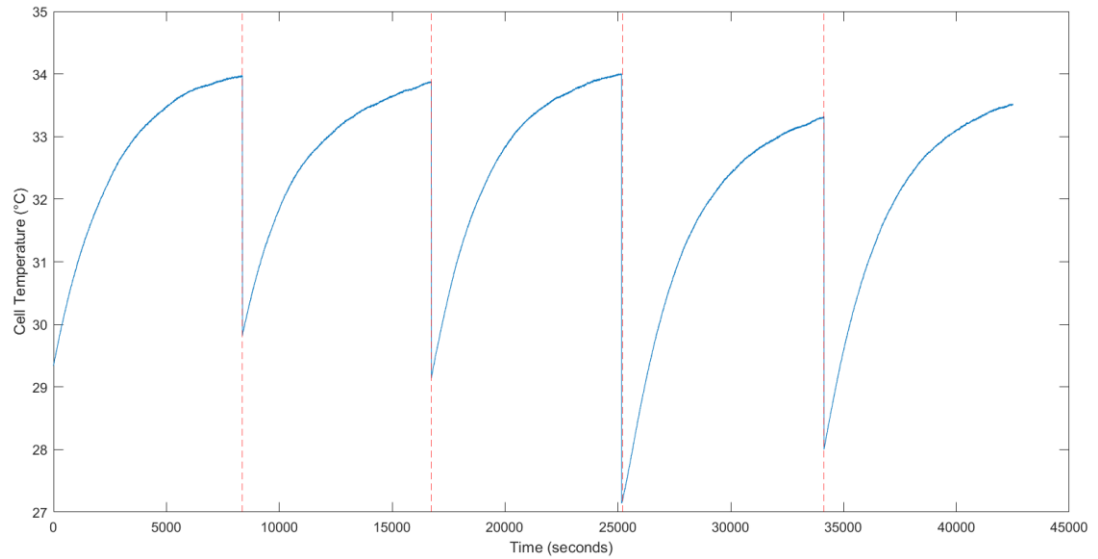


Figure 4.6. The device's internal temperature (cell temperature) across all five measurement sessions.

The temperature within each measurement session also follows a logarithmic trend.

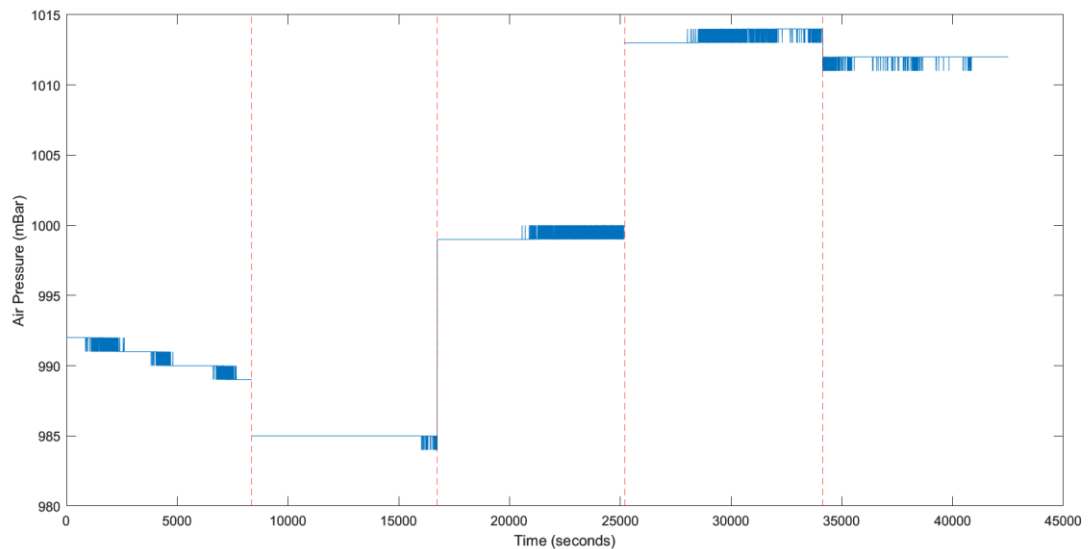


Figure 4.7. Air pressure across all five measurement sessions

Small differences in air pressure were also observed between the measurement sessions.

The data indicated that stable conditions are reached only near the end of the session, after approximately 2 hours and 20 minutes. This indicates that, to ensure stable measurement conditions, the device should be kept running for at least approximately 2.5 hours before starting actual measurements.

The impact of varying environmental conditions on the final classification results is challenging to assess. However, given that the achieved accuracy is not perfect, it can be concluded that the method used in this study should be applied under controlled indoor conditions, at least until further improvements in accuracy are achieved. The field applicability of the method is difficult to estimate but it is expected that the classification accuracy would decrease if measurements were taken in varying environments.

5 CONCLUSION AND OUTLOOK

This study explored the feasibility of IMS combined with machine learning algorithms to classify whisky-glass combinations. Based on the results, it can be concluded that classification using IMS and machine learning is possible, although the achieved accuracies reveal certain limitations.

A total of nine different whisky-glass combinations were tested. Feature extraction and preprocessing were applied to the raw data to improve the model performance, and among the tested algorithms, Support Vector Classification (SVC) provided the best results. The best classification accuracies achieved were 87% for whisky identification, 69% for glass identification, and 67% for whisky-glass combination classification. The results show that across all tested classifiers, the classification of whiskies was consistently more accurate than the classification of glasses. To potentially improve performance further, future studies could explore combining multiple classifiers and a final decision could be made using a majority voting scheme, as described in more detail in [20]. This approach could better capture the complementary strengths of individual algorithms and enhance overall accuracy.

It is important to recognize that this study was a small-scale investigation. Therefore, the validity and generalizability of the findings should be interpreted with caution. The measurement approach simulated a practical sniffing situation, but future studies could explore classification using VOCs collected from closed containers. Such an approach could minimize the effects of external environmental factors and improve the consistency of the measurements. The method could have practical applications if the classification accuracy could first be improved, and the testing process accelerated. High classification accuracy could enable commercial operations where whiskies are reliably and rapidly identified and distinguished from counterfeit products, even under field conditions.

Additionally, it was observed during the experiments that the ChemPro 100i measurement device should be kept running for approximately 2.5 hours prior to data collection in order to stabilize internal temperature and humidity. This stabilization is necessary to ensure more consistent and reliable measurements.

All in all, this study demonstrated that IMS combined with machine learning has promising potential for whisky analysis under controlled conditions. Nevertheless, further research and development are required before the method can be reliably applied on a broader scale.

REFERENCES

- [1] Whisky Market Report 2025, Research and Markets, 2023. Available: <https://www.researchandmarkets.com/reports/5806931/whisky-market-repo>
- [2] C. A. Okolo, K. N. Kilcawley, C. O'Connor, *Recent advances in whiskey analysis for authentication, discrimination, and quality control*, *Comprehensive Reviews in Food Science and Food Safety*, vol. 22, no. 1, pp. 123–145, 2023. Available: <https://doi.org/10.1111/1541-4337.13249>
- [3] Z. Zhai, et al., *Electronic Noses: From Gas-Sensitive Components and Practical Applications to Data Processing*, *Sensors*, vol. 24, no. 15, p. 4806, 2024. Available: <https://doi.org/10.3390/s24154806>
- [4] G. A. Eiceman and Z. Karpas, *Ion Mobility Spectrometry*, 2nd ed., CRC Press, Boca Raton, FL, 2005. Preview version used; page numbers refer to the preview PDF version. Available: <https://www.taylorfrancis.com/books/mono/10.1201/9781420038972/ion-mobility-spectrometry-eiceman-karpas>
- [5] M. A. Mäkinen, O. A. Anttalainen, M. E. T. Sillanpää, *Ion mobility spectrometry and its applications in detection of chemical warfare agents*, *Analytical Chemistry*, vol. 82, no. 23, pp. 9594–9600, 2010. Available: <https://doi.org/10.1021/ac100931n>
- [6] R. Fernandez-Maestre, *Ion mobility spectrometry: history, characteristics and applications*, *Revista UDCA de Actualidad y Divulgación Científica*, vol. 15, pp. 467–479, 2012. Available: <https://doi.org/10.31910/rudca.v15.n2.2012.848>
- [7] E. te Brinke, A. Arrizabalaga-Larrañaga, M. H. Blokland, *Insights of Ion Mobility Spectrometry and Its Application on Food Safety and Authenticity: A Review*, *Analytica Chimica Acta*, vol. 1222, 2022, pp. 340039–340039. Available: <https://doi.org/10.1016/j.aca.2022.340039>
- [8] G. Minaev, et al., *A Comparison of Various Algorithms for Classification of Food Scents Measured with an Ion Mobility Spectrometry*, *Sensors*, vol. 22, no. 5, p. 4420, 2022. Available: <https://doi.org/10.3390/s21020361>
- [9] D. Karakaya, A. Ulucan, M. Turkan, *Electronic Nose and Its Applications: A Survey*, *International Journal of Automation and Computing*, vol. 17, no. 2, pp. 179–209, Apr. 2020. Available: <https://doi.org/10.1007/s11633-019-1212-9>

- [10] F. Röck, N. Barsan, U. Weimar, *Electronic nose: Current status and future trends*, *Chemical Reviews*, vol. 108, no. 2, pp. 705–725, 2008. Available: <https://doi.org/10.1021/cr068121q>
- [11] T. T. Dung, et al., *Applications and Advances in Bioelectronic Noses for Odour Sensing*, *Sensors*, vol. 18, no. 1, p. 103, 2018. Available: <https://doi.org/10.3390/s18010103>
- [12] A. Roine, et al., *Rapid and Accurate Detection of Urinary Pathogens by Mobile IMS-Based Electronic Nose: A Proof-of-Principle Study*, *PLOS ONE*, vol. 9, no. 12, p. e114279, 2014. Available: <https://doi.org/10.1371/journal.pone.0114279>
- [13] M. Hernández-Mesa, et al., *Ion Mobility Spectrometry in Food Analysis: Principles, Current Applications and Future Trends*, *Molecules*, vol. 24, no. 15, p. 2706, 2019. Available: <https://doi.org/10.3390/molecules24152706>
- [14] C. A. Browne, T. P. Forbes, E. Sisco, *Detection and Identification of Sugar Alcohol Sweeteners by Ion Mobility Spectrometry*, *Analytical Methods*, vol. 8, no. 28, pp. 5611–5618, 2016. Available: <https://doi.org/10.1039/c6ay01554a>
- [15] M. Alikord, et al., *Food Safety and Quality Assessment: Comprehensive Review and Recent Trends in the Applications of Ion Mobility Spectrometry (IMS)*, *Critical Reviews in Food Science and Nutrition*, vol. 62, no. 18, pp. 4833–4866, 2022. Available: <https://doi.org/10.1080/10408398.2021.1879003>
- [16] P. Wiśniewska, et al., *Chemical composition analysis and authentication of whisky*, *Journal of the Science of Food and Agriculture*, vol. 95, no. 11, pp. 2159–2166, 2014. Available: <https://doi.org/10.1002/jsfa.6960>
- [17] J. Macias et al., *Whisky tasting using a bimetallic nanoplasmonic tongue*, *Nanoscale*, vol. 11, no. 36, pp. 16706–16715, 2019. Available: <https://doi.org/10.1039/C9NR04583J>
- [18] T. Liu et al., *A Multiscale Wavelet Kernel Regularization-Based Feature Extraction Method for Electronic Nose*, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 11, pp. 7078–7089, Nov. 2022. Available: <https://doi.org/10.1109/TSMC.2022.3151761>
- [19] A. R. Martins, et al., *Discrimination of whisky brands and counterfeit identification by UV–Vis spectroscopy and multivariate data analysis*, *Food Chemistry*, vol. 229, pp. 142–151, 2017. Available: <https://doi.org/10.1016/j.foodchem.2017.02.024>
- [20] W. Zhang, et al., *The Use of Electronic Nose for the Classification of Blended and Single Malt Scotch Whisky*, *IEEE Sensors Journal*, vol. 22, no. 7, pp. 7015–7021, Apr. 2022. Available: <https://doi.org/10.1109/JSEN.2022.3147185>

- [21] H. Borsdorf, P. Fielder, T. Mayer, *The effect of humidity on gas sensing with ion mobility spectrometry*, *Sensors and Actuators B: Chemical*, vol. 218, pp. 184–190, 2015. Available: <https://doi.org/10.1016/j.snb.2015.04.102>
- [22] P. Müller et al., *Online Scent Classification by Ion-Mobility Spectrometry Sequences*, *Frontiers in Applied Mathematics and Statistics*, vol. 5, p. 39, 2019. Available: <https://doi.org/10.3389/fams.2019.00039>
- [23] T. J. Kelly, C. O'Connor, K. N. Kilcawley, *Sources of Volatile Aromatic Congeners in Whiskey*, *Beverages*, vol. 9, no. 3, p. 64, 2023. Available: <https://doi.org/10.3390/beverages9030064>
- [24] SciPy, *scipy.signal.detrend* — *SciPy v1.13.0 Manual*, 2024. Available: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.detrend.html>
- [25] T. D. Pham, *Time–Frequency Time–Space LSTM for Robust Classification of Physiological Signals*, *Scientific Reports*, vol. 11, no. 1, p. 6936, 2021. Available: <https://doi.org/10.1038/s41598-021-86432-7>
- [26] D. Huang, et al., *Comparison of linear discriminant analysis methods for the classification of cancer based on gene expression data*, *Journal of Experimental & Clinical Cancer Research*, vol. 28, no. 1, p. 149, 2009. Available: <https://doi.org/10.1186/1756-9966-28-149>
- [27] T. Cover, P. Hart, *Nearest neighbor pattern classification*, *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967. Available: <https://doi.org/10.1109/TIT.1967.1053964>
- [28] C. Cortes, V. Vapnik, *Support-vector networks*, *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995. Available: <https://doi.org/10.1007/BF00994018>
- [29] Y. Qin, *A Review of Quadratic Discriminant Analysis for High-Dimensional Data*, *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 13, no. 2, e1434, 2021. Available: <https://doi.org/10.1002/wics.1434>
- [30] M. Islam, G. Chen, S. Jin, *An Overview of Neural Network*, *American Journal of Neural Networks and Applications*, vol. 5, no. 1, pp. 7–11, 2019. Available: <https://doi.org/10.11648/j.ajnna.20190501.12>
- [31] I. D. Mienye, T. G. Swart, G. Obaido, *Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications*, *Information*, vol. 15, no. 9, p. 517, 2024. Available: <https://doi.org/10.3390/info15090517>

- [32] A. Tharwat, et al., *Linear discriminant analysis: A detailed tutorial*, *AI Communications*, vol. 30, no. 2, pp. 169–190, 2017. Available: <https://doi.org/10.3233/AIC-170729>
- [33] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., Springer, 2009. Available: <https://doi.org/10.1007/978-0-387-84858-7>
- [34] J. Allgaier, R. Pryss, *Cross-Validation Visualized: A Narrative Guide to Advanced Methods*, *Machine Learning and Knowledge Extraction*, vol. 6, no. 2, pp. 1378–1388, 2024. Available: <https://doi.org/10.3390/make6020065>
- [35] T. Boulmaiz, M. Guermoui, B. Hamouda, *Impact of training data size on the LSTM performances for rainfall–runoff modeling*, *Modeling Earth Systems and Environment*, vol. 6, pp. 1–11, 2020. Available: <https://doi.org/10.1007/s40808-020-00830-w>
- [36] A. Kondratev et al., *A comparison of online methods for change point detection in ion-mobility spectrometry data*, *Array*, vol. 14, 2022. Available: <https://doi.org/10.1016/j.array.2022.100140>