LASER RAMAN SPECTROSCOPIC ASSESSMENT OF HONEY ADULTERATION BY MOLASSES

BY

ROBERT ISAARA OPATI IKEDI
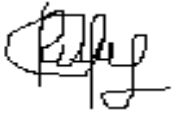
I56/11152/2018

B.Sc. (Hons)

A thesis submitted for examination in partial fulfillment of the requirements of the award of the degree of Master of Science in Physics of the University of Nairobi.

© August, 2022

DECLARATION

I declare that this Thesis is my original work and has not been submitted elsewhere for examination. Where other people's work or my own work has been used, this has properly been acknowledged and referenced in accordance with the University of Nairobi's requirements.

Signature ……… ………… ……   Date: **24/08/2022**

<div align="center">

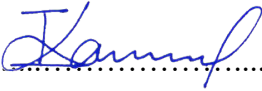**ROBERT ISAARA OPATI IKEDI**

I56/11152/2018

Department of Physics

School of Physical Sciences

University of Nairobi

</div>

This thesis is submitted for examination with our approval as research supervisors:

	Signature	Date
Dr. Zephania Birech Department of Physics University of Nairobi birech@uonbi.ac.ke	…….……………….…	24/08/2022 …….…
Dr. Ian Kaniu Department of Physics University of Nairobi ikaniu@uonbi.ac.ke	….…………………..…	24-08-2022 ………………………

DEDICATION

"Great things are done when men and mountains meet."

– William Blake –

To my dear parents, Mr. and Mrs. Ikedi, my brother, Ambrose Ikedi, my sisters, Immaculate, Catherine, Edith, Salome and Scovia and my niece Hope. I am grateful to be a son, a brother and an uncle in this family.

ACKNOWLEDGEMENT

ABSTRACT

Honey adulteration by cheaper sweeteners such as sugar syrups, synthetic honey, molasses, and sugar beet has become a common vice thus negatively affecting the quality of honey production, and diminishing its market value. Lack of label–free, easy to use and rapid quality assessment honey adulteration detection techniques in the market has encouraged honey producers and processors to cheat on its quality. Furthermore, the current honey adulteration detection techniques such as, Stable Carbon Isotope Ratio Analysis (SCIRA), Liquid Chromatography (LC), Gas Chromatography (GC), and High Performance Liquid Chromatography (HPLC) suffer from the disadvantages that include being less rapid and expensive to use. Hence the need for rapid and affordable honey adulteration detection techniques. In this research, laser Raman spectroscopy robustness as an emergent technique for definitive molecular fingerprint analysis was explored to study honey adulteration. Authentic honey was intentionally adulterated by molasses in varying concentration ranges. Raman spectra was collected separately with each done under 60 seconds from small quantities of 1 g of authentic honey, molasses and molasses - adulterated honey samples. PCA was employed to perform exploratory analysis of the combined authentic and adulterated Raman spectral data sets, while machine learning techniques namely, random forest (RF), and support vector machine (SVM), and artificial neural networks (ANN) were used to create multivariate classification and regression models for forecasting authentic honey and molasses - adulterated honey samples. The most variant bands between authentic honey and molasses that were confirmed using ANOVA and PCA showed characteristic bands centered around: 690 cm^{-1} (stretching of CO and CCO, and bending of OCO); 732 cm^{-1} (glucose ν(C-C) vibrations); 754cm^{-1} (weak (C = O) bond vibrations); 845 cm^{-1} (glucose spectrum); 970 cm^{-1} (glucose, ν (C- O) vibrations). Furthermore, high classification accuracies ranging from 86 – 100 % were achieved using RF and SVM classification models. Artificial Neural Networks (ANN) was built as a regression model using the concentration ranges of 0 – 10%. The coefficient of determination (R –squared) was $R^2 = 0.5786$ and the mean absolute error (MAE) was 1.51. In order to calculate the limit of detection (LOD), the training data set obtained from the ANN regression model were used to determine the LOD. The median and mean absolute deviation values of the samples with known concertation versus samples whose concertation were predicted were used to calculate the LOD

because they were found to be statistically stable, thus they yielded minimized error bars. Using the ANN model an LOD value lower than 1% was obtained. Thus, the results discussed in this research demonstrate the capability of Raman spectroscopy coupled with PCA, RF, and SVM, and ANN for molecular distinction of authentic and molasses - adulterated honey using the Raman spectral data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANN   Artificial Neural Network

CARS   Coherent Anti – stokes Raman Spectroscopy

GC   Gas Chromatography

HMF   Hydroxymethylfurfural

HMM   Honey Molasses Mixture

HFCS   High Fructose Corn Syrup

HPLC   High Performance Liquid Chromatography

IUPAC   International Union of Pure and Applied Chemistry

LC   Liquid Chromatography

LOD   Limit of Detection

MAE   Mean Absolute Error

MIR   Mid Infrared

MSS   Maltose Sugar Syrup

NIR   Near Infrared

NMR   Nuclear Magnetic Resonance

PCA   Principal Components Analysis

RF   Random Forest

RRS   Resonance Raman Spectroscopy

RMSEP   Root Mean Square Error of Prediction

SCIRA   Stable Carbon Isotope Ratio Analysis

SERS   Surface Enhanced Raman Spectroscopy

SVM   Support Vector Machines

TERS   Tip Enhanced Raman Spectroscopy

UV   Ultraviolet

CHAPTER 1: INTRODUCTION

1.1: Background to the Study

Laser Raman spectroscopy is a spectroscopic technique typically used to determine vibrational modes of molecules when such molecules interact with laser source of light (Yang and Yi, 2011). Raman scattering of light arises when incident radiation is scattered by molecules resulting in a frequency shift that is either the same as incident radiation or a shifted frequency (Smith and Dent, 2019). The frequency shifts in the molecular transitions of scattered radiation lies between rotational, electronic and vibrational level, and they constitute about 0.0001 % of the incident radiation (Wiley, 2006). The frequency shifts results into Raman effects corresponding to wavelength shifts of a vibrating molecular bond (Long, 2005). Hence, unique molecular structures that form the molecular fingerprints of the sample under study are revealed by the shifts in wavelength (Py *et al.*, 2015).

As a non- destructive technique, Raman spectroscopy involves simple sample preparation procedures that do not need dissolving of samples, or pressing of pellets, or alter the physical or chemical structure of a sample (Zeitler *et al.*, 2007). This limits the possibility of sample cross – contamination (Korth and Ralston, 2002) and reduces clean ups. Moreover, the rich and informative Raman spectra has instrumentally aced Raman spectroscopy capabilities over other analytical methods (Alula *et al.*, 2018) in definitive molecular analyses. In addition, Raman spectroscopy technique is advantageous since, it operates at wavelengths that are independent of vibration modes being studied, with tunable wavelength ranges spanning from UV to NIR being possible (Gaft and Nagli, 2008) even the far – infrared which are very difficult to access (Petry *et al.*, 2003). Furthermore, Raman spectroscopy combined analytical techniques have incredibly facilitated the progressive qualitative and quantitative assessment of food components, thus the quality of food (Yang *et al.*, 2005; Se *et al.*, 2019).

In assessing adulteration, Laser Raman spectroscopy is employed to predict levels of adulteration in virgin oil using soya bean, corn and olive residue (Baeten and Meurens, 1996). Moreover, the technique has also been used in the discrimination of olive oil from various vegetable oils as well as detecting adulteration (Donfack and Materny, 2009). The robustness of laser Raman

spectroscopy has been extensively harnessed in detecting urea adulteration in milk (Khan and Krishna, 2015), analysis of chemical profiling of medical counterfeits (Dégardin *et al.*, 2011), and the detection of butter adulteration by margarine (Selin *et al.*, 2013).

1.2: Honey Adulteration

Honey adulteration is the process of lowering the quality of honey by combining it with low-cost sweeteners that have similar properties but are of lower quality than honey (Tura and Seboka, 2020). The use of cheap sweeteners to adulterate honey is credited with the ability to artificially manufacture cheap sweeteners with profiles that resemble authentic honey (Yaacob, *et al.*, 2019). (Se *et al.,* 2019) specifically mentioned starch syrups, inverted syrups, jaggery syrup, molasses, and date syrup as common cheap sweeteners used to adulterate honey. For example, the distinct dark brown color of jaggery syrup has made it difficult to distinguish it from natural honey on many occasions (Mishra *et al.*, 2010), thus its wide use as a honey adulterant. For instance, molasses, a product of sugarcane refining is commonly used to adulterate honey. The presence of such low-quality honey on the market may have lowered the market price of genuine honey (Zábrodská *et al*., 2014), thus, limiting the country's economic growth.

Honey adulteration crept into honey production in recent decades as a fraudulent act to meet the insecure global honey supply (Zábrodská *et al*., 2014). Honey adulteration, according to (Ertelli *et al.,*2010), has become an appealing vice that helps honey fraudsters meet the ever-increasing market demand for honey and honey products. In a case study of the Czech market, (Zábrodská *et al.,* 2014) found that consumers frequently encounter bogus honey as well as honeys that are occasionally tainted in some way by deadly chemicals such as antibiotics, colorings, and Hydroxymethylfurfural. Excessive consumption of such phony honey is harmful to people's health, particularly diabetics (Fakhlaei *et al.*, 2020). Adulterants like molasses can also cause digestive issues like loose stools, according to the researchers (Gupta, 2020). Ill-health, such as obesity and high blood sugar levels (Ismail *et al*., 2018), are some of the consequences of consuming tainted honey.

1.3: Conventional Methods of Honey Adulteration Sensing.

Adulteration sensing in honey includes not only the detection of adulterants directly added to honey, but also the detection of adulterants through the indirect feeding of sugars to honey colonies (Guler *et al.*, 2014). In honey adulteration sensing, conventional methods such as thin layer chromatography (TLC), stable carbon isotope ratio analysis (SCIRA), gas chromatography (GC), liquid chromatography (LC), and high performance liquid chromatography (HPLC) are popular. TLC has been used to detect HFCS (Yaacob, *et al.*, 2019) and to investigate the authenticity of honey by looking into the ratio of fructose to glucose (Cimpoiu *et al.*, 2013). This method, however, is unreliable because more extensive works needs to be done to assess its reliability in detecting (Se *et al.*, 2019a). SCIRA, on the other hand, has proven useful in distinguishing honeys from various botanical sources (Bontempo *et al.*, 2015). GC has proven to be accurate in detecting sugar adulterants and the aroma of honey (Yaacob, *et al.*, 2019). However, using this technique necessitates time-consuming sample preparation (Matute *et al.*, 2007). While LC has been found to be effective in sensing C3 and C4 adulterants (Yaacob, *et al.*, 2019), HPLC has been found to be reliable for routine monitoring of large numbers of samples, despite being labor intensive and requiring an expert (Roussel *et al.*, 2003).

In this research, robustness of Laser Raman spectroscopy as an emergent technique for definitive molecular fingerprint analysis was explored in the detection of the molecular distinction of honey and molasses (honey adulterant) which are two compounds with almost similar molecular composition. The Raman spectra obtained from both authentic honey and honey samples adulterated with molasses were subjected to exploratory analysis to ascertain unique properties of authentic samples and adulterated samples. Similarly, classification analysis was done to check on the ability of laser Raman spectroscopy to in molecular distinction of similar compound. Furthermore, regression analysis to help in building a model that could be used confirming the authenticity of honey samples. Using principal component analysis (PCA) loadings plot, definitive spectral bands characterizing authentic honey, molasses and honey adulterated by molasses at different concentrations were identified. Moreover, for the purposes of building rapid models, principle components (PCs) were used as inputs for both classification and regression models. This greatly reduced the architecture of the models, thus making the modes to be rapid. Classification analysis were done using support vector machine (SVM) and random forest (RF) models, high classification accuracies ranging from 86 – 100 % were achieved. In addition, regression analysis

was done using artificial neural network (ANN) model in the concentration range of $0 - 10$ %. This yielded a coefficient of determination $R^2 = 0.5786$ and mean absolute error (MAE) of 1.51.

This research is ordered into six chapters: the introduction entails background information and captures a statement of the problem, the objectives, significance and justification of the study and the hypothesis of this study. The chapter on literature review presents a review of literature on honey adulteration detection by Raman spectroscopy and other spectroscopic techniques, a review of machine learning tools applied in extraction of Raman data in honey adulteration studies as well as the discussion of honey adulteration studies. The chapter on theoretical framework discusses Raman spectroscopy and its utility in various applications by considering emissions that are relevant in the interpretation of molecular vibrations that are Raman active. In addition, discussions of Chemometric techniques, namely PCA, RF, SVM, and ANN are also outlined. In the materials and methods chapter, sample preparation approaches are described in detail as well as how Raman spectral data was acquired from each sample set and preprocessed before analysis. In addition, data analysis approaches utilized in interpreting the spectral data are described. Results and discussion chapter entails a description of the collected results, analyses using various techniques and the discussion of the findings. Conclusion and recommendations chapter contains a summary of the main findings and directions for future research.

1.4: Statement of the problem

The current techniques used in assessing quality of honey are cumbersome and expensive, this makes it hard to frequently carry out quality assessments of honey at every stage of processing, marketing and selling (Kružík, 2017). Thus there is a compelling demand to study and create techniques that are rapid, non-destructive and portable that can easily be used for routine assessments of the quality of honey. If more rapid techniques are developed, it will significantly ensure quality food products are sold in the market, thus the economic viability of each food product such as honey will be realized (Zábrodská *et al*.,2014).

1.5: Research objectives

1.5.1: Main objective

To create an effective quantification and prediction model for assessing honey adulteration through analysis of Raman spectra using PCA, RF, SVM, and ANN

1.5.2: Specific Objectives

 i. To obtain characteristic Raman profiles and identify unique Raman bands of molasses, authentic honey and molasses – adulterated honey.
 ii. To perform exploratory data analysis using multivariate calibration on the collected Raman spectra of authentic honey and adulterated honey using selected Chemometrics techniques
 iii. Build a prediction model from the collected Raman spectra of authentic and adulterated honey and employ it in determining limit of detection.

1.6: Significance and Justification of the study

The consequences of honey adulteration as a vice range from health risks such as high blood sugar levels and obesity to economic risks which include and not limited to brake – down of the economy due to low prices of adulterated honey and loss of consumer confidence on the quality of honey (Ismail *et al*., 2018). The shortcomings due to ineffective honey adulteration detection techniques in the markets today has also greatly contributed to the increase in honey adulteration fraud (Jaafar *et al.*, 2020). The current techniques are unable to cope with the newest and sophisticated adulteration methods, and in so doing low quality honey gets its way into the market thus diminishing the market price of authentic honey (Zábrodská *et al*., 2014). In addition, the current techniques of honey adulteration detection are cumbersome, use expensive chemicals with some being able to detect only one adulterant (Cotte *et al*., 2007; Bougrini *et al*., 2016). These pertinent issues call for a solution. It is for such reasons that this study explores the adequacy of laser Raman spectroscopy coupled with machine learning techniques as a robust definitive molecular analysis method in the assessment of honey adulteration.

The scientific significance of this work is geared towards improving and adding to the existing spectral libraries on honey adulteration using laser Raman spectroscopy. By identifying unique and prominent spectral bands that significantly contribute to honey adulterated by molasses, this

research has proved the possibility of laser Raman to be used as a hands on technique for routine monitoring and assessment of honey quality.

1.7: Scope and limitation of the study

Production of natural honey under varying climatic conditions has made it an easy target of adulteration by low – cost sweeteners. In the market today, honey adulterants range from industrially manufactured syrups as well as indirect feeding of honey bees on sugar syrups. In this work, molasses, a highly viscous by – product from sugar refining is studied as honey adulterant. On its own, the molecular composition of molasses is majorly dominated by fructose, glucose, and sucrose which are also the basic compounds characterizing the honey profile.

1.8: Hypothesis of the Study

It is possible to create a quantification scheme for honey adulteration detection lower than 1%. Using the distinct vibrational bands present in honey together with the target adulterant, the prediction of trace adulteration levels is achievable with prediction models being developed from a range of low adulteration levels of honey.

CHAPTER 2: LITERATURE REVIEW

2.1: Chapter Overview

This chapter presents a review of literature on honey adulteration detection by Raman spectroscopy and other spectroscopic techniques. A review of Chemometric techniques applied in extraction of Raman data in honey adulteration studies has also been presented. A discussion of honey adulteration studies has also been done.

2.2: Honey Adulteration Detection Using Raman Spectroscopy

Raman spectroscopy has been utilized in the detection of honey adulterated by fructose, insulin syrup, glucose, and malt must (Oroian *et al.*, 2018). Groups of authentic samples and samples at different adulteration levels were created for analysis. The classification analysis using PLS – LDA yielded a total accuracy of 96.54 % and 90.00 % for authentic vs. adulterated and adulterated honey respectively (Oroian *et al.*, 2018). When authentic honey was examined further, it revealed bands at: 346, 353, 408, 498, 606, 681 and 793 cm^{-1} which were assigned to respective molecular components and vibrations. In addition, the band at 1048 cm^{-1} was associated to glucose ring vibration, at 1054 cm^{-1} CH and COH bending vibrations in carbohydrates were found to be responsible with some CN bond vibrations in proteins and amino acids also contributing, and at 1238 cm^{-1} COH bond vibrations were prominent (Corvucci *et al.*, 2015; Li *et al*., 2017; Özbalci *et al*., 2013).

Honey tainted by fructose exhibited skeletal intensity at 606 cm^{-1}, ring vibration and stretch of C-OH at 750 – 850 cm^{-1} and C – O – C cyclic alkyl ethers at 1074 cm^{-1}. For honey adulterated with glucose, glucose specific wavenumbers were noted at 498 and 918 cm^{-1}. For honey adulterated by hydrolyzed insulin syrup, characteristic band with high magnitudes in intensity were realized at 450 , 680, 793, and 1253 cm^{-1} (Li *et al*., 2017). Furthermore, honey adulteration by inverted sugar showed characteristic bands at 681 cm^{-1} and 793 cm^{-1} which corresponded to fructose molecule, and glucose ring vibrations respectively (Corvucci *et al.*, 2015; Li *et al*., 2017; Özbalci *et al*., 2013).

2.3: Detection of Honey Adulteration by high fructose corn syrup (HFCS) and maltose syrup (MS) using Raman spectroscopy

Li *et al*., (2012) applied Raman spectroscopy in uncovering high fructose corn syrup (HFCS) and maltose syrup (MS) adulterants in honey. Authentic honey and HFCS and MS were mixed in different ratios. Analysis of the Raman spectroscopy results was done using PLS – LDA. The results proved the viability of Raman spectroscopy coupled with PLS –LDA as a probable method for disclosing adulterants in honey (Shan, *et al.*, 2012). As noted by Paradkar and Irudayaraj. (2001), the Raman band found present at 705 cm^{-1} corresponded to CO and CCO stretching, and bending of OCO. CH vibrations characterized the bands at 865 and 824 cm^{-1} while CH and COH vibrations contributed to the signal at 915 cm^{-1} and 1065 cm^{-1}. Raman bands at 1127 cm^{-1} were linked to C – O stretching vibrations whereas COH vibrations contributed to 1264 cm^{-1} and at 1373 cm^{-1} was linked to CH and OH bonds bending. The combination of CH bending vibrations and the vibration of COO- group was associated with the band at 1461 cm^{-1} (Shan, *et al.*, 2012).

2.4: Detection of Honey Adulteration Using Chemometrics – Integrated Raman Spectroscopy

Raman spectroscopy coupled with Chemometrics has effectively been utilized in the quantification of various sugar adulterants (Özbalci *et al.*, 2013). For instance, Li *et al* (2012), employed Raman spectroscopy in studying honey adulterated by HFCS and MSS. Using the PLS –LDA discriminant model, high total accuracies of 91.1 % and 97.8 % were achieved for HFCS and MSS samples respectively. Furthermore, (Kneipp *et al.,* 2002) hinted on the progressive utilization of Raman spectroscopy in assessment of foodstuff safety and quality. In this study, Kneipp harnessed the advantageous uniqueness of Raman spectroscopy and its ability to combine both infrared spectroscopies alongside NIR spectroscopy. Furthermore, (Mignani *et al.,* 2016) devised an ingenious compact dispersive Raman spectroscopy (at 1064 nm stimulated) combined with Chemometrics data treatment to check out traits of honey depending on diverse markers such as carbohydrate profile. The quantitative and qualitative analyses of excess sugars in comparison to data acquired from Raman spectra proved the viability of the technique in predicting main sugars (Mignani *et al.*, 2016).

2.5: Other Spectroscopic Techniques Used in Honey Adulteration Detection

2.5.1: Chemometrics-Integrated Infrared Spectroscopy

The combination of IR spectroscopy with Chemometrics has become an instrumental tool in the detection of honey adulteration. This method is fancied for a variety of reasons, including its speed, ease of sample preparation procedures, cheaper, non-destructive, and suitability in actual monitoring (Velázquez *et al*., 2009; Anguebes *et al*., 2016), thus making it a famous method in honey adulteration detection. Frequently, IR spectra can reveal concealed details about a sample's constitution. Matrix effect and overlapping, that may cause standalone IR spectroscopy less capable compared to aforementioned methods for detecting adulterated honeys, could be overcome by bringing it together with Chemometrics statistical analysis. Moreover, NIR spectroscopy, coupled with PLS – LDA and Coherent Anti –Stokes Raman Scattering (CARS) have been effectively applied in examining quality of honey samples containing high fructose corn syrup (HFCS) and maltose sugar syrup (MSS), coupling the method with CARS has helped improve the classification accuracies (Ma *et al.*, 2017). Although this method could not predict the degree of HFCS adulterant, it was found to be accurate for MSS prediction (Yaacob, *et al.*, 2019).

2.5.2: Nuclear Magnetic Resonance Spectroscopy (NMRS)

NMRS utilization in combination with statistical models to evaluate the validity of honeys from different flora and geographical locations has been fruitful (Spiteri *et al.*, 2014). When compared to other techniques, NMRS has been insufficiently sensitive in detecting adulterants of ISS and HFCS in honey. For instance, (Ribeiro *et al*., 2014) used NMRS to distinguish HFCS adulterated honey from pure blossom honey, and discovered significant differences in transverse relaxation time between the two honeys (Oliveira *et al*., 2014; Spiteri *et al*., 2014). In general, NMRS enables the quick detection (under 5 minutes to receive one NMRS spectrum), the simultaneous measurement of adulterants and several chemical compounds from the same spectrum. However, the high cost of operation of the NMR that needs an expert restrains its large use as do other high-end analytical techniques (Ertelli *et al.*, 2010).

2.5.3: Chemometrics-Integrated High-performance Anion-Exchange Chromatography Coupled with Pulsed Amperometric Detection (HPAEC –PAD)

HPAEC-PAD technique is well-established in determining carbohydrate content of compounds. Majority of HPAEC-PAD applications use a high pH mobile phase because numerous carbohydrates are not charged at pH 7. Additionally, using a gold working electrode and oxidation at a high pH, carbohydrates can be detected (Corradini *et al.*, 2012). Electron flow through the working electrodes over time generate charge that yield to a chromatogram. As such the potentials created overtime result to pulses Amperometric detection (PAD) (Mellado-mojica *et al.*, 2016). In an adulteration study, HPAEC – PAD was used to identify thirteen different sugars syrups that had been used to spike pure honey. It was discovered that, combining supervised classification with LDA was effective in differentiating pure honey from adulterated honey, categorizing distinct varieties of honey, together with designating a sample to an actual category (Cordella and Milit, 2005). High classification accuracies at 96.5 percent using the LDA model proved to be successful in categorizing the various types of honeys.

2.6: Tools Used to Extract Raman Spectral Data for Analysis

Principal Component analysis (PCA) as a multivariate tool has been applied in the comparative examination of Raman spectra and IR spectra of commercialized honey in the markets of Ecuador. This was done to verify the quality of the honey. Of the 8 samples that were analyzed, two of them were found to have a huge disparity in the composition of sucrose and reducing sugars, thus shaping two clear clusters that varied in relation to samples analyzed (Salvador *et al.*, 2019). In another study involving (Vis-NIRS) coupled with Chemometric tools, PCA was utilized as a linear technique for dimensionality reduction. Through the representation of loading values, the wavelengths that had the great significance of the first three PCs looked associated to the level of adulteration for almost all the wavelengths. Although there was no complete difference that was possible using this technique, the obtained results suggested that PCA displayed a trend for the adulterated samples to be grouped in relation to the level of adulteration (González *et al.*, 2018).

In a study of virgin oil adulteration, PCA was employed to scale down the multidimensional data acquired from Raman spectroscopy. Thus, the discrimination of oil samples based on origin or

their composition was realized (Ez *et al.*, 2003). Likewise, SIMCA, Chemometric analysis was successfully applied for the correct classification of the origin of Mexican honeys from four dissimilar regions (Velázquez *et al.*, 2009). Furthermore, Vibrational spectroscopy coupled with multivariate methods as PCA and HCA can be used for laying out investigated elements into categories on the basis of their comparison, thus can be employed in analysis of spectral data for differentiation of unadulterated wheat flour and wheat flour adulterated with L-Cysteine and Cysteine (Cebi *et al.*, 2017).

In a study of forecasting beef adulteration by blight beef using Vis – NIR, support vector machine (SVM) was used to create a segregation hyperplane of the valid beef and blighted beef (Zhao *et al.*, 2019). Similarly, in studying honey adulteration by hyperspectral imaging, the SVM algorithm had an accuracy of 92 % (Shafiee *et al.*, 2016). Moreover, using Gas Chromatography, the SVM algorithm has been used to build a model with 100 percent classification accuracy to distinguish between authentic sesame oils and adulterated sesame oils adulterated (Peng *et al.*, 2015).

Random Forest (RF) has been used in food adulteration detection as a single-class classifier and in infrared spectroscopy. In order to use the random forest, artificial outliers from the target samples had to be generated, and all possible adulterations had to be simulated. The algorithm used affluently generated invariably scattered artificial samples in all directions close to the intended categories, involving the entire realm of variations and allowing the expansion of a suitable random forest model (Bachion *et al.*, 2019). Researchers have used e-nose measurements and RF in differentiating distinct samples of apple juice (Ro, 2018). In a study involving rapid segregation of authentic and adulterated Andiroba Oil, RF was used to build classification models, which showed 100% of correct classification. Thus, this method could be crucial in periodic assessments of adulteration (Santana *et al.*, 2018).

Using electronic nose, the application of ANN to unearth the adulteration levels in camellia seed oil and sesame oil was insightful. The outcome indicated that, while the electronic nose can't forecast the adulteration percentage in camellia seed oil using ANN as a pattern identification technique, it can be applied to quantify sesame oil (Hai and Wang, 2006). In adulteration sensing in saffron samples using electronic nose, ANN affirmed the possibility of distinguishing saffron from saffron mixtures containing graft materials with accuracies of 86 percent (Heidarbeigi *et al.*,

2015). Similarly, ANN has been applied successfully in detecting cow ghee adulteration by of margarine using e –nose machine with accuracies as high as 85.6 and 97.2 percent (Ayari and Ghaleh, 2018). In a study on coconut oil adulteration, ANN data sets were developed to validate and test the ANN models. This system was found to uncover coconut oil adulteration by sunflower with an accuracy of 99.53 percent and 98.82 percent. Thus, suggesting that this technique could be useful in the creation of a portable sensor for scanning of adulteration in edible oil (George *et al.*, 2017).

2.7: A Summary of Honey Adulteration Detection Techniques Studies

Exploration of rapid and sensitive honey adulteration techniques has been an ongoing process in research (Yaacob, *et al.*, 2019). Lately, Wu *et al.*, (2017) comprehensively analysed the applications of chromatographic and spectroscopic techniques in honey adulteration detection. He established that, chromatographic techniques namely using GC (Matute *et al.*, 2010) , LC (Wang *et al.*, 2015), HPLC (Du *et al.*, 2015; Wang *et al.*, 2015)) and HPAEC- PAD (Morales *et al.*, 2008) had been applied successfully in the identification of a number of honey adulterants. Notably, using SPE – HPAEC – PAD (Megherbi *et al.*, 2009) a 1 % limit of detection of honey adulteration had been achieved. Conclusively, Wu established that, the practical applications of spectroscopic techniques was more robust and remarkable than the chromatographic techniques when identifying adulterants present in honey (Yaacob, *et al.*, 2019). This was because spectroscopic techniques are rapid and simple to use.

In comparison to spectroscopic methods, traditional chromatographic methods are tedious, limited to expertise operation and utilize toxic and expensive chemicals in their analysis. In addition, some of the methods can only identify one honey adulterant (Cotte *et al*., 2007; Bougrini *et al*., 2016). Hence limiting their application in qualitative and quantitative assessment of honey adulteration. On the other hand, spectroscopic techniques cannot be applied singly in the identification of adulterant markers in honey, hence their integration with Chemometrics methods to enable derivation of important spectral information. In literature, Chemometric – integrated spectroscopic techniques have been used to test for honey authenticity (Sivakesava and Irudayaraj, 2002) as well as identification of microscopic figurative markers that show honey adulteration by sugar syrups (Cordella *et al.*, 2005). Thus robust systems built using Chemometric – integrated

spectroscopic techniques has caught the attention of researchers in classifying honey and detecting honey adulterants (Yaacob, *et al.*, 2019).

In summary, Se *et al.,* (2019) did a comparative distinction of the techniques that have already been used for honey adulteration detection as shown in **Table 2.1**.

Table 2.1: Comparative study of distinctive methods of honey adulteration detection

Method	Chemometric method	Adulterants	Limit of Detection	Reference
SCIRA	No Tool	HFCS, glucose syrup and saccharose syrup	LOD > 7%	Tosun 2013 (Cotte *et al.*, 2007)
NMR	ICA	Beet and cane full High fructose corn syrup, High fructose inverted syrup and Inverted syrup	LOD > 10%	(Spiteri *et al.*,2015) (Oliveira, Ribeiro, Mársico, *et al.*, 2014)
FT-Raman spectroscopy combined with Chemometrics	PLS, PCR, LDA, and CVA	Inverted beet and cane syrup	96% Correct classification accuracy acquired.	(Paradkar *et al.*, 2001) (Cannavan, 2014) (Pierna *et al.*, 2011)
NIR integrated with Chemometrics	CARS; PLS-LDA and PLSR	HFCS and maltose syrup	HFCS and MS adulterants classified with 96.3 and 96.1%	(Zhu *et al.*, 2010) (Li *et al.*, 2017)
FTIR-ATR coupled with Chemometrics	PLS, PCR	Corn Syrup, High fructose corn syrup and Inverted sugar syrup	$R^2 > 0.976$, SEP < 3.	(Velazquez *et al.*, 2009) (Ranjan *et al.*, 2021)
Reversed phase SPEHPAEC-PAD		COSS	LOD > 1%	(Megherbi *et al.*, 2009) (Taylor *et al.*, 2014.)
HPLC-RID		Starch syrup	LOD > 2.5%	(Wang *et al.*, 2015) (Tahsin *et al.*, 2014)
E-tongue coupled with Chemometrics		Sucrose and saccharose syrup	LOD > 2%	Bougrini *et al.*, 2016 (Rifna *et al.*, 2022)

2.8: Summary of Literature Review

The application of the aforementioned Chemometric – coupled Raman spectroscopic techniques and other techniques reported in **Table 2.1** have remarkably proven the viability of these systems in detection of adulterants in honey. However, for spectroscopic techniques, fluorescence caused by excitation from the laser radiation can swamp the weak phenomena of Raman spectra thus making collection of quality data difficult, thus posing as a challenge to the quality of Raman spectral data. Similarly, in a study done by Paradkar *et al.,* (2001), mining of Raman data was a major challenge. Thus in this work, Raman spectral preprocessing entailed smoothening using Vancouver Algorithm based on fifth – order polynomial fitting method (Birech *et al.*, 2020). In addition, spectral normalization was done using R software. Furthermore, the development of better and effective classification and regression models was done using robust machine learning techniques, namely; Random Forest (RF), Support Vector Machines (SVM), and Artificial Neural Network (ANN).

CHAPTER THREE 3: THEORETICAL FRAMEWORK

3.1: Chapter Overview

This chapter discusses Raman spectroscopy and its utility in various applications by considering emissions that are relevant in the interpretation of molecular vibrations that are Raman active. Types of Raman spectroscopy and their applications are also discussed in brief. In addition, discussions of Chemometric techniques, namely PCA, RF, SVM, and ANN are also outlined.

3.2: Theory of Laser Raman Spectroscopy

Raman scattering is an inelastic phenomenon (Li and Church, 2014a) that occurs when incident laser light interacts with a sample. When the electric field of the laser interacts with a sample, molecules at the ground vibrational electronic states are excited and are promoted to a virtual state (a non – existent state) below an electronic state (Li and Church, 2014b). Relaxation from the virtual state may occur through Rayleigh scattering giving out the same energy as incident laser, or stokes Raman shift scatter which yields energy lower than the incident laser or anti – stokes Raman scatter which emits energy higher in magnitude than that of the incident laser (Kudelski, 2008).

Conventionally, when incident laser light interacts with a sample of wave number V_o most of it is transmitted through Rayleigh scattering, but some of it is scattered. New wave numbers of the type $V_o^{'} = V_o + V_m$ will be observed with V_m associated with Raman scattering which entails, rotational, vibrational and electronic levels (Wiley., 2006). The sample interaction with incident laser of a wavenumber V_o causes a photon to be absorbed, thus it is promoted from a low to a high energy level E1 and E2 respectively, therefore ΔE = E2 – E1. The energy change ΔE between the two levels is denoted by a wavenumber V_m, where ΔE = hcv_m. Likewise, when interaction involve energy transitions from a higher level E2 to a lower level E1, energy is made available and E2 – E1 = hcv_m (McCreery., 2001). Thus the quantization of losses in photon energy (Kudelski., 2008).
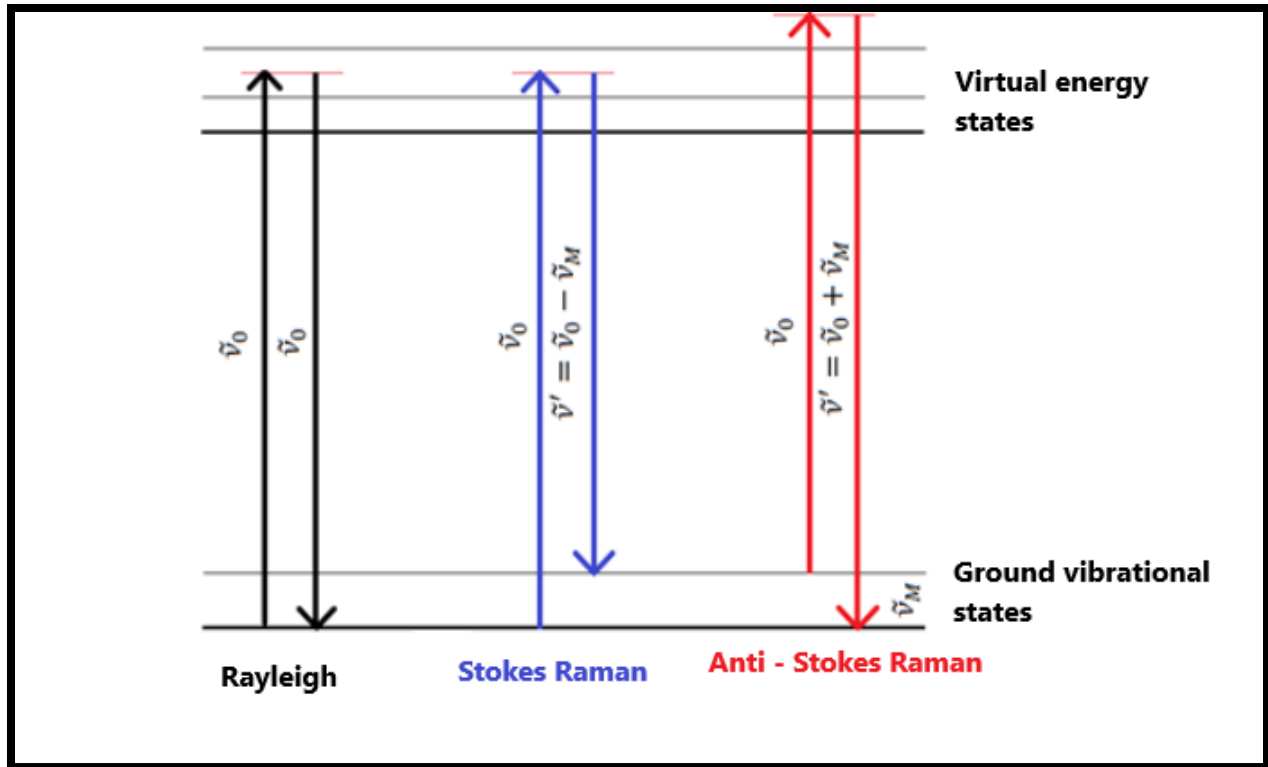
Figure 3.1: Diagrammatic representation of energy transitions between ground vibrations sates and the virtual states of Rayleigh scattering, Stokes Raman and Anti –stokes Raman scattering (Source: Krishnan, 2019).

Descriptively, when the electric field of a laser combines with the molecules of a sample, the molecules are excited in a specified geometrical orientation, thus polarization. Relaxation of molecules from the polarized condition occurs almost immediately and molecules return to their initial state at a process called Rayleigh scattering. Relaxation of the molecule to the first excited vibrational level results in Stokes – Raman shifts whereas relaxation from an excited vibration level to the ground state produces anti – stokes Raman scattering (Tan *et al.*, 2019). It is from such relaxation process that Raman active molecules with polarizability that have even symmetry, stretching vibrations as well as vibrations that do not generate electric field become possible (Bosshard *et al.*, 2002). Therefore, Raman spectroscopy becomes capable of carrying out both qualitative and quantitative analysis of samples by exploring the spectral signatures (Araujo *et al.*, 2018).

3.3: Sample and Raman Intensity

Raman scattering intensity is an important quantity in Raman measurements because it depends mainly on how many molecules are present per unit volume of sample examined by Raman instrument (Zhao *et al.*, 2021). Raman spectroscopy is a technique for determining a molecule's intensity by examining its Raman intensity. The association between Raman intensity and molecule concentration lays foundation of quantitative Raman spectroscopic analysis whereas qualitative Raman analysis depends on the weighted sum of the Raman spectra of the components constituting a mixture (Haynes *et al.*, 2005).

3.4: Variants of Raman Spectroscopy and Their Advantages

The laser wavelength chosen is vital for performance of Raman and has a significant impact on the spectrum obtained (Krafft *et al.*, 2009). Field of view for Raman measurements is dependent on the laser wavelengths, thus resonance effects (Pettinger *et al.*, 2012). Shorter wavelengths are more easily noticeable with lower scatter effects and quantum efficiency. Because more electronic transitions occur in the ultraviolet (UV) than in the infrared regions of the spectrum, shorter wavelengths are also more likely to excite fluorescence. The wavelength of the laser should be as long as the sensitivity demands allow (Nakamoto and Brown, 2003).

3.4.1: Frequency Precision and Accuracy of Raman Shifts

Assignment of Raman features, library searching, and spectral subtraction depend on reproducible Raman shift values (Raanan *et al.*, 2018). Frequency accuracy becomes increasingly important as spectral databases become available, since spectra from a wide variety of spectrometers will contribute to the database. Unfortunately, peak frequencies in the literature often vary by several reciprocal centimeters for a given Raman feature, but accuracy of substantially better than ±1 cm-1is routinely achievable. In some spectrometers, < 1 cm-1 accuracy is both routine and automatic (McCreery, 2001).

3.4.2: Reproducibility of Relative and Absolute Peak Intensities of Raman Shifts

Although less important than peak frequencies, peak intensities obviously play a role in quantitative analysis and, in many cases, qualitative identification (Blum *et al.*, 2014). Multivariate calibration techniques and their transferability depend on reproducible relative peak heights. A possibly lengthy method development procedure may fail when a different spectrometer is used,

if the observed intensities vary. Reproducibility of absolute signal is difficult to achieve between labs or even between instruments of the same design, but it is important for a particular instrument. Absolute intensities can at least be used to evaluate day-to-day instrument performance and to detect hardware or alignment problems (McCreery, 2001).

3.4.3: Available Laser Power

The figures of merit are normalized to a given laser power since they apply to collection rather than excitation of Raman scattering (Müller and Sumpf, 2020). While high laser power is vital in this facet, there are limitations. High power, or more precisely high power density, can cause thermal or photolytic damage to the sample, ultimately limiting the acceptable power density. A second limitation on laser power is more pragmatic, involving cost or utility requirements. With today's technology, lasers with outputs of greater than about 500 mW require cooling water or external heat exchangers and are impractical for many requirements (Liu *et al.*, 2019).

3.5: Utility of Raman Spectroscopy in Molecular Analysis

As an analytical technique, Raman spectroscopy blends both pros of IR and NIR spectroscopy for molecular analysis (Das and Agrawal, 2011). Likewise, its operation based on frequency shifts that correspond to molecular vibrations of the sample under study have given Raman spectroscopy an edge in the study of molecular composition of materials (Kneipp *et al.*, 2002). Furthermore, Raman enjoys molecular analysis since no two molecules can yield exactly the same Raman spectrum, thus enabling qualitative and quantitative analysis of samples for correct spectral interpretation (Wiley, 2006). In summary, as a key vibrational spectroscopy technique, Raman spectroscopy has enabled the identification and study of molecular composition of materials through observation of the Raman spectral bands (Schmitt and Popp, 2006).

Raman spectroscopy comes in various types depending on the interaction between the laser and the sample. Non – resonance Raman scattering occurs when electrons in molecules are polarized after interacting with incident radiation (Lin *et al.*, 2021). Conversely, resonance Raman scattering entails very close excitation frequencies between the incident radiation and the electronic transition of the molecule of interest (Robert, 2009). Whenever a sample interacts with incident radiation of very high irradiance it results into nonlinear processes during scattering. This is called non – linear Raman spectroscopy (Snežzana, Leo and Biljan, 2007).

3.6: Surface Enhanced Raman Scattering (SERS)

SERS is a very sensitive method that increases the Raman scattering of molecules by utilizing some nanostructured metallic substrates (Iqbal *et al.*, 2020). The adoption of SERS for spectral measurements emanates from the increase in the observed cross – section area of the sample under study by magnitude as far as 15 orders in comparison to the normal Raman scattering (Zheng *et al.*, 2020). The SERS effect causes an increase in Raman intensity as a consequence of an enhanced electric field at the metal surface (McNay *et al.*, 2011) thus, chemical and electrical field enhancement is realized (Haynes *et al.*, 2005).

Currently, the application of SERS in spectral analysis has led to the possibility of declaring Raman spectrum for a single molecule (Etchegoin *et al*., 2003). Interestingly, SERS spectroscopy has been applied in analyzing complex mixtures by measuring many SERS spectra from just a few molecules of a sample randomly selected (Kudelski, 2008). Furthermore, the applications of SERS have been demonstrated to be beyond laboratory research problems (Sharma *et al.*, 2012). For instance, UV SERS has been effectively applied in analysis of protein residues and DNA bases (Cui *et al.*, 2010) with ultrafast spectroscopies capable of generating enhancement factors that could exceed 10^{20} (Oliver, 2018). In addition, investigations on plasmon – enhanced photocatalytic reactions and ultrafast reaction dynamics have been realized through ultrafast spectroscopies (Chuntonov and Rubtsov, 2020).

3.7: Tip –enhanced Raman spectroscopy (TERS)

TERS is a technology that couples Raman spectroscopy with aperture-less scanning near-field optical microscopy (s-SNOM) to achieve Raman spectroscopy and Raman imaging beyond the diffraction limit of the probing light (Verma, 2017). The efficiency of TERS rides on scanning resolution of the probe and sensitivity (Kumar *et al.*, 2015). TERS functions on the surface of the sharp metal tip to excite Raman scattering from small volume (Verma, 2017). Thus, enhanced Raman intensity of molecules place close to such mental structures (Yeo *et al.*, 2009).

As such, the interrogative capabilities of TERS has been applied successfully in analysis of samples with sub – diffraction – limited capabilities (Meyer *et al.*, 2017) as well as analysis of involving solid-state physics, organic or inorganic chemistry or biochemical and cell biological issues (Deckert and Wiley, 2009).

3.8: Machine Learning Techniques Applicable for Raman Spectroscopy Analysis

3.8.1 Principal Component Analysis (PCA)

PCA is the real workhouse of exploratory data analysis (Abdelghany, 2016). This technique is used to show disparities between huge amounts of data obtained from a spectrum. By reducing the data to a minimum number of principle components (Sivakesava *et al*., 2001), PCA boils down the entire data set into essentials that describe the variation in the data. Thus, as an unsupervised statistical method, most of the linear latent variables that explain for the variance in the observed variables are computed (Lenhardt *et al.*, 2014). The triumph of this technique leans on the selection of preferred scope of spectrums and how many variables are assigned for use in the model (Paradkar *et al*., 2002).

PCA is the upmost frequently used multivariate analysis calibration method employed in spectroscopy for extracting relevant features from complex spectral data containing overlapping regions (Paradkar and Irudayaraj, 2002). In addition, PCA is efficient in distinctive identification of data patterns that reveal similarities and differences in the data by using correlations between the variables of interest (Lenhardt *et al.*, 2014). Thus, the few principle components obtained after a PCA gives a full representation of the entire data set (Sivakesava and Irudayaraj, 2001). Conclusively, PCA triumphs with proper selection of spectral ranges besides the number of variables engaged in the model (Paradkar and Irudayaraj, 2002).

PCA is capable to put up with all variables in an entire data set structure with applications to an X – matrix data with no Y – data properties being possible (Eriksson *et al.*, 2006). Although outliers and scaling methods greatly influence PCA results, constant variables have been found to have insignificant effects on PCA results (Brown, 2010). Given a data set, the first principal component is that latent variable responsible for conserving relative distances between the objects and displays the highest variance of the scores. This first principal component is defined by a loading vector;

$$p_1 = (p_1, p_2, \ldots\ldots, p_m) \tag{3.7.1}$$

Where p is loadings in PCA while m is variables' number.

In Chemometrics, the loading vectors lengths are normalized to 1: that is

$$p_1^T p_1 = 1 \qquad\qquad\qquad (3.7.2)$$

For example, given an object $\dot{\imath}$, defined by a vector x_i containing elements x_{il} to x_{im} , its PC 1 score t_{il} is expressed as:

$$t_{i1} = x_{i1} + x_{i2} p_2 + \ldots + x_{im} p_m = x_i^T . p_1 \qquad\qquad (3.7.3)$$

Hence, in the X – data matrix, all the objects arranged in rows, the score vector t_1 is defined as

$$t_1 = x.p_1 \qquad\qquad\qquad (3.7.4)$$

Furthermore, the second principal component (PC2) characteristically defines the highest possible variance of scores and is orthogonal to PC1. Generally, any subsequent PCs computed are notably orthogonal to the previous PCs and they cover the highest possible variance of the data in their direction of projection. Mathematically, the scalar product of orthogonal vectors is zero, and so considering all pairs of PCA loading vectors we have:

$$p_j^T . p_k = 0 \qquad j,k = 1, \ldots\ldots m \qquad\qquad (3.7.5)$$

Conclusively, as the number of PCs increase, the variance becomes very small or zero. Thus, for many practical data sets the first two or three PCs containing the highest variance are used for data interpretation and for scatter plots.

3.8.2: Random Forest (RF)

Random Forest is an excellent technique for classification and quantification problems (Genuer *et al.*, 2012). RF is a classifier that consists of a group of tree-structured classifiers with the following structure: $\{h(x, \Theta_k), k = 1, \ldots\}$ where the Θ_k are the independent identically distributed random vectors. Every tree in the group draws one unit of vote to the class that is more prevalent at input x. (Liu *et al.*, 2012).

As a member of the ensemble methods, it utilizes boosting and bagging of classification of trees for results aggregation (Liaw and Wiener, 2014; Genuer *et al.*, 2012). Expressly, boosting utilizes the additional weights yielded by consecutive trees in predicting incorrect points, and in doing so, its prediction relies on a weighted vote. On its path, bagging independently constructs trees using

bootstrap sample of the data, thus a simple majority vote is taken as the end result for prediction (Breiman., 2001). According to Breiman's RF model, every tree is put using a training data set and a random variable. The random variable for the kth tree is designated as Θ_k , and any two of these random variables can be used to create a classifier $h(x, \Theta_k)$, where x is the input vector. The classifiers sequence $\{h_1(x), h_2(x), \ldots\ldots\ldots h_k(x)\}$ is obtained after k times of running, and we use it to create multiple classification model systems (Rodriguez-Galiano *et al.*, 2015) (Liu *et al.*, 2012). The final outcome of these systems is determined by a simple highest vote, with the decision function

$$H(x) = argmax \sum_{i=1}^{k} I(h_1(x) = Y) \tag{3.7.6}$$

where Y is the output variable, I(.) is the indicator function, H(x) is a combination of classification models, and h_i is a single decision tree model. Each tree has the right to vote for the best classification outcome for a particular input variable. Specific process shown in **Figure 3.2.**
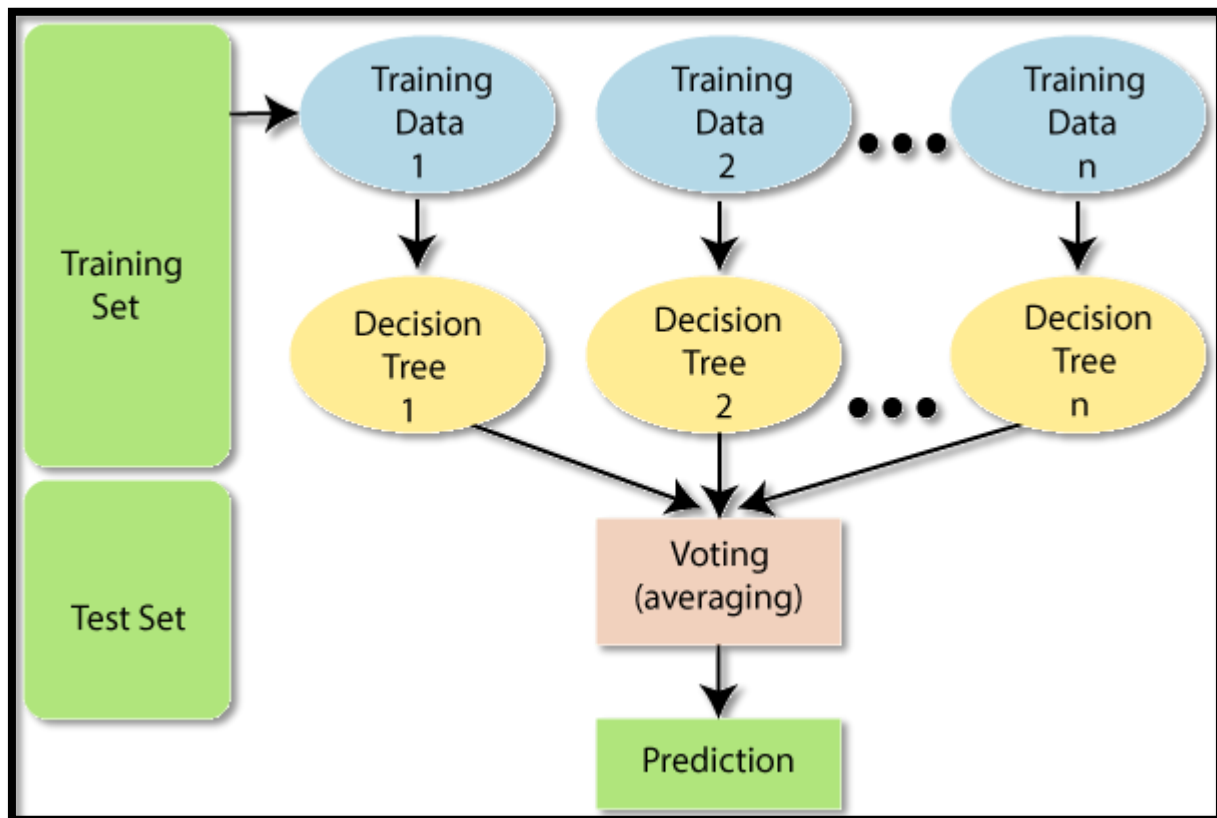
Figure 3.2: Random Forest schematic

As a classification method, random forest applies a collection of decision trees where training is done on an arbitrary subgroup of data included in the training set using arbitrarily selected subset variables (Statnikov *et al*., 2008; Anghelone *et al*, 2015). Inside the random forest, splitting at each node highly depends on the best predictors randomly chosen from that node (Breiman, 2001). Although there are nonlinear convoluted high order interaction effects, it yields variable relevance for each predictor variable (Strobl *et al.*, 2007). Thus making random forests an authoritative tool to use in multivariate classification to naturally measure similarity within different samples (Segal, 2004;Breiman, 2001).

The choice of the variable of importance for a random forest model is based on three different issues: firstly is the sensitivity of the samples size and the number of variables. Secondly, the sensitivity of mtry and ntree as method parameters are central in the optimization of a random forest model. Thirdly, for highly correlated data sets, the determination of the variable of importance determines the robustness and effectiveness of the developed model (Liaw and Wiener, 2014). In constructing a random forest model, the two main parameters, mtry and ntree are of great importance. For instance, if good quality trees are chosen, the out – of – bag (OOB) error is minimized. This yield a model with good performance (Mitchell, 2011). The OOB of samples is used in defining the correlation among the tress, estimation of error of prediction, and evaluating the variables of importance (Genuer *et al.*, 2012). Therefore, the choice of good RF model parameters is important for building better discrimination models as well as stabilizing the score of the variables of importance (Liaw and Wiener, 2014).

3.8.3: Support Vector Machine (SVM)

SVM technique is majorly utilized for classifying and quantifying analyses (Chang and Lin, 2011). As a classification technique, SVM models transform original data spaces into higher dimensions' spaces. This enables linear division of the data groups into classes, hence solving pattern recognition problems (Vandewalle, 1999). In its operation, SVM models create hyperplane that ensures a maximum margin exists between the closest support vectors separating classes of data (Lenhardt *et al.*, 2014).

Conventionally, classification models have been utilizing the density of classes in finding a separating surface between classes. This approach however is limited by the Hughes effect which calls for dimensionality reduction of data by feature selection (Qi *et al.*, 2011). Reportedly, SVM

models do not suffer this limitation because they seek an optimized separating surface that finds support vectors that set the demarcation of the classes (Gualtieri and Chettri, 2000). Generally, in p dimensional space, the hyperplane corresponds to a space given by p -1 (Brown, 2010). Remarkably, in 2- dimensions the equation of a separating hyperplane is given by

$$\beta_o + \beta_1 X_1 + \beta_2 X_2 = 0 \tag{3.7.6}$$

Here β_o, β_1, and β_2 are parameters and if there is a value $X = (X_1, X_2)^T$ for which the equation holds true, the point can be located on the hyperplane. Conversely, if X does not satisfy the above equation

$$\beta_o + \beta_1 X_1 + \beta_2 X_2 > 0 \tag{3.7.7}$$

Or, X can also take the form

$$\beta_o + \beta_1 X_1 + \beta_2 X_2 < 0 \tag{3.7.8}$$

In which case, the value lies on the other side of the equation

Therefore, for a p dimensional space, the equation of the hyperplane takes the form

$$\beta_o + \beta_1 X_1 + \beta_2 X_2 + ... + \beta X_p = 0 \tag{3.7.9}$$

Hence equally, a value of $X = (X_1, X_2......X_p)^T$ for which the equation above is satisfied locates the point on the hyperplane (Lin and Chen, 2013).

In real – life applications, SVM has been utilized in medical tests to determine which patients suffer from particular diseases and at the same time recommending a specific type of treatment (Yoshioka *et al.*, 2007). Moreover, SVM has been used extensively in classification of text (Joachims, 2005), recognition of facial expression (Michel and Kaliouby, 2003), as well as gene analysis (Le Thi *et al.*, 2008). Hence, Support Vector Machines can be thought as a linear classifier technique, with a peculiar kind of rule that guarantees good predictive performances on unseen data(Carey *et al.*, 2018).

3.8.4: Artificial Neural Network (ANN)

ANNs are structural systems designed to connect and create impression between input data and output data similar to the way the human brain analyzes and processes information (Deng *et al.*, 2020). ANNs models train and learn from input data. The self – learning capabilities enable them to yield better results as more data becomes available (Plumb *et al.*, 2005). Informatively, Neural network modelling has been instrumental in solving analytical problems, expressly in the fitting of multivariate data (Zupan, 2008). Thus it's application in food quality assessment (Özbalci *et al.*, 2013).

Categorically, ANNs can be applied in classification, time series forecasting and regression analyses (Gallo and Bonis, 2013). In regression analysis, neural networks are used to precisely describe the input – output relations in all situations lacking functional form whereas for time series applications, neural networks utilize past available data to predict future values. For classification problems, NNs can be trained to efficiently discriminate classes that are linearly separable along a straight line (Kwon, 2011).

To build a robust ANN model, it is vital to choose and optimize the basic parameters such as number of layers, number of neurons in the hidden layer, the learning rate, and the number of epochs of the training model (Lee *et al.*, 2016). For instance, ANN model with a few number of neurons leads to non –linearity problem whereas if too many neurons are used it results to overfitting problem (Kim *et al.*, 2019). Generally, ANN models are programmed via a learning algorithm to perform a particular task. The learning algorithm conditions the neural network to provide answers to a specific problem. The learning of the model can be unsupervised learning, where the network is trained only based on a set of inputs, beyond the respective outputs in which case a set of both the inputs and outputs are fed into the network and it learns from them (Rocchetta *et al.*, 2019). Conversely, supervised learning utilizes the backpropagation algorithm since it allows the modification of the connecting weights such that it reduces error function E. (Otchere *et al.*, 2021). To obtain a most favorable network model for exploring a given data set, it is prudent to partition the data into subsets, namely, the training set and the test set. Basically, the network learns from the training set, and audits how it goes in hand with the test set and administers itself to a set that has not been checked on (Kwon, 2011).

3.8.4.1: Back Propagation ANN Learning Algorithm

Multilayer feed – forward net with the back – propagation (BP –ANN) learning algorithm is the commonly used ANN models (Zupan, 2008). A simple flowchart followed by the back propagation algorithm involves neurons arranged in a multi – layer network consisting of an input layer (Layer 1), hidden layer (Layer 2), and an output layer (layer 3) as shown in **Figure 3.3**. In fact BP algorithm monitors learning of a network by utilizing the mean square error and the gradient descent to achieve a minimum error sum of squares so us to obtain the modification to the connection weight of the network (Li and Cheng *et al.*, 2012). Expressly, the back propagation learning process involves either the forward propagation or the back propagation of error signal. In forward propagation an input signal is sent from an input layer through a hidden layer to the output layer. For back propagation of error signal, the signal's error is defined as the difference between the actual output and the expected output of the network. The propagation of the error signal is from the output layer to the input layer in a layer by layer manner (Wang and Wu, 2006).
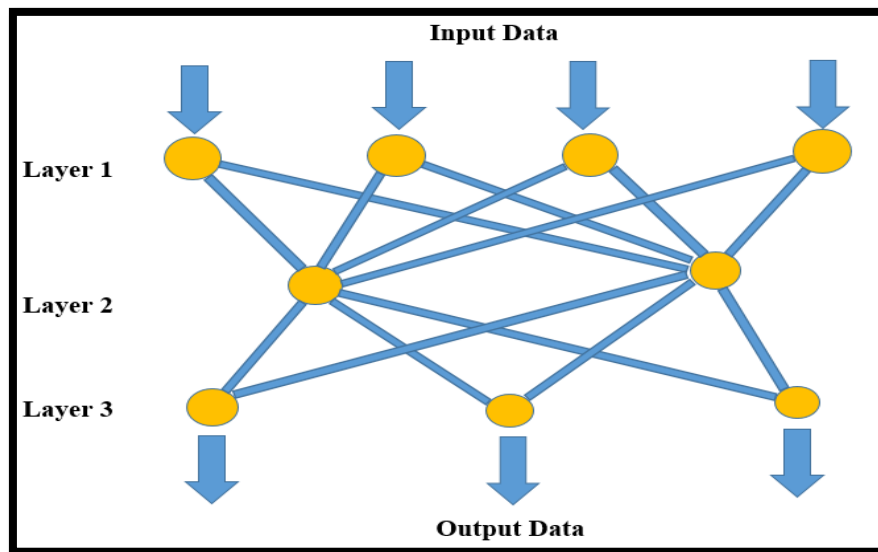
Figure 3.3: A multi – layered neuron network architecture with an input layer (layer 1), hidden layer (layer 2), and an output – layer (layer 3) with the back propagation algorithm (Source: (Kadhm *et al.*, 2021).

The work of a neural network is not only verified graphically but also quantitatively using the error indicators such as; the determination index (R^2) which statistically measures the fit that shows how much variation of a dependent variable is explained by the independent variable(s) in a regression model. The mean absolute error (MAE) is a measure that eliminates the effects of

outliers on the regression model whereas, the mean absolute percentage error (MAPE) measures prediction accuracy of forecasting of the regression model. The mean square error (MSE) measures how close a regression line is to a set of data points, and the root mean square error (RMSE) tells you how concentrated the data is around the line of best fit (Kwon, 2011). These indices measure in many ways, the spread between the original and the estimated output of the network with low values of these indictors, indicating a well optimized neural network model (Chau, 2004).

CHAPTER 4: MATERIALS AND METHODS

4.1: Chapter Overview

This chapter entails sample preparation approaches described in detail as well as how Raman spectral data was acquired from each sample set and preprocessed before analysis. In addition, data analysis approaches utilized in interpreting the spectral data are described.

4.2: Collection of Honey and Molasses Samples

In this study, 100 ml authentic honey sample was obtained from the bee department of ICIPE-Kenya and used without any treatment. The honey sample was directly harvested from a beehive, put into a cylindrical container and stored in the laboratory before experiment (**Figure 4.4 (a)**). In addition, 100 ml of Molasses, a honey adulterant collected in a bottle, was commercially acquired from a local dealer in Nairobi Kenya (**Figure 4.4 (b)**). All the samples were kept at 23 –25 ºC in the laboratory before measurements were taken.

(a)   (b)

Figure 4.4: Authentic honey sample obtained from the bee department of ICIPE-Kenya (a), and Molasses, a honey adulterant (b).

4.3: Raman Sample Substrates Used in the Experiment

Conductive silver paint/paste used for this study is as reported by (Birech *et al*., 2022). The conductive silver paint was smeared on cleaned glass slides to provide a surface over which the samples to be studied were placed before Raman spectral measurements were taken. The Silver paint was used as a Raman signal enhancer to enable acquisition of pronounced Raman signals for easy differentiation of spectral bands in the samples studied.

(a)   (b)

Figure 4.5: Conductive silver paint (a), and three cleaned glass slides (b)

Before applying the Silver paint on a clean glass slide, the bottle was shaken well to ensure all the paint is well liquefied. On the other hand, the glass slides were thoroughly cleaned using distilled water, left to dry before they were rinsed by dipping in a solution of ethanol. The Silver paint was then applied onto the clean glass slides using a smooth brush fitted on its bottle top.

4.4: Sample Preparation

The authentic honey sample acquired from ICIPE and the purchased molasses were each considered to be 100% in concentration. By calculating the concentration of molasses by mass in a mixture of honey and molasses (HMM), 27 samples of honey adulterated by molasses were made by mixing authentic honey sample and molasses at different masses (**Table 4.2**). The masses of the honey and molasses were measured using an electronic balance (**Figure 4.6**). To obtain a desired percentage concentration of molasses in the mixture of honey and molasses samples, the following formula was used:

$$\% \text{ conc. of molasses in HMM} = \frac{\text{Mass of molasses in grams}}{\text{Mass of molasses in grams} + \text{Mass of honey in grams}} \times 100\% \quad (4.4.1)$$

(a)   (b)
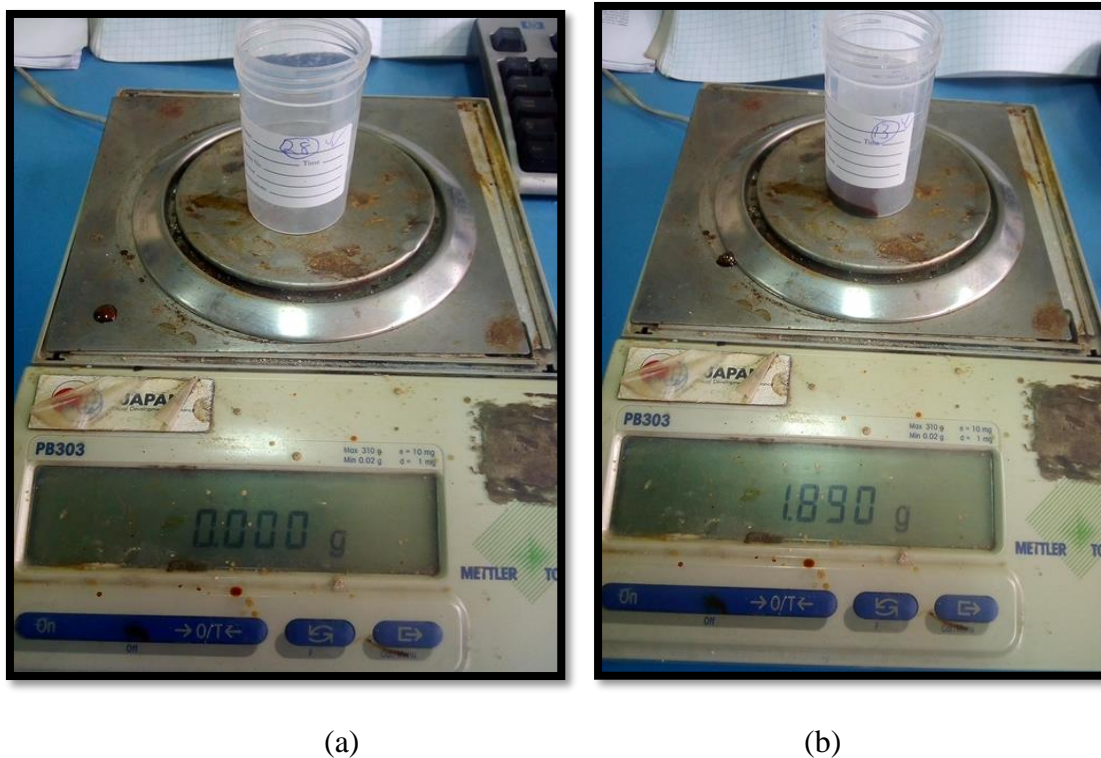
Figure 4.6: Electronic balance used to measure, the weight of an empty sample bottle (a), and the weight of a sample bottle with honey and molasses in it (b)

Table 4.1: Mass of molasses expressed as a percentage of the mixture of honey and molasses

Sample label	Mass of molasses (g)	Mass of honey (g)	Mass of molasses in mixture (%)
Honey	0.000	1.0	0.0
BB	0.005	1.0	0.5
CC	0.010	1.0	1.0
DD	0.015	1.0	1.5
EE	0.020	1.0	2.0
FF	0.026	1.0	2.5
GG	0.031	1.0	3.0
HH	0.036	1.0	3.5
IT	0.042	1.0	4.0
JE	0.047	1.0	4.5
KK	0.053	1.0	5.0
LL	0.058	1.0	5.5
MM	0.064	1.0	6.0
NN	0.070	1.0	6.5
OR	0.075	1.0	7.0
PP	0.081	1.0	7.5
QQ	0.087	1.0	8.0
RR	0.093	1.0	8.5
SS	0.105	1.0	9.5
TT	0.111	1.0	10.0
UU	0.176	1.0	15.0
VV	0.250	1.0	20.0
WW	0.333	1.0	25.0
XX	0.429	1.0	30.0
YY	0.538	1.0	35.0
ZZ	0.667	1.0	40.0
PT	0.819	1.0	45.0
ET	1.000	1.0	50.0
Molasses	1.000	0.0	100.0

The prepared mixtures of honey—molasses (HMM) at different percentage concentrations of molasses were each stirred vigorously using a rod to obtain a homogeneous mixture before storing at the laboratory in the open (**Figure 4.7 (a) and (b)**). The solutions were stirred again using a glass rod before smearing onto a glass slide with conductive silver paint awaiting measurements.

(a)   (b)

Figure 4.7: Prepared mixtures of honey – molasses (HMM) samples at different percentage concentrations of molasses (a), the prepared mixtures stored in the open on a laboratory bench (b).

Three spots of silver paint smears were made on each glass slide and left for 7 minutes to dry (**Figure 4.8 (a)**). A drop < 1 gram of HMM at each percentage concentration was smeared over the three spots of respective labeled glass slides and spread to form a uniform thin layer before measurements were done (**Figure 4.8 (b)**). Thirty Raman spectra were acquired from each percentage concentration of the HMMs. (what was thickness and uniformity? How does it affect results? thus also the 1 drop of HMM) – best answer is, single layer smear, of silver paste, and single layer smear of HMM for consistent results and many spectra were taken to take off any biasness of results and big variation between spectra – this was done to make sure at least the data collected is consistent from sample to sample – at each % concentration 30 spectra were collected.
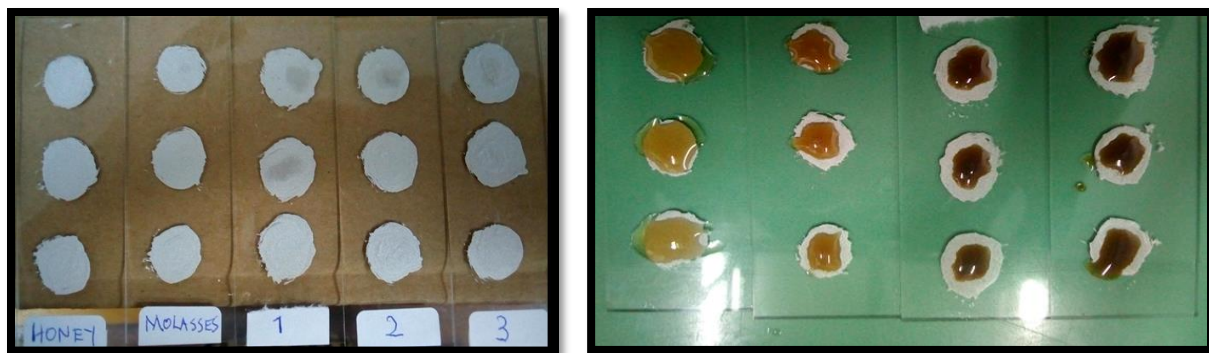
<center>(a)   (b)</center>

Figure 4.8: Glass slides with smears of silver paint left to dry in the open (a), honey—molasses smears (HMM) over dry silver paint (b)

4.5: Instrumental Optimization and Sample Analysis

STR Raman spectrometer equipped with a 785 nm Near Infrared diode laser, a thermoelectric cooled CCD detector with 2048 pixels, a manually controlled microscope stage was used (**Figure 4.9**). Instrumental parameters were set as follows: density filter was set to allow 50% of laser power to pass to the sample (with a maximum power output of 18.20 milliwatts and a spot size of 68.47 micrometers), X10 (0.30) objective lens was used to focus the sample under study at the microscope stage, acquisition time was 10 seconds, 5 accumulations per sample spot and spectra was acquired in the region of 97 to 1846 cm^{-1} at a resolution of 1.78. The 600 BLZ grating was used. Before starting the experiment, the Raman system was first cooled to -76 ^{o}C and calibrated using a silicon wafer peaked at 520.5 cm^{-1} and centered at 1050 cm^{-1}. All experimental measurements were done at room temperature and a dark room.

<center>37</center>
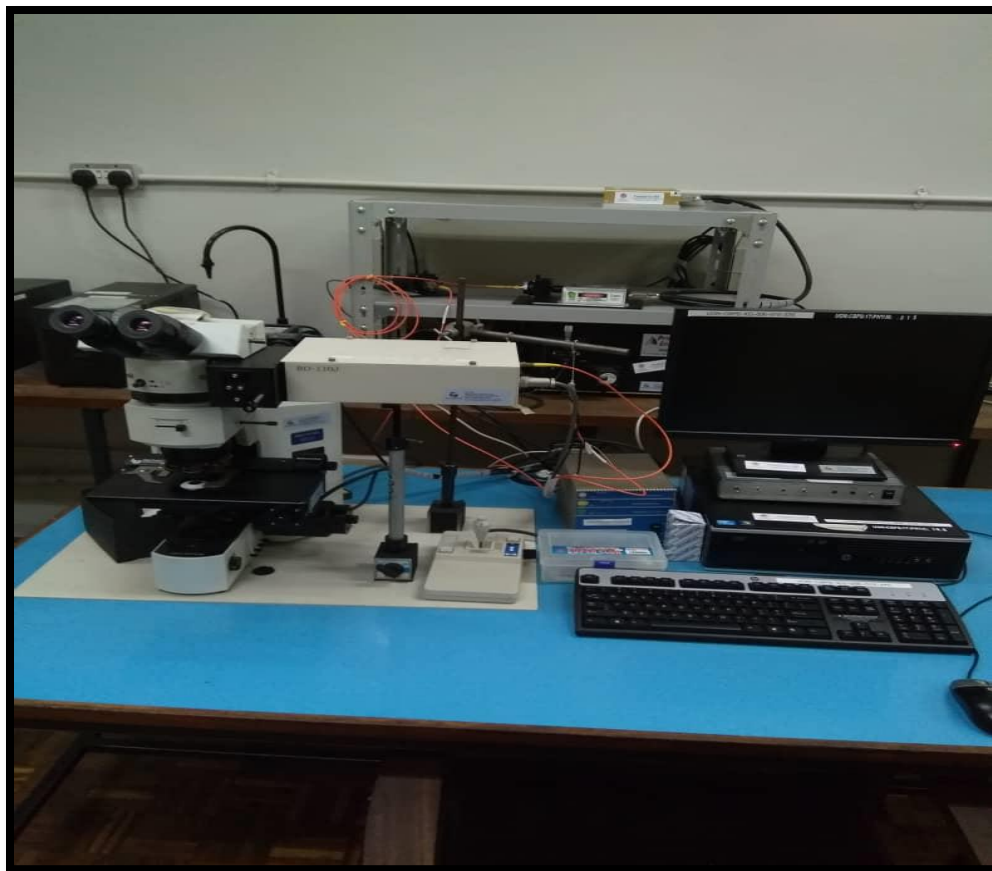
Figure 4.9: Laser Raman Spectrometer equipment in the department of Physics at the University of Nairobi.

4.6: Raman Spectral Data Acquisition and Analysis

A drop of 1 gram of pure molasses and pure honey, as well as honey — molasses mixtures were each placed on a glass slide smeared with a thin layer of conductive silver paste and placed on the microscope stage of the STR Raman spectrometer. Once on the stage, each sample was first focused to obtain a clear view before the spectra were captured and stored using the STR software fitted in the computer. The STR Raman spectrometer captured a single spectrum under 10 seconds. A total of thirty spectra was captured from each sample. The collected Raman spectra was preprocessed using Vancouver Raman algorithm in order to remove fluorescence and background signals while R software was utilized for spectral normalization.

4.6.1: Analysis of Collected Raman Spectra

In this study, PCA was executed using Chemo Spec package in R software to eliminate repetitious data and to extract useful information about the collected spectra. Robust PCA method was used with the scaling options set at no scaling, median absolute deviation and the activation function r_pcaSpectra were chosen for analysis. Two data sets were prepared for the analysis. Set A comprised of authentic honey and molasses and Set B comprised of honey adulterated by molasses at different adulteration levels.

On the other hand, RF and SVM were comparatively used for classification analysis of authentic honey and adulterated honey samples. Five data sets, Set. 1, Set. 2, Set .3, Set. 4, and Set. 5 discussed in the results section were developed and used for classification analysis by RF and SVM. Principal Components were used as inputs for the RF and SVM classification models. For the RF model, the training phase comprised 70 % of the data set, while the test set phase comprised 30 % of the data set. Using the first 10 PCs, both models, RF and SVM were found to yield high classification accuracies, thus utilizing in the building of the models. The RF model was constructed using 15 trees with the mtry value set at 4 for all the data sets. Moreover, the SVM model was constructed using the linear kernel function with a cost of 5 and gamma of 1. For SVM, the hyperplane with highest correct classification accuracies was obtained when the data partition was 60 % for training set and 40 % for testing set. The confusion matrix for all the classification data sets involving RF and SVM are as presented in **Tables 5.4 and Table 5.5** displayed in results section. To know the machine learning technique best suited for classification analysis of honey adulterated by molasses, the classification accuracies of RF and SVM were simultaneously compared, thus validating the results of each model. As such, RF could be preferred to SVM for classification analysis involving honey adulterated by molasses because it yielded better results than SVM as shown in **Table 5.4 and Table 5.5**

ANN regression model was built for the purpose of predicting the various concentrations of molasses as an adulterant. In this study, PC's were used as inputs for the neural net since this led to efficient reduction of the net architecture thus a rapid training phase (Wu and Massart, 1996). In order to build the ANN model, the data was first randomly partitioned into two sets, 70 % for the model training and 30 % for model testing. Model optimization was done by training the model using different number of PC's. It was realized that, using the first 5 PC's, the training phase took

the least time with better R^2 values being realized. Using the tune Grid package in R, the first hidden layer of the ANN model was randomized from 2 to 8 hidden layers with the second and the third set at 16 and 3 respectively. Through the training phase, the model was able to determine the best neural net architecture by selecting the architecture that yielded the lowest RMSEP value. In addition, the model was built using the rprop + algorithm, with the threshold set at 0.1 and the stepmax at e^{+05}. Moreover, the ANN model training set values were utilized in calculating the limit of detection using Origin software.

CHAPTER 5: RESULTS AND DISCUSSIONS

5.1: Chapter Overview

Analysis of huge amount of spectral data to simultaneously determinate the specific molecular fingerprints in complex media such as honey and molasses is very difficult. Hence, the application of machine learning techniques to eliminate repetitious data and extract essential features from a data set become necessary. Before analyzing the data using multivariate techniques, the spectral window of 300 to 1800 cm^{-1} was chosen for analysis (Yaacob, *et al.*, 2019). This region was found to give relevant information for the interpretation of the spectral data. In this work, PCA was used to perform exploratory analysis. Using PCA, glucose, fructose, sucrose, protein and amide spectral bands were established in the spectral profile of authentic honey, molasses, and molasses - adulterated honey samples thus enabling their spectral fingerprint distinction. In addition, RF and SVM were comparatively used in classification analysis of authentic and molasses - adulterated honey samples. High classification accuracies ranging from 86 – 100 % were obtained from the various samples sets that were analysed. ANN was used to construct a prediction model that could be employed for rapid forecasting of molasses – adulteration in honey at various levels. In addition, the ANN values were used in the calculation of limit of detection.

5.2: Characteristic Raman Spectra of Authentic Honey and Molasses

Due to the close molecular composition of both molasses and honey, it is expected that their Raman profiles do not significantly differ. In order to identify differing bands between the two, ANOVA was used. **Figure 5.10**, displays the significantly variant Raman bands of authentic honey and molasses and were found to be within 670 cm^{-1} to 1050 cm^{-1} spectral range. The bands with the highest variance were recorded as bands that could be used for distinction of honey and molasses. As such, the prominent unique peaks with the highest significant variance were noted at 690**,** 732, 754, 790, 793, 845, 880, 970, 1001, and 1645 cm^{-1}.
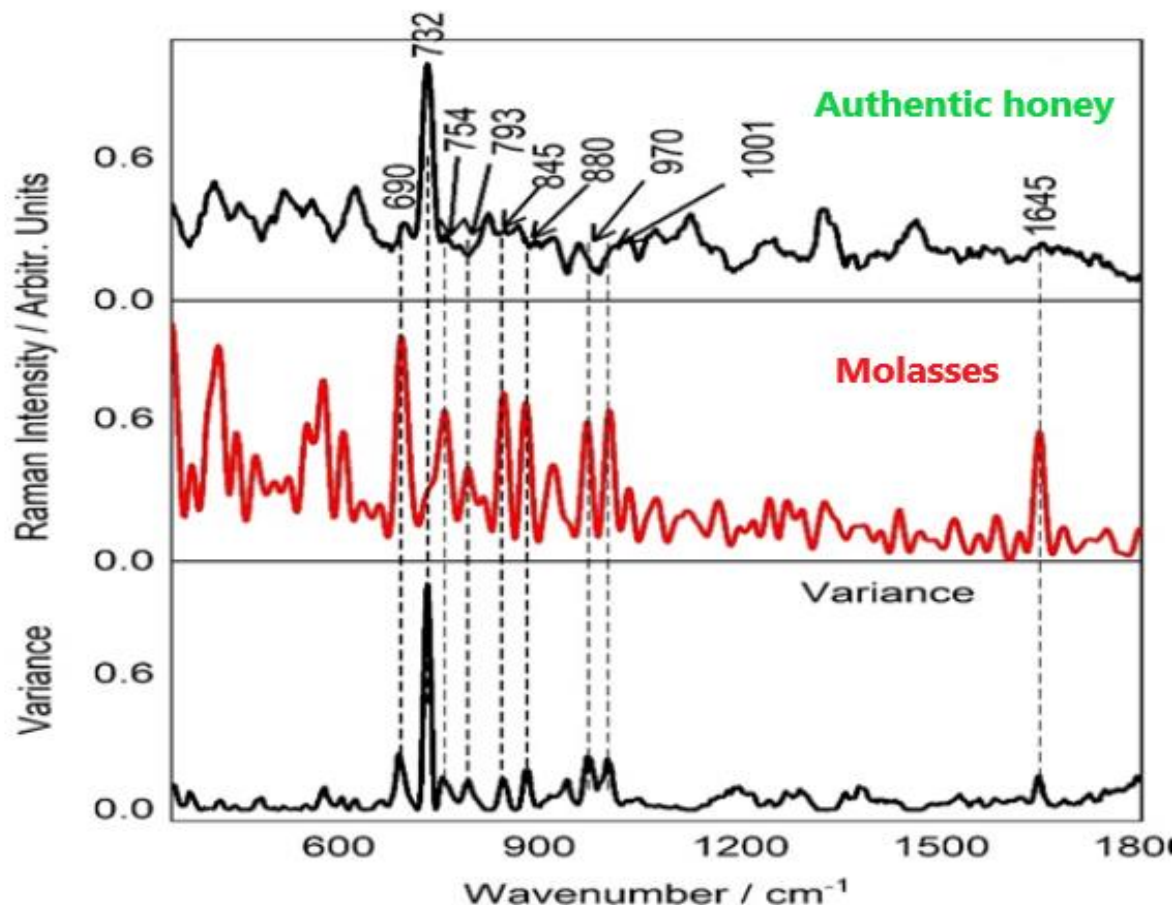
Figure 5.10: Displaying Characteristic Raman spectral profiles of authentic and molasses (average of – each) in the range of 300 – 1800 cm^{-1}. The variance plot is also plotted with significantly variant bands indicated with dotted lines as an eye guide.

As shown in **Figure 5.10**, the unique band at 732 cm^{-1} was noted in authentic honey spectra and the bands at 690, 754, 793, 845,880, 1001, and 1645 cm^{-1} were noted in molasses spectral profile. The prominent band at 732 cm^{-1} was tentatively associated with glucose ν(C-C) vibrations and the band at 790 cm^{-1} was linked to ν(C-C) and δ (C – H) vibrations in α-glucose. Also, the ν(C-C) and ν(C-O) stretching vibrations on glucose were found to have a spectral range of 1000 to 1200 cm^{-1} thus responsible for the band at 1001 cm^{-1} (Özbalci *et al.*, 2013). The band at 1645 cm^{-1} was attributed to Amides –C=O stretching which dominate the signal window of 1640 cm^{-1} to 1670 cm^{-1} (Ghosh and Jayas, 2009). The 845 cm^{-1} band was discovered to correspond to glucose spectrum while the bands at 793 cm^{-1} and 970 cm^{-1} were attributed to glucose spectrum and ν(C-

42

O) vibrations in glucose (Oroian *et al.*, 2018; Özbalci *et al.*, 2013). Stretching of CO and CCO, OCO bending were found responsible for the band at 690 cm[-1] (Oroian *et al.*, 2018). Weak (C = O) bond vibrations and wagging (NH -O) bond vibrations are responsible for the bands at 754 cm[-1] and 880 cm[-1], respectively (Wurzburg, 2006).

Conclusively, 690**,** 732, 754, 793, 845, 880, 1001, and 1645 cm[-1] may be used as Raman marker bands in the distinction of authentic honey and molasses.
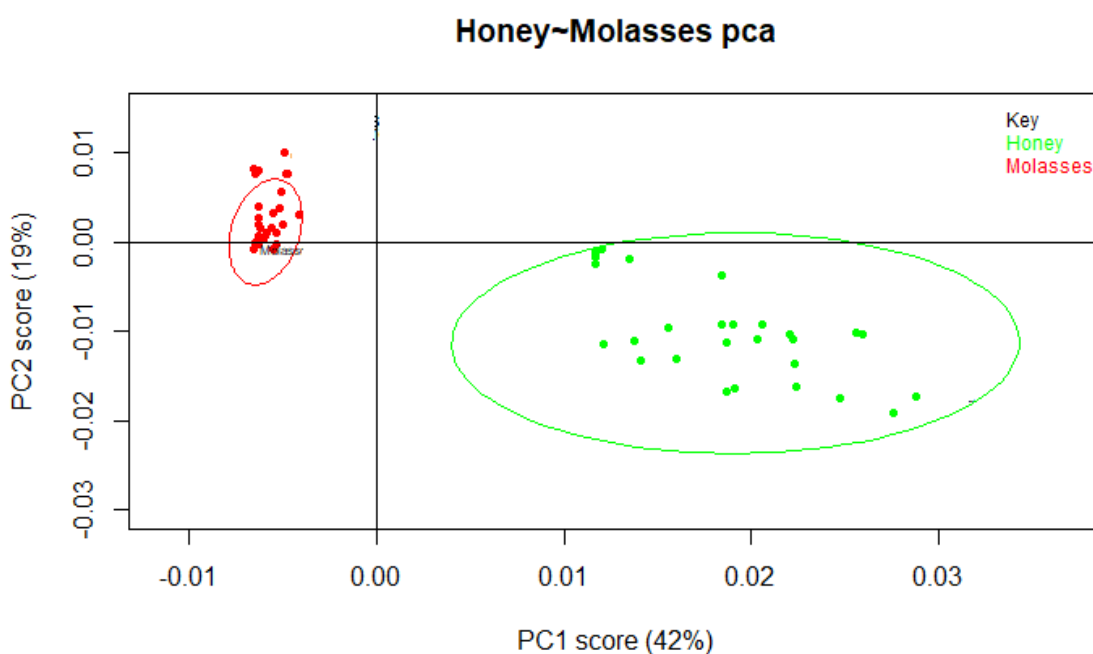
Figure 5.11(a): PCA score plot of authentic honey and molasses samples shows close clustering of honey samples and molasses samples, indicating that each sample has the its own unique characteristics.
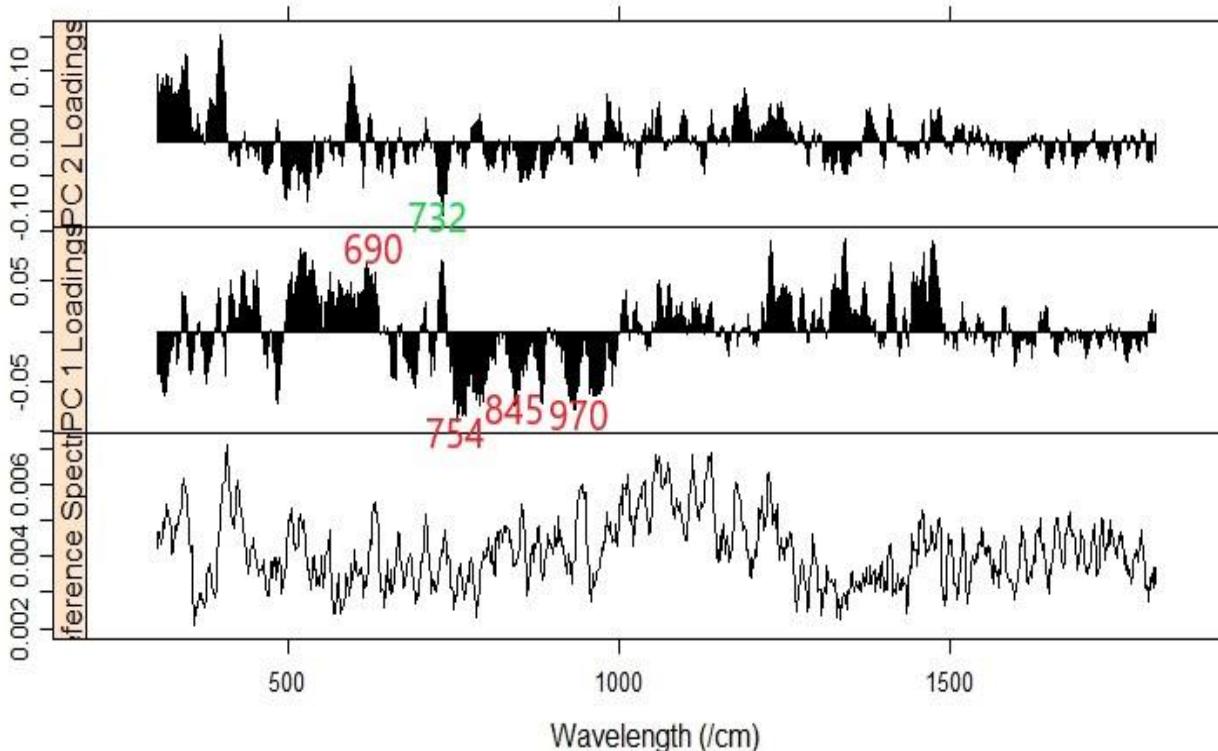
Figure 5.12 (b): Plot of PCA loadings for authentic honey and molasses samples. Unique bands responsible for honey and molasses' segregation have been identified, these bands are the ones with the highest loading value.

To further elucidate the significantly variant Raman spectral bands between authentic honey and molasses, PCA was done. As displayed in **Figure 5.11 (a) and 5.12 (b)**, for PCA score plot of the combined Raman spectral data sets of the two samples, distinct clusters of the two sets was observed between positive PC1 and negative PC1.However negative PC2 showed some clusters were similar between authentic honey and molasses. The segregation displayed in positive and negative PC1 indicated that the spectral patterns of the two compounds (i.e. authentic honey and molasses) were distinctively different **(Figure 5.11 (a)). Figure 5.12 (b)** thus were used to confirm the bands liable for the distinct clustering. These bands were found centered at 690, 732, 754, 845, and 970 cm^{-1} thus confirming the spectral bands identified using ANOVA. The component and vibrational assignments of these bands are as shown in **Table 5.3**. The prominent bands that influenced the segregation were those with large loadings value and are as labeled in **Figure 5.12 (b).**

44

Table 5.3: Component and Vibrational assignments of Raman bands/peaks identified in authentic honey and molasses using ANOVA and PCA

Raman band/peak (cm^{-1})	Component assignment	Vibrational assignment	Authentic honey	Molasses	References
690	fructose	Stretching of CO and CCO, OCO bending	X	✓	(Oroian *et al*., 2018)
754	fructose	C – OH	X	✓	
732	glucose	ν(C-C)	✓	X	(Özbalci *et al*., 2013) (Gelder *et al*., 2007)
845 and 880	fructose and glucose,	ν(C–O), δ(C–C–H), ν(C–C) and δ(C–C–O)	X	✓	
1001	glucose	ν(C-C) and ν(C-O) stretching vibrations	X	✓	
793 and 970	glucose	ν(C- O)	X	✓	(Oroian *et al*., 2018; Özbalci *et al*., 2013)
1645	Amides	–C=O stretching	X	✓	(Ghosh and Jayas., 2009) (Oliveira and Bara, 2002)

5.3: Identification of Raman Marker Bands for Molasses Adulteration in Honey
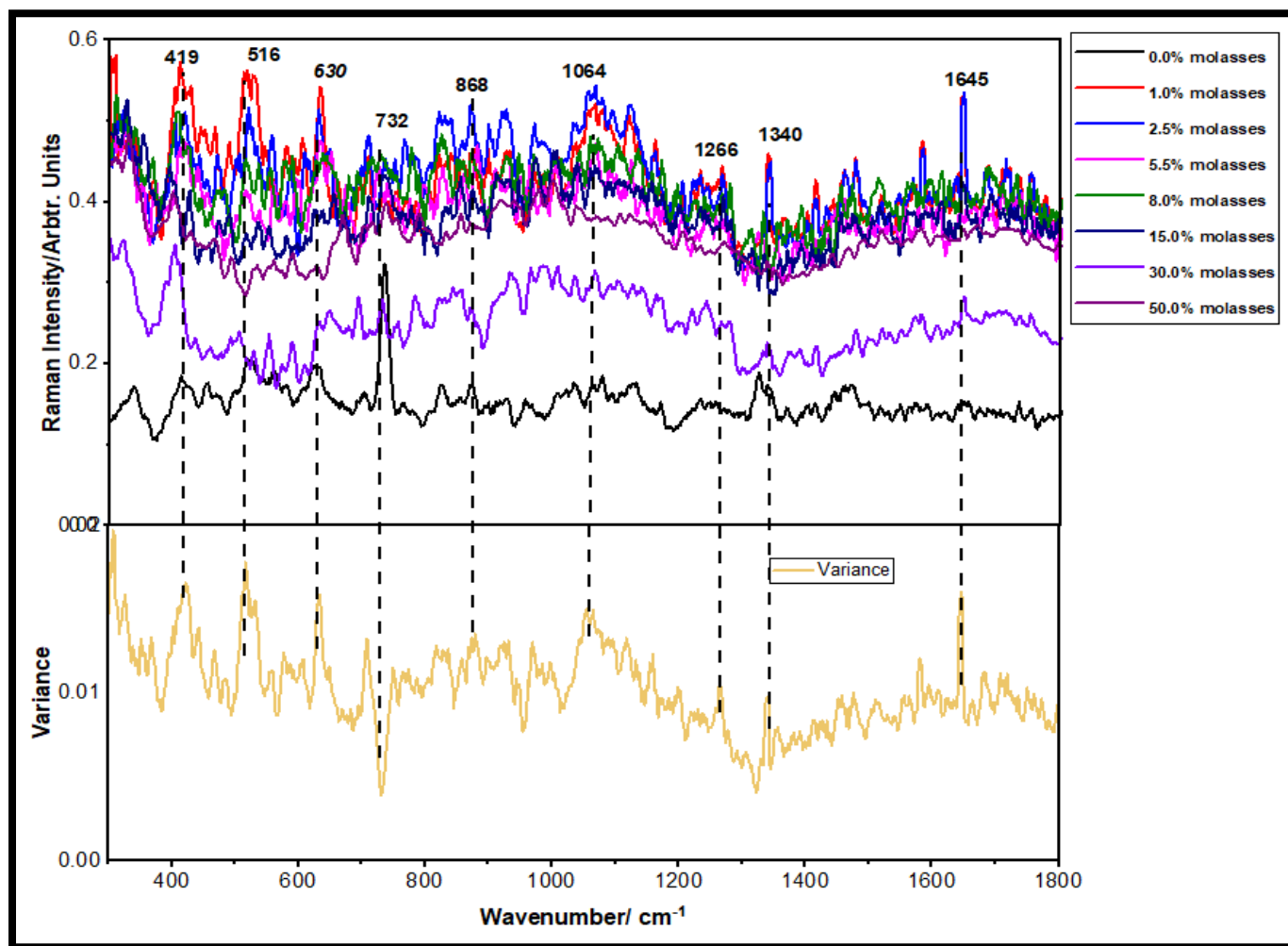
Figure 5.13: ANOVA on averaged Raman spectra of authentic honey and honey adulterated by molasses at various concentrations. The variance plot has been plotted showing significantly variant bands indicated with dotted lines as an eye guide.

Using ANOVA as shown in **Figure 5.13**, the significantly distinct and variant Raman spectral bands at various concentrations were found to be centered at 419, 516, 630, 732, 868, 1064, 1266, 1340, 1582 and 1645 cm^{-1}. The bands at 419 and 630 cm^{-1} were linked to fructose δ (C – C – O) ring vibration in the pyranoid ring and to ring deformation respectively (Salvador *et al.*, 2019). The band at 516 was linked to exocyclic deformation band in β – glucose that occurs within the signal range of 477 to 542 cm^{-1}. The prominent band at 732 cm^{-1} was tentatively attributed to glucose ν(C-C) vibrations (Özbalci *et al.*, 2013). The band at 868 cm^{-1} was linked to C – O – C

46

cyclic alkyl ethers in fructose bending of fructose as well us it may have arose due to υ(COH), υ(CCH) and υ(OCH) side group deformations in β – glucose that dominate the signal window of 800 cm^{-1} to 950 cm^{-1} (Miljanić, Frkanec and Biljan, 2007). (Goodacre *et al.*, 2002). The ν(C-C) and ν(C-O) stretching vibrations on glucose were found to have a signal range of 1000 - 1200 cm^{-1} thus responsible for the band at 1064 cm^{-1} (Özbalci *et al.*, 2013). The band at 1266 cm^{-1} could be limked to D (–) fructose δ (=CH) (Gelder *et al.*, 2007). In the signal ranges of 1335 cm^{-1} to 1435 cm^{-1}, the band at 1340 cm^{-1} was found to be connected to carbohydrates molecules that are vibrationally dominated by C – H and O – H bonds (Li and Shan *et al.*, 2012a) (Oliveira and Bara, 2002). Amides –C=O stretching, which dominate the spectral range of 1640 cm^{-1} to 1670 cm^{-1}, maybe responsible for the band at 1645 cm^{-1} (Ghosh and Jayas, 2009) (Oliveira and Bara, 2002).

5.4: Results of Classification Analysis Using random forest (RF) and support vector machine (SVM)

Adulteration levels were determined as low, medium and high based on the concentration ranges of the adulterated simulates prepared as shown in **Table 4.2.** Concentrations up to 10% were considered low adulteration levels, whereas concentrations between 15 – 35% were considered to be medium concentrations and concentration between 40 – 50% were considered to be high concentrations.

Table 5.4: **Classification Models of Set.1, Set.2, and Set.3 using RF and SVM**

	Random Forest Test Set				Support Vector Machine Test Set			
Set. 1	1	2	Total	% correct	1	2	Total	% correct
1	9	0	9	100.00	10	2	12	83.33
2	1	8	9	88.89	0	6	6	100.0
Total	10	8	18	**94.44**	10	8	18	**88.89**
Set. 2	1	2	Total	% correct	1	2	Total	% correct
1	10	0	10	100.00	10	0	10	100.0
2	0	8	8	100.00	0	8	8	100.0
Total	10	8	18	**100.00**	10	8	18	**100.00**
Set. 3	1	2	Total	% correct	1	2	Total	% correct
1	10	0	10	100.00	10	0	10	100.0
2	0	8	8	100.00	0	8	8	100.0
Total	10	8	18	**100.00**	10	8	18	**100.00**

Set 1: 1 – honey, 2 – honey with 0.5 % molasses, **Set 2:** 1 – honey, 2 – honey with 25 % molasses,

Set 3: 1 – honey, 2 – honey with 50 % molasses

Table 5.5: **Classification models of Set. 4 and Set. 5 using RF and SVM**

	Random Forest Test Set						**Support Vector Machine Test Set**					
Set. 4	1	2	3	4	Total	% correct	1	2	3	4	Total	% correct
1	10	0	1	0	11	90.91	10	0	0	0	10	100.0
2	0	8	0	0	8	100.0	0	8	1	0	9	88.89
3	0	0	9	0	9	100.0	0	0	9	0	0	100.0
4	0	0	0	8	8	100.0	0	0	0	8	8	100.0
Total	10	8	10	8	36	**97.22**	10	8	10	8	36	**97.22**
Set. 5	1	2	3	4	Total	% correct	1	2	3	4	Total	% correct
1	9	1	0	0	10	90.00	9	2	0	0	11	75.00
2	1	162	18	1	182	89.01	1	161	15	1	178	82.42
3	0	3	29	5	37	78.38	0	3	31	7	41	78.95
4	0	0	4	15	19	78.95	0	0	5	13	18	68.42
Total	10	166	51	21	248	**86.69**	10	166	51	21	248	**86.29**

Set 4: 1 – honey, 2 – honey with 0.5 % molasses, 3 – honey with 25 % molasses and 4 – honey with 50 % molasses.

Set 5: 1 – honey, 2 – low adulteration concentrations (0.5% – 10%), 3 – medium adulteration concentrations (15% – 35%), 4-high adulteration concentrations (40% – 50%)

The effective application of a random forest model for classification analysis is anchored on the choice of two main parameters; the mtry and the ntree. A random forest model is said to have good performance if the OOB (out of bag) error is minimized (Breiman., 2001). In this study, an optimized random forest model was established with mtry at set 4 and ntree set at 15 for all the five data sets developed for classification analysis. High correct classification accuracies ranging from 86 – 100 % were realized when using the random forest method as shown in **Table 5.4** and **Table 5.5.**

Other variables that had greater influence on the accuracies of RF model were as shown in **Table 5.6** below.

Table 5.6: Details of variables from the RF Classification Analysis

	OOB Error (%)	**p – value**	**kappa**
Set 1	14.29	0.0003914	0.8889
Set 2	0	2.542 e^{-05}	1
Set 3	0	2.542 e^{-05}	1
Set 4	8.33	2.2 e^{-16}	0.9628
Set 5	10.15	8.38 e^{-13}	0.7169

In **Table 5.6,** Set 1 is noted to have the highest OOB error of 14.29 % than the other sets. This set comprised of authentic honey and honey at 0.5 % adulteration level. With this, the two sets had close similarities, thus the high OOB error and classification accuracies lower than 100 % as reported in **Table 5.4**. Furthermore, the OOB error of Set 2 and Set 3 were zero owing to the fact that these sets comprised of authentic honey with a 25 % adulterated sample for Set 2, and authentic honey with a 50 % adulterated sample for Set 3. With this high variation in characteristics due to adulteration, the random forest model correctly classified authentic honey from adulterated honey samples, thus the zero OOB error and high classification accuracies of 100 % as displayed in **Table 5.4**.

Notably, the main principal of operation of an SVM classification model is in creating a hyperplane that maximizes on the margin from the closest support vectors separating any two classes. In addition, the effectiveness of an SVM classier lies in the choice of the kernel type, the gamma (γ) and the cost (C). The cost regulates the degree of violating the margin (Y. Wang *et al.*, 2006; Nyairo, 2018). A low C values yields wide margins that can hold more support vectors and vice versa (Lin and Chen., 2013). For radial basis kernels (RBK) functions, γ is utilized as a measure of similarity between two given points. In this study, the SVM model was able to yield a hyperplane when the parameters were optimized at a cost = 5 and $\gamma = 1$.
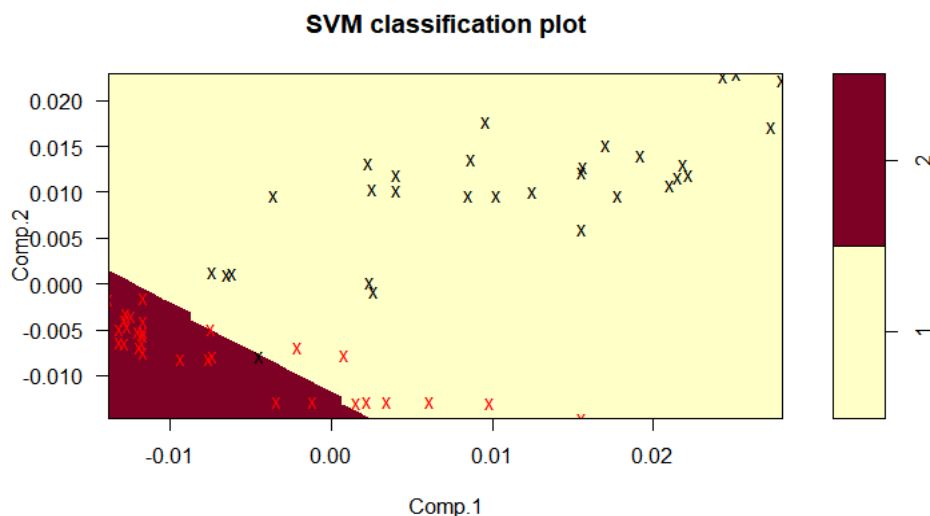
Figure 5.14: SVM plot utilizing linear kernel function

5.5: Results of Prediction Model Using ANN

The ANN model was constructed using the tune Grid package in R. The first hidden layers of the model were set from 2 to 8 allowing the training phase of the model to automatically have several architectures. The model was able to select the best neural architecture based on the layers that yielded the lowest RMSEP value as shown in **Table 5.7**. In the training phase, the best neural architecture with the lowest RMSEP value of 1.86 was realized when layer 1 had 7 inputs. Using this architecture, the test phase yielded a RMSEP value of 2.05 with R^2 value of 0.5786 and MAE of 1.51.

Table 5.7: Training Phase of ANN Model

Layer 1	RMSE	R – Squared	MAE
2	NaN	NaN	NaN
3	NaN	NaN	NaN
4	NaN	NaN	NaN
5	1.86	0.65	1.41
6	1.97	0.61	1.44
7	**1.86**	**0.65**	**1.29**
8	2.26	0.55	1.62

5.5.1: Calculation of Limit of Detection (LOD):

In analytical techniques, the limit of detection is the least amount of an analyte that can be traced and established to be present in a particular sample (Vogt, 2015). According to IUPAC, the statistical significance of a spectrum is when the signal is raised by 3 standard deviation above the background reference (Chang, 2011). Thus

$$LOD = \frac{3\sigma}{S}$$
(5.11)

σ is the standard deviation of the multiple samples and S is the sensitivity.
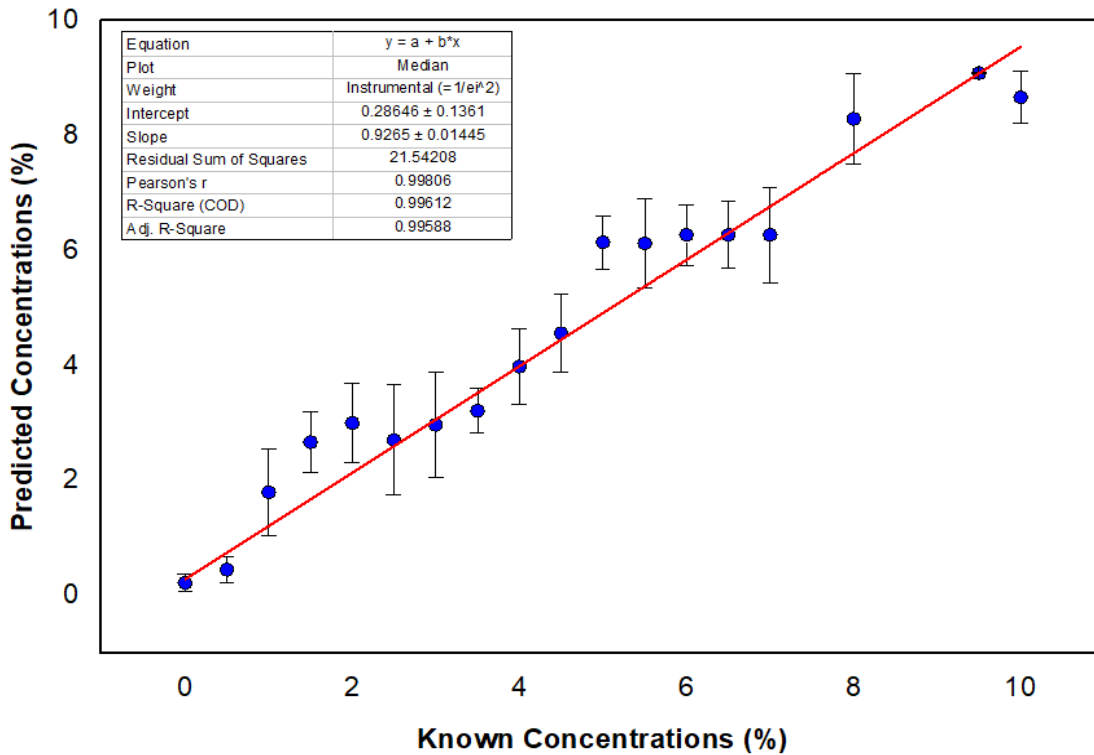
Figure 5.15 (a): Linear fit plot using the median and mean absolute deviation from the training data set of the ANN model.

ANN model was executed using tune Grid package in R software. Due to wide range in consecutive adulteration levels above 10%, the concentration ranges from 0 – 10% were used for the calculation of LOD. However, when testing the ANN model, it was realized that a linear fit using the mean and standard deviation yielded high errors bars. Therefore, linear fit plots shown in **Figure 5.15 (a) and Figure 5.16 (b)** were obtained by plotting the median values of known concentration against predicted concentrations. The median was used because it has been reported to be statistically more stable than the mean (Seo, 2006). The error bars were derived from the mean absolute deviation (MAD) calculated using the predicted values at different concentrations using the formula:

$$MAD = MEDIAN(ABS(R - MEDIAN(R)))$$ (5.12)
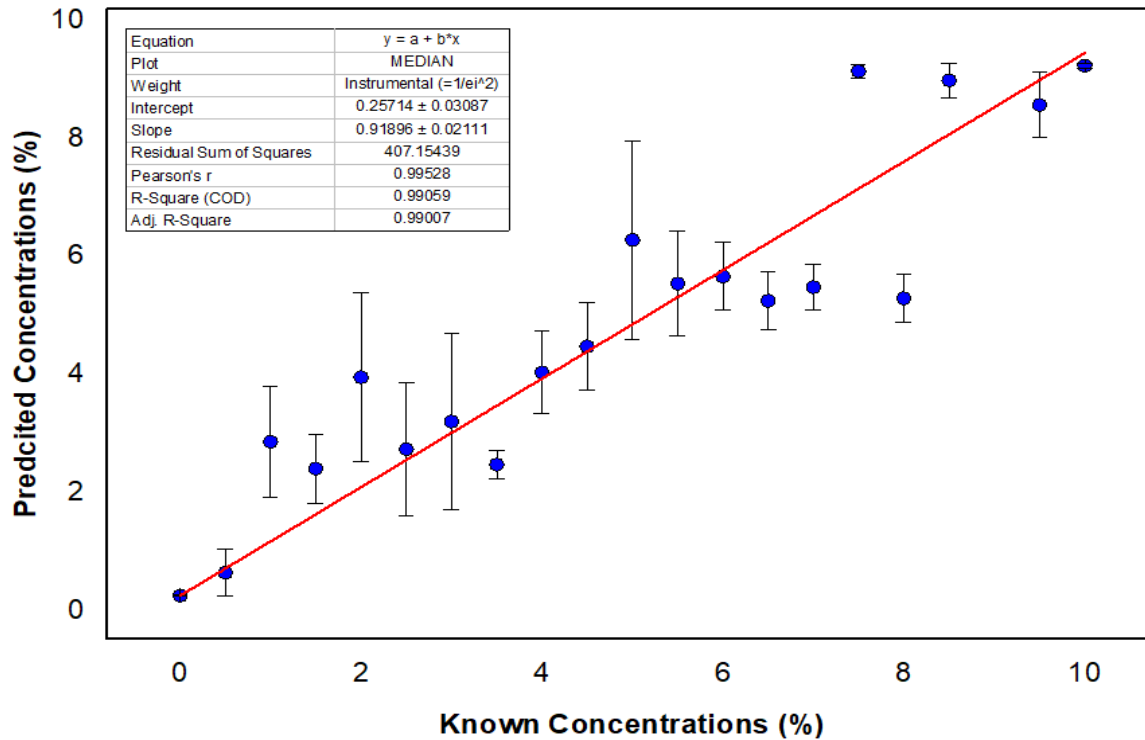
Figure 5.16 (b): Linear fit plot using the median and mean absolute deviation from the test data set of the ANN model.

Utilizing **equation 5.11**, the LOD of molasses in honey in reference to the collected Raman spectral data of authentic honey, molasses and molasses - adulterated honey was found to be:

$$LOD = \frac{3\sigma}{S} = \frac{3 \times 0.1361}{0.9265} = \frac{0.4083}{0.9265}$$

$$LOD = 0.44069\%$$

An LOD of 0.44069 % shows that laser Raman spectroscopy is able to accurately detect 4 grams of molasses present in 1000 grams of authentic honey, thus proving the viability of laser Raman as a rapid and robust technique that could be used to detect subtle adulteration levels.

CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS

6.1: Conclusions

The objective of this study was to discover definitive molecular fingerprints that distinguish authentic honey from honey adulterated with molasses by detecting the vibrational bands of molecular compounds found in both honey and molasses, such as fructose, glucose, and sucrose. The analysis of authentic honey and molasses molecular profile using ANOVA and principal components produced definitive and distinct spectral bands which characterize authentic honey, molasses, and honey under different adulterations levels of molasses. Using ANOVA, the band at 732 cm^{-1} was found to be unique to honey profile whereas the bands at 690, 754, 793, 845,880, 1001, and 1645 cm^{-1} were noted in molasses spectral profile. The aforementioned bands may be considered as Raman marker bands for detecting molasses adulteration in honey.

To check on the segregation ability of Raman spectroscopy to identify definitive spectral bands of specific molecular compounds was explored using Random Forest and Support Vector Machines as classification models. The use of principal components as inputs in both models helped in the classification models' rapid analysis as well as their high correct accuracies. Both RF and SVM models achieved high correct classification accuracies ranging from 86 to 100 percent, confirming the robustness and potential of Chemometric –Integrated Raman spectroscopy as a viable method for conclusive molecular analysis.

Principle components were used as inputs in the construction of ANN prediction model. The use of principal components as inputs significantly helped the model to have a less complicated architecture, thus improving on its learning rate and accuracy of prediction. Developed with an aim of predicting low levels of adulteration, low concentrations ranging from $0 – 10$ % were considered. The model yielded R^2 value of 0. 5786 with RMSEP of 2.05 I and MAE of 1.51. Although the R^2 value of the prediction model was at 0.5786, it clearly illustrates the possibility of better models being built whenever negligible adulteration ranges between consecutive samples are put into consideration. In addition, the utilization of the ANN model values, yielded a limit of detection lower than 1 %.

6.2: Recommendations

The market demand of honey and honey products is ever growing, so are the methods of developing sophisticated honey adulterants. The methodology developed in this work using molasses as one of the major honey adulterants can be extended to include honey from different geographical regions as well as more of the cheaply available and commonly used honey adulterants. In so doing, comprehensive, spectral libraries of distinct spectral bands characterizing, the honeys and the various adulterants can be developed and act as a reference for spectral analysis of honey in our country.

Notably, the main setback experienced in this work involved colleting good quality Raman spectra that is reproducible on repeat measurements as well as a signal substrate that has negligible influence on the acquired Raman signal. The author proposes a methodology where obtaining reproducible Raman spectra from a spot of a sample is possible, this can be achieved by using automated laser systems with the ability to fully scan all the sample at once to obtain multiple spectra and at the same time maintaining the amount of power delivered to the sample under study. Although background subtraction is recommended whenever substrates are used for signal enhancements, it is however prudent to consider the use of substrates that have minimal contribution to the Raman signal of a sample especially when dealing with samples that are partly opaque and partly translucent.

In a nutshell, although the robustness of Chemometric – integrated spectroscopic techniques has been remarkable in detecting honey adulterants (Se *et al.*, 2019b), there is need to explore the effectiveness of biosensor techniques such as e –nose and optical fibre sensors in the detection of the different types of adulterants in honey. Comparative to spectroscopic techniques, biosensors are also simple, rapid and accurate to use (Ghasemi-Varnamkhasti *et al.*, 2018). Similarly, biosensors coupled with Chemometrics data analysis has been reported to be capable of real time adulteration analysis on large scale (Bougrini *et al.*, 2016).

REFERENCES

Abdelghany, S.F. (2016), An Introduction to Mammals AN INTRODUCTION TO, **Vol. 5 No. April**, pp. 1–19.

Aleboyeh, A., Kasiri, M.B., Olya, M.E. and Aleboyeh, H. (2008), Prediction of azo dye decolorization by UV/H2O2 using artificial neural networks, *Dye. Pigment.*, **Vol. 77 No. 2**, pp. 288–294.

Alula, M.T., Mengesha, Z.T. and Mwenesongole, E. (2018), Advances in surface-enhanced Raman spectroscopy for analysis of pharmaceuticals: A review, *Vib. Spectrosc.*, Elsevier, **Vol. 98 No. March**, pp. 50–63.

Anghelone, M., Jembrih-Simbürger, D., & Schreiner, M. (2015). Identification of copper phthalocyanine blue polymorphs in unaged and aged paint systems by means of micro-Raman spectroscopy and Random Forest. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **Vol *149***, 419-425.

Anguebes, F., Pat, L., Ali, B., Guerrero, A., Córdova, A. V., Abatal, M., & Garduza, J. P. (2016). Application of multivariable analysis and FTIR-ATR spectroscopy to the prediction of properties in campeche honey. *Journal of Analytical Methods in Chemistry*, **Vol.*2016 pp. 20 - 31***

Araujo, C.F., Nolasco, M.M., Ribeiro, A.M.P. and Ribeiro-Claro, P.J.A. (2018), Identification of microplastics using Raman spectroscopy: Latest developments and future prospects, *Water Res.*, Elsevier Ltd, **Vol. 142**, pp. 426–440.

Ayari, F., Mirzaee-Ghaleh, E., Rabbani, H., & Heidarbeigi, K. (2018). Using an E-nose machine for detection the adulteration of margarine in cow ghee. *Journal of Food Process Engineering*, **Vol. *41*(6)**, e12806.

Bachion, F., Santana, D., Borges, W. and Poppi, R.J. (2019), Random forest as one-class classi fi er and infrared spectroscopy for food adulteration detection, *Food Chem.*, Elsevier, **Vol. 293 No. July 2018**, pp. 323–332.

Baeten, V., Meurens, M., Morales, M. T., & Aparicio, R. (1996). Detection of virgin olive oil adulteration by Fourier transform Raman spectroscopy. *Journal of Agricultural and Food Chemistry*, **Vol.*44***, pp.2225-2230.

Birech, Z., Mwangi, P.W., Sehmi, P.K. and Nyaga, N.M. (2020), Application of Raman spectroscopy in comparative study of antiobesity influence of oxytocin and freeze-dried extracts of Uvariodendron anisatum Verdeck (Annonaceae) in Sprague Dawley rats, *J. Raman Spectrosc.*, **Vol. 51 No. 3**, pp. 398–405.

Blum, C., Opilik, L., Atkin, J. M., Braun, K., Kämmer, S. B., Kravtsov, V., ... & Zenobi, R. (2014). Tip-enhanced Raman spectroscopy–an interlaboratory reproducibility and comparison study. *Journal of Raman Spectroscopy*, **Vol. *45***, pp.22-31.

Bogdanov, S. (2017), Honey composition, *Honey B.*, **Vol. 1**, pp. 1–10.

Bontempo, L., Camin, F., Ziller, L., Perini, M., Nicolini, G., & Larcher, R. (2017). Isotopic and elemental composition of selected types of Italian honey. *Measurement*, **Vol. *98*,** pp. 283-

289.

Bosshard, C., Spreiter, R., Degiorgi, L. and Günter, P. (2002), Infrared and Raman spectroscopy of the organic crystal DAST: Polarization dependence and contribution of molecular vibrations to the linear electro-optic effect, *Phys. Rev. B - Condens. Matter Mater. Phys.*, **Vol. 66 No. 20**, pp. 1–9.

Bougrini, M., Florea, A., Cristea, C., Sandulescu, R., Vocanson, F., Errachid, A., Bouchikhi, B., *et al.* (2016), Development of a novel sensitive molecularly imprinted polymer sensor based on electropolymerization of a microporous-metal-organic framework for tetracycline detection in honey, *Food Control*, Elsevier Ltd, **Vol. 59**, pp. 424–429.

Breiman, L.E.O. (2001), Random Forests, **Vol. 1**, pp. 5–32.

Brown, S. D. (2010). Introduction to multivariate statistical analysis in chemometrics. *Applied Spectroscopy*, **Vol. 64,** pp. 112A-112A.

Jandrić, Z., Haughey, S. A., Frew, R. D., McComb, K., Galvin-King, P., Elliott, C. T., & Cannavan, A. (2015). Discrimination of honey of different floral origins by a combination of various chemical parameters. *Food Chemistry*, **Vol. 189**, 52-59.

Carey, D. L., Ong, K. L., Whiteley, R., Crossley, K. M., Crow, J., & Morris, M. E. (2017). Predictive modelling of training loads and injury in Australian football. **Vol.1706**. pp. 04336.

Cebi, N., Ekinci, C., Esra, M., Yayla, A. and Sagdic, O. (2017), Detection of L -Cysteine in wheat flour by Raman microspectroscopy combined chemometrics of HCA and PCA, **Vol. 228**, pp. 116–124.

Chih-Chung, C. (2011). LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology*, Vol. *2*, pp. 27-1.

Chang, K.H. (2011), Limit of Detection and Its Establishment in Analytical Chemistry, *Heal. Environ. J.*, **Vol. 2 No. 1**, pp. 38–43.

Chau, K. (2004), *Advances in Neural Networks - ISNN 2004*, **Vol. 3174, pp. 45 - 56**

Chuntonov, L. and Rubtsov, I. V. (2020), Surface-enhanced ultrafast two-dimensional vibrational spectroscopy with engineered plasmonic nano-antennas, *J. Chem. Phys.*, AIP Publishing, LLC, **Vol. 153 No. 5**, p. 050902.

Cimpoiu, C., Hosu, A., Miclaus, V. and Puscas, A. (2013), Spectrochimica Acta Part A : Molecular and Biomolecular Spectroscopy Determination of the floral origin of some Romanian honeys on the basis of physical and biochemical properties, *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, Elsevier B.V., **Vol. 100**, pp. 149–154.

Cordella, C., Militao, J. S., Clément, M. C., Drajnudel, P., & Cabrol-Bass, D. (2005). Detection and quantification of honey adulteration via direct incorporation of sugar syrups or bee-feeding: preliminary study using high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) and chemometrics. *Analytica Chimica Acta*, **Vol. 531,** pp. 239-248.

Cordella, C., Milit, J.S.L.T., Drajnudel, P. and Cabrol-bass, D. (2005), Detection and

quantification of honey adulteration via direct incorporation of sugar syrups or bee-feeding : preliminary study using high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) and chemometrics, **Vol. 531**, pp. 239–248.

Corradini, C., Cavazza, A., & Bignardi, C. (2012). High-performance anion-exchange chromatography coupled with pulsed electrochemical detection as a powerful tool to evaluate carbohydrates of food interest: principles and applications. *International Journal of Carbohydrate Chemistry*, **Vol. 2012,** pp. *2012*.

Corvucci, F., Nobili, L., Melucci, D. and Grillenzoni, F.V. (2015), The discrimination of honey origin using melissopalynology and Raman spectroscopy techniques coupled with multivariate analysis, *Food Chem.*, Elsevier Ltd, **Vol. 169**, pp. 297–304.

Cotte, J.F., Casabianca, H., Lhéritier, J., Perrucchietti, C., Sanglar, C., Waton, H. and Grenier-Loustalot, M.F. (2007), Study and validity of 13C stable carbon isotopic ratio analysis by mass spectrometry and 2H site-specific natural isotopic fractionation by nuclear magnetic resonance isotopic measurements to characterize and control the authenticity of honey, *Anal. Chim. Acta*, **Vol. 582 No. 1**, pp. 125–136.

Cui, L., Wu, D.Y., Wang, A., Ren, B. and Tian, Z.Q. (2010), Charge-transfer enhancement involved in the SERS of adenine on Rh and Pd demonstrated by ultraviolet to visible laser excitation, *J. Phys. Chem. C*, **Vol. 114 No. 39**, pp. 16588–16595.

Das, R.S. and Agrawal, Y.K. (2011), Raman spectroscopy: Recent advancements, techniques and applications, *Vib. Spectrosc.*, Elsevier B.V., **Vol. 57 No. 2**, pp. 163–176.

Deckert, V. and Wiley, J. (2009), Tip-Enhanced Raman Spectroscopy, **Vol. 6**, pp. 1336–1337.

Dégardin, K., Roggo, Y., Been, F. and Margot, P. (2011), Analytica Chimica Acta Detection and chemical profiling of medicine counterfeits by Raman spectroscopy and chemometrics, **Vol. 705**, pp. 334–341.

Deng, L., Wu, Y., Hu, X., Liang, L., Ding, Y., Li, G., ... & Xie, Y. (2020). Rethinking the performance comparison between SNNS and ANNS. *Neural networks*, **Vol. *121*,** pp. 294-307.

Donfack, P. and Materny, A. (2009), Visible Raman spectroscopy for the discrimination of olive oils from different vegetable oils and the detection of adulteration, **Vol. 2009 No. April**, pp. 1284–1289.

Du, B., Wu, L., Xue, X., Chen, L., Li, Y., Zhao, J. and Cao, W. (2015), Rapid Screening of Multiclass Syrup Adulterants in Honey by Ultrahigh-Performance Liquid Chromatography/Quadrupole Time of Flight Mass Spectrometry, *J. Agric. Food Chem.*, **Vol. 63 No. 29**, pp. 6614–6623.

Eriksson, L., Andersson, P.L., Johansson, E. and Tysklind, M. (2006), Megavariate analysis of environmental QSAR data. Part I - A basic framework founded on principal component analysis (PCA), partial least squares (PLS), and statistical molecular design (SMD), *Mol. Divers.*, **Vol. 10 No. 2**, pp. 169–186.

Bertelli, D., Lolli, M., Papotti, G., Bortolotti, L., Serra, G., & Plessi, M. (2010). Detection of honey adulteration by sugar syrups using one-dimensional and two-dimensional high-resolution

nuclear magnetic resonance. *Journal of agricultural and food chemistry*, **Vol. 58,** pp. 8495-8501.

Etchegoin, P., Maher, R.C., Cohen, L.F., Hartigan, H., Brown, R.J.C., Milton, M.J.T. and Gallop, J.C. (2003), New limits in ultrasensitive trace detection by surface enhanced Raman scattering (SERS), *Chem. Phys. Lett.*, **Vol. 375 No. 1–2**, pp. 84–90.

López-Díez, E. C., Bianchi, G., & Goodacre, R. (2003). Rapid quantitative assessment of the adulteration of virgin olive oils with hazelnut oils using Raman spectroscopy and chemometrics. *Journal of agricultural and food chemistry*, **Vol. 51,** pp. 6145-6150.

Fakhlaei, R., Selamat, J., Khatib, A., Razis, A. F. A., Sukor, R., Ahmad, S., & Babadi, A. A. (2020). The toxic impact of honey adulteration: A review. *Foods*, **Vol. 9**, pp. 1538.

Ferreiro-gonzález, M., Espada-bellido, E., Ferreiro-gonzález, M., Espada-bellido, E., Guillén-cueto, L., Palma, M., Barroso, C.G., *et al.* (2018), Rapid quantification of honey adulteration by visible-near infrared spectroscopy combined with chemometrics Talanta Rapid quanti fi cation of honey adulteration by visible-near infrared spectroscopy combined with chemometrics, *Talanta*, Elsevier B.V., **Vol. 188 No. February 2019**, pp. 288–292.

Gaft, M. and Nagli, L. (2008), UV gated Raman spectroscopy for standoff detection of explosives, *Opt. Mater. (Amst).*, **Vol. 30 No. 11**, pp. 1739–1746.

Gallardo-Velázquez, T., Osorio-Revilla, G., Loa, M.Z. de and Rivera-Espinoza, Y. (2009), Application of FTIR-HATR spectroscopy and multivariate analysis to the quantification of adulterants in Mexican honeys, *Food Res. Int.*, Elsevier Ltd, **Vol. 42 No. 3**, pp. 313–318.

Gallo, C. and Bonis, M. De. (2013), A neural network model for forecasting photovoltaic deployment in italy, *Int. J. Sustain. Energy ...*, **Vol. 1 No. 1**, pp. 1–13.

De Gelder, J., De Gussem, K., Vandenabeele, P., & Moens, L. (2007). Reference database of Raman spectra of biological molecules. *Journal of Raman Spectroscopy: An International Journal for Original Work in all Aspects of Raman Spectroscopy, Including Higher Order Processes, and also Brillouin and Rayleigh Scattering*, **Vol. 38**, pp. 1133-1147.

Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern recognition letters*, **Vol. 31,** pp. 2225-2236.

George, T., Rufus, E. and Alex, Z.C. (2017), ARTIFICIAL NEURAL NETWORK BASED ULTRASONIC SENSOR SYSTEM FOR DETECTION OF ADULTERATION IN EDIBLE OIL, **Vol. 12 No. 6**, pp. 1568–1579.

Ghasemi-Varnamkhasti, M., Apetrei, C., Lozano, J. and Anyogu, A. (2018), Potential use of electronic noses, electronic tongues and biosensors as multisensor systems for spoilage examination in foods, *Trends Food Sci. Technol.*, **Vol. 80**, pp. 71–92.

Ghosh, P. K., & Jayas, D. S. (2009). Use of spectroscopic data for automation in food processing industry. *Sensing and Instrumentation for Food Quality and Safety*, **Vol. 3**, pp. 3-11.

Tura, A. G., & Seboka, D. B. (2019). Review on honey adulteration and detection of adulterants in honey. *Gastroenterology*, **Vol. 4**, pp. 1-6.

Goodacre, R., Radovic, B.S. and Anklam, E. (2002), Progress toward the rapid nondestructive assessment of the floral origin of European honey using dispersive Raman spectroscopy, *Appl. Spectrosc.*, **Vol. 56 No. 4**, pp. 521–527.

Grazia Mignani, A., Ciaccheri, L., Mencaglia, A.A., Di Sanzo, R., Carabetta, S. and Russo, M. (2016), Dispersive raman spectroscopy for the nondestructive and rapid assessment of the quality of southern Italian honey types, *J. Light. Technol.*, **Vol. 34 No. 19**, pp. 4479–4485.

Gualtieri, J.A. and Chettri, S. (2000), Support Vector Machines for classification of hyperspectral data, *Int. Geosci. Remote Sens. Symp.*, **Vol. 2 No. March**, pp. 813–815.

Guler, A., Kocaokutgen, H., Garipoglu, A. V., Onder, H., Ekinci, D. and Biyik, S. (2014), Detection of adulterated honey produced by honeybee (Apis mellifera L.) colonies fed with different levels of commercial industrial sugar (C3 and C4 plants) syrups by the carbon isotope ratio analysis, *Food Chem.*, **Vol. 155**, pp. 155–160.

Choudhary, A., Gupta, N., Hameed, F., & Choton, S. (2020). An overview of food adulteration: Concept, sources, impact, challenges and detection. *International Journal of Chemical Studies*, **Vol. 8**, pp. 2564-2573.

Hai, Z., & Wang, J. (2006). Detection of adulteration in camellia seed oil and sesame oil using an electronic nose. *European Journal of Lipid Science and Technology*, **Vol. 108**, pp. 116-124.

Haynes, C. L., McFarland, A. D., & Van Duyne, R. P. (2005). Surface-enhanced Raman spectroscopy **Vol. 77 No. 17,** pp. 338 -A.

Heidarbeigi, K., Mohtasebi, S.S., Foroughirad, A., Rafiee, S. and Rezaei, K. (2015), Detection of Adulteration in Saffron Samples Using Electronic Nose Detection of Adulteration in Saffron Samples Using Electronic Nose, *Int. J. Food Prop.*, Taylor & Francis, **Vol. 18 No. 7**, pp. 1391–1401.

Iqbal, T., Ashfaq, Z., Afsheen, S., Ijaz, M., Khan, M. Y., Rafique, M., & Nabi, G. (2020). Surface-enhanced Raman scattering (SERS) on 1D nano-gratings. *Plasmonics*, **Vol. 15,** pp. 1053-1059.

Ismail, W.I.W., Samat, S., Enchang, F.K., Razak, A.A. and Nor Hussein, F. (2018), Adulterated Honey Consumption can Induce Obesity, Increase Blood Glucose Level and Demonstrate Toxicity Effects, *Sains Malaysiana*, **Vol. 47 No. 2**, pp. 353–365.

Jaafar, M. B., Othman, M. B., Yaacob, M., Talip, B. A., Ilyas, M. A., Ngajikin, N. H., & Fauzi, N. A. M. (2020). A review on honey adulteration and the available detection approaches. *International Journal of Integrated Engineering*, Vol. *12* No. 2, pp. 125-131.

Joachims, T. (2005), Text Categorization with SVM: Learning with Many Relevant Features, *Eur. Conf. Mach. Learn.*, **Vol. 29**, pp. 137–142.

Kadhm, M. S., Mohammed, M. J., & Ayad, H. (2021). An accurate signature verification system based on proposed HSC approach and ANN architecture. *Indonesian Journal of Electrical Engineering and Computer Science*, **Vol. *21* No. 1,** pp. 215-223.

Khan, K. M., Krishna, H., Majumder, S. K., & Gupta, P. K. (2015). Detection of urea adulteration in milk using near-infrared Raman spectroscopy. *Food analytical methods*, **Vol. *8 No.* 1,** pp.

93-102.

Kim, M., Yun, J., Cho, Y., Shin, K., Jang, R., Bae, H. J., & Kim, N. (2019). Deep learning in medical imaging. *Neurospine*, **Vol. *16* No. 4,** pp. 657.

Kneipp, K., Kneipp, H., Itzkan, I., Dasari, R.R. and Feld, M.S. (2002), **Vol. 14**, pp. 597–624.

Korth, W., & Ralston, J. (2002). Techniques for the detection of adulterated honey. *RIRDC publication*, **Vol. 02**, pp. 47

Krafft, C., Dietzek, B., & Popp, J. (2009). Raman and CARS microspectroscopy of cells and tissues. *Analyst*, **Vol. *134* No. 6**, pp. 1046-1057.

Krishnan, S. (2019), heory of raman spectroscopy, **Vol 18**, pp. 1–5.

Kružík, V., Grégrová, A., Rajchl, A. and Čížková, H. (2017), Study on Honey Quality Evaluation and Detection of Adulteration by Analysis of Volatile Compounds, *J. Apic. Sci.*, **Vol. 61 No. 1**, pp. 17–27.

Kudelski, A. (2008), Analytical applications of Raman spectroscopy, *Talanta*, **Vol. 76 No. 1**, pp. 1–8.

Kumar, N., Mignuzzi, S., Su, W., & Roy, D. (2015). Tip-enhanced Raman spectroscopy: principles and applications. *EPJ Techniques and Instrumentation*, **Vol. *2* No. 1**, pp. 9.

Kwon, S.J. (2011), Artificial neural networks, *Artif. Neural Networks*, **Vol .7**, pp. 1–426.

Lee, K.Y., Chung, N. and Hwang, S. (2016), Application of an artificial neural network (ANN) model for predicting mosquito abundances in urban areas, *Ecol. Inform.*, The Authors, **Vol. 36**, pp. 172–180.

Lenhardt, L., Zeković, I., Dramićanin, T., Tešić, Ž., Milojković-Opsenica, D., & Dramićanin, M. D. (2014). Authentication of the botanical origin of unifloral honey by infrared spectroscopy coupled with support vector machine algorithm. *Physica Scripta*, *Vol.* **T162**, pp. 014042.

Li, J., Cheng, J.H., Shi, J.Y. and Huang, F. (2012), Brief introduction of back propagation (BP) neural network algorithm and its improvement, *Adv. Intell. Soft Comput.*, **Vol. 169 AISC No. VOL. 2**, pp. 553–558.

Li, S., Shan, Y., Zhu, X., Zhang, X. and Ling, G. (2012), Detection of honey adulteration by high fructose corn syrup and maltose syrup using Raman spectroscopy, *J. Food Compos. Anal.*, Elsevier Inc., **Vol. 28 No. 1**, pp. 69–74.

Li, S., Zhang, X., Shan, Y., Su, D., Ma, Q., Wen, R. and Li, J. (2017), Qualitative and quantitative detection of honey adulterated with high-fructose corn syrup and maltose syrup by using near-infrared spectroscopy, *Food Chem.*, **Vol. 218**, pp. 231–236.

Li, Y. S., & Church, J. S. (2014). Raman spectroscopy in the analysis of food and pharmaceutical nanomaterials. *Journal of food and drug analysis*, **Vol. *22* No. 1**, pp. 29-48.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, **Vol. 2 No. 3**, pp. 18-22.

Lin, W.J. and Chen, J.J. (2013), Class-imbalanced classifiers for high-dimensional data, *Brief.*

Bioinform., **Vol. 14 No. 1**, pp. 13–26.

Lin, Y. K., Leong, H. Y., Ling, T. C., Lin, D. Q., & Yao, S. J. (2021). Raman spectroscopy as process analytical tool in downstream processing of biotechnology. *Chinese Journal of Chemical Engineering*, **Vol. *30***, pp. 204-211.

Liu, H., Xue, Y., Li, J., Wu, W., & Lan, J. (2019). Investigation of laser power output and its effect on Raman spectrum for marine metal corrosion cleaning. *Energies*, **Vol. 13 No. 1**, pp. 12

Liu, Y., Wang, Y. and Zhang, J. (2012), New machine learning algorithm: Random forest, *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, **Vol. 7473 LNCS**, pp. 246–252.

Long, D.A. (2005), Introductory Raman Spectroscopy. John R. Ferraro, Kazuo Nakamoto and Chris W. Brown. Academic Press, Amsterdam, Second Edition, 2003. xiii + 434, *J. Raman Spectrosc.*, **Vol. 36 No. 10**, pp. 1012–1012.

Ma, L., Li, M., Ma, X., Cheng, L., Du, P. and Liu, Y. (2017), A review of supervised object-based land-cover image classification, *ISPRS J. Photogramm. Remote Sens.*, The Authors, **Vol. 130**, pp. 277–293.

McCreery, R. L. (2001). Raman spectroscopy for chemical analysis. *Measurement science and technology*, **Vol. *12* No. 5**, pp. 653.

McNay, G., Eustace, D., Smith, W. E., Faulds, K., & Graham, D. (2011). Surface-enhanced Raman scattering (SERS) and surface-enhanced resonance Raman scattering (SERRS): a review of applications. *Applied spectroscopy*, **Vol. *65* No. 8**, pp. 825-837.

Megherbi, M., Herbreteau, B., Faure, R. and Salvador, A. (2009), Polysaccharides as a marker for detection of corn sugar syrup addition in honey, *J. Agric. Food Chem.*, **Vol. 57 No. 6**, pp. 2105–2111.

Mellado-mojica, E., Seeram, N.P. and López, M.G. (2016), Journal of Food Composition and Analysis Comparative analysis of maple syrups and natural sweeteners : Carbohydrates composition and classi fi cation (differentiation) by, *J. Food Compos. Anal.*, Elsevier Inc., **Vol. 52**, pp. 1–8.

Meyer, C., Hühn, S., Jungbauer, M., Merten, S., Damaschke, B., Samwer, K., & Moshnyaga, V. (2017). Tip-enhanced Raman spectroscopy (TERS) on double perovskite La2CoMnO6 thin films: field enhancement and depolarization effects. *Journal of Raman Spectroscopy*, **Vol. *48 No.* 1**, pp. 46-52.

Michel, P., & El Kaliouby, R. (2003). Real time facial expression recognition in video using support vector machines. In *Proceedings of the 5th international conference on Multimodal interfaces* **Vol. 2003**, pp. 258-264.

Mishra, S., Kamboj, U., Kaur, H. and Kapur, P. (2010), Detection of jaggery syrup in honey using near-infrared spectroscopy, *Int. J. Food Sci. Nutr.*, **Vol. 61 No. 3**, pp. 306–315.

Matthew, W. (2011). Bias of the random forest out-of-bag (OOB) error for certain input parameters. *Open Journal of Statistics*, **Vol. 01 No. 03**, pp. *2011*.

Morales, V., Corzo, N. and Sanz, M.L. (2008), HPAEC-PAD oligosaccharide analysis to detect adulterations of honey with sugar syrups, *Food Chem.*, **Vol. 107 No. 2**, pp. 922–928.

Müller, A., & Sumpf, B. (2020). Compact diode laser based light source with alternating dual-wavelength emission at 532 nm. *Applied Physics B*, **Vol. 126 No. 8**, pp. 1-7.

Nakamoto, K. and Brown, C.W. (2003), *Introductory Raman Spectroscopy Vol 3. pp. 3 -15*.

Onkangi, J. N. (2018). *Nuclear Forensics Analysis Of Fission Products By Means Of Chemometric Laser Induced Breakdown Spectroscopy* (Doctoral dissertation, University of Nairobi).

Oliveira, L.F.C.D.E. and Bara, R.C. (2002), Fourier Transform Raman Spectroscopy of Honey, **Vol. 56 No. 3**, pp. 0–5.

Ribeiro, R. D. O. R., Mársico, E. T., da Silva Carneiro, C., Monteiro, M. L. G., Júnior, C. C., & de Jesus, E. F. O. (2014). Detection of honey adulteration of high fructose corn syrup by Low Field Nuclear Magnetic Resonance (LF 1H NMR). *Journal of Food Engineering*, **Vol. *135***, pp. 39-43.

Oliveira, R. De, Ribeiro, R., Teixeira, E., Carneiro, S., Lúcia, M., Monteiro, G., Adam, C., *et al.* (2014), LWT - Food Science and Technology Classi fi cation of Brazilian honeys by physical and chemical analytical methods and low fi eld nuclear magnetic resonance (LF 1 H NMR), *LWT - Food Sci. Technol.*, Elsevier Ltd, **Vol. 55 No. 1**, pp. 90–95.

Oliver, T. A. (2018). Recent advances in multidimensional ultrafast spectroscopy. *Royal Society Open Science*, **Vol. *5 No.* 1,** pp. 171425.

Oroian, M., Ropciuc, S. and Paduret, S. (2018), Honey Adulteration Detection Using Raman Spectroscopy, *Food Anal. Methods*, Food Analytical Methods, **Vol. 11 No. 4**, pp. 959–968.

Otchere, D. A., Ganat, T. O. A., Gholami, R., & Ridha, S. (2021). Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models. *Journal of Petroleum Science and Engineering*, **Vol. *200***, pp. 108182.

Özbalci, B., Boyaci, I.H., Topcu, A., Kadilar, C. and Tamer, U. (2013), Rapid analysis of sugars in honey by processing Raman spectrum using chemometric methods and artificial neural networks, *Food Chem.*, **Vol. 136 No. 3–4**, pp. 1444–1452.

Pandey, R., Paidi, S.K., Kang, J.W., Spegazzini, N., Dasari, R.R., Valdez, T.A. and Barman, I. (2015), Discerning the differential molecular pathology of proliferative middle ear lesions using Raman spectroscopy, *Sci. Rep.*, Nature Publishing Group, **Vol. 5 No. August**, pp. 1–8.

Paradkar, M.M. and Irudayaraj, J. (2002), Discrimination and classification of beet and cane inverts in honey by FT-Raman spectroscopy, *Food Chem.*, **Vol. 76 No. 2**, pp. 231–239.

Peng, D., Bi, Y., Ren, X., Yang, G., Sun, S. and Wang, X. (2015), Detection and quantification of adulteration of sesame oils with vegetable oils using gas chromatography and multivariate data analysis, *FOOD Chem.*, Elsevier Ltd, **Vol. 188**, pp. 415–421.

Petry, R., Schmitt, M. and Popp, J. (2003), Raman spectroscopy - A prospective tool in the life sciences, *ChemPhysChem*, **Vol. 4 No. 1**, pp. 14–30.

Pettinger, B., Schambach, P., Villagómez, C. J., & Scott, N. (2012). Tip-enhanced Raman spectroscopy: near-fields acting on a few molecules. *Annual review of physical chemistry*, **Vol. 63,** pp. 379-399.

Pierna, J.A.F., Abbas, O., Dardenne, P. and Baeten, V. (2011), Discrimination of Corsican honey by FT-Raman spectroscopy and chemometrics, *Biotechnol. Agron. Soc. Environ.*, **Vol. 15 No. 1**, pp. 75–84.

Plumb, A.P., Rowe, R.C., York, P. and Brown, M. (2005), Optimisation of the predictive ability of artificial neural network (ANN) models : A comparison of three ANN programs and four classes of training algorithm, **Vol. 25**, pp. 395–405.

Qi, B., Zhao, C., Youn, E. and Nansen, C. (2011), Use of weighting algorithms to improve traditional support vector machine based classifications of reflectance data, *Opt. Express*, **Vol. 19 No. 27**.

Raanan, D., Ren, L., Oron, D. and Silberberg, Y. (2018), Impulsive Raman spectroscopy via precision measurement of frequency shift with low energy excitation, *Opt. Lett.*, **Vol. 43 No. 3**

Kumar, N., Ranjan, R., Kumar, Y., Patel, S. S., Krishna, V. S., Appaiah, A., ... & Panchariya, P. C. (2021, March). Discrimination of Various pure Honey samples and its Adulterants using FTIR Spectroscopy Coupled with Chemometrics. In *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)* **Vol. 1**, pp. 808-811.

Rifna, E.J., Pandiselvam, R., Kothakota, A., Rao, K.V.S., Dwivedi, M., Kumar, M., Thirumdas, R., *et al.* (2022), Advanced process analytical tools for identification of adulterants in edible oils – A review, *Food Chem.*, Elsevier Ltd, **Vol. 369 No. July 2021**, p. 130898.

Różańska, A., Dymerski, T., & Namieśnik, J. (2018). Novel analytical method for detection of orange juice adulteration based on ultra-fast gas chromatography. *Monatshefte für Chemie-Chemical Monthly*, **Vol. 149 No. 9**, pp. 1615-1621.

Robert, B. (2009), Resonance Raman spectroscopy, *Photosynth. Res.*, **Vol. 101 No. 2–3**, pp. 147–155.

Rocchetta, R., Bellani, L., Compare, M., Zio, E., & Patelli, E. (2019). A reinforcement learning framework for optimal operation and maintenance of power grids. *Applied energy*, **Vol. 241,** pp. 291-301.

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M. and Chica-Rivas, M. (2015), Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines, *Ore Geol. Rev.*, Elsevier B.V., **Vol. 71**, pp. 804–818.

Roussel, C., Vanthuyne, N., Serradeil-albalat, M. and Vallejos, J. (2003), T rue or apparent reversal of elution order during chiral high-performance liquid chromatography monitored by a polarimetric detector under different mobile phase conditions, **Vol. 995**, pp. 79–85.

Ruiz-Matute, A.I., Brokl, M., Soria, A.C., Sanz, M.L. and Martínez-Castro, I. (2010), Gas chromatographic-mass spectrometric characterisation of tri- and tetrasaccharides in honey, *Food Chem.*, Elsevier Ltd, **Vol. 120 No. 2**, pp. 637–642.

Ruiz-Matute, A.I., Soria, A.C., Martínez-Castro, I. and Sanz, M.L. (2007), A new methodology based on GC-MS to detect honey adulteration with commercial syrups, *J. Agric. Food Chem.*, **Vol. 55 No. 18**, pp. 7264–7269.

Salvador, L., Guijarro, M., Rubio, D., Aucatoma, B., Guillén, T., Vargas Jentzsch, P., ... & Ramos Guerrero, L. (2019). Exploratory monitoring of the quality and authenticity of commercial honey in Ecuador. *Foods*, **Vol. *8* No. 3**, pp. 105.

de Santana, F. B., Mazivila, S. J., Gontijo, L. C., Neto, W. B., & Poppi, R. J. (2018). Rapid discrimination between authentic and adulterated andiroba oil using FTIR-HATR spectroscopy and random forest. *Food Analytical Methods*, Vol. *11* No. 7, pp. 1927-1935.

Schmitt, M. and Popp, J. (2006), Raman spectroscopy at the beginning of the twenty-first century, *J. Raman Spectrosc.*, **Vol. 37 No. 1–3**, pp. 20–28.

Se, K.W., Wahab, R.A., Syed Yaacob, S.N. and Ghoshal, S.K. (2019a), Detection techniques for adulterants in honey: Challenges and recent trends, *J. Food Compos. Anal.*, Elsevier Inc., **Vol. 80**, pp. 16–32.

Se, K. W., Wahab, R. A., Yaacob, S. N. S., & Ghoshal, S. K. (2019). Detection techniques for adulterants in honey: Challenges and recent trends. *Journal of Food Composition and Analysis*, **Vol. *80*,** pp. 16-32.

Segal, M.R. (2004), Machine Learning Benchmarks and Random Forest Regression Publication Date Machine Learning Benchmarks and Random Forest Regression, *Cent. Bioinforma. Mol. Biostat.*, **Vol. 2004**. pp. 15.

Selin, R., Hakki, I., Efe, H. and Tamer, U. (2013), Determination of butter adulteration with margarine using Raman spectroscopy, *Food Chem.*, Elsevier Ltd, **Vol. 141 No. 4**, pp. 4397–4403.

Seo, S. (2006), A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets Vol. 127 No. 1, pp. 65 - 74.

Shafiee, S., Polder, G., Minaei, S., Moghadam-Charkari, N., van Ruth, S. and Kuś, P.M. (2016), Detection of Honey Adulteration using Hyperspectral Imaging, *IFAC-PapersOnLine*, **Vol. 49 No. 16**, pp. 311–314.

Sharma, B., Frontiera, R.R., Henry, A.I., Ringe, E. and Van Duyne, R.P. (2012), SERS: Materials, applications, and the future, *Mater. Today*, Elsevier Ltd, **Vol. 15 No. 1–2**, pp. 16–25.

Sivakesava, S. and Irudayaraj, J. (2001), Detection of inverted beet sugar adulteration of honey by FTIR spectroscopy, *J. Sci. Food Agric.*, **Vol. 81 No. 8**, pp. 683–690.

Sivakesava, S., & Irudayaraj, J. (2002). Classification of simple and complex sugar adulterants in honey by mid-infrared spectroscopy. *International journal of food science & technology*, **Vol. *37* No. 4**, pp. 351-360.

Smith, E. and Dent, G. (2019), *Modern Raman Spectroscopy*, *Mod. Raman Spectrosc Vol.4, pp 1-10.*

Snežzana Miljanić, Leo Frkanec, Tomislav Biljan, 3 Zlatko Meić Mladen Žini ć. (2007), Recent

Advances in linear and nonlinear Raman spectroscopy I, *J. Raman Spectrosc.*, **Vol. 38 No. April**, pp. 1538–1553.

Spiteri, M., Jamin, E., Thomas, F., Rebours, A., Lees, M., Rogers, K. M., & Rutledge, D. N. (2015). Fast and global authenticity screening of honey using 1H-NMR profiling. *Food Chemistry*, **Vol. 189**, pp. 60-66.

Statnikov, A., Wang, L. and Aliferis, C.F. (2008), A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification, **Vol. 10**, pp. 1–10.

Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics*, **Vol. 8 No. 1,** pp. 1-21.

Tahsin, M., Berna, N., Said, O., Karaman, S., Dertli, E., Sagdic, O. and Arici, M. (2014), Steady , dynamic and creep rheological analysis as a novel approach to detect honey adulteration by fructose and saccharose syrups : Correlations with HPLC-RID results, *FRIN*, Elsevier B.V., **Vol. 64**, pp. 634–646.

Tan, R., Kelley, D. F., & Kelley, A. M. (2019). Resonance hyper-Raman scattering from CdSe and CdS nanocrystals. *The Journal of Physical Chemistry C*, **Vol. 123 No. 26,** pp. 16400-16405.

Megherbi, M., Herbreteau, B., Faure, R., Dessalces, G., & Grenier-Loustalot, M. F. (2008). Solid phase extraction of oligo-and polysaccharides; application to maltodextrins and honey qualitative analysis. *Journal of liquid chromatography & related technologies*, **Vol. 31 No. 7**, pp. 1033-1046.

Le Thi, H.A., Nguyen, V.V. and Ouchani, S. (2008), Gene selection for cancer classification using DCA, *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, **Vol. 5139 LNAI**, pp. 62–72.

Vandewalle, J. (1999), SpringerLink - Neural Processing Letters, Volume 9, Number 3, *Springerlink.Com*, **Vol. 9 No. 3**, pp. 293–300.

Verma, P. (2017). Tip-enhanced Raman spectroscopy: technique and recent advances. *Chemical reviews*, **Vol. 117 No. 9,** pp. 6447-6466.

Peica, N. (2006). *Vibrational spectroscopy and density functional theory calculations on biological molecules* (Doctoral dissertation, Universität Würzburg).

Views, T.H.E., The, O.F., The, P., That, V., Society, M., Views, F. and Entertainment, C. (2019), Know Why Blackstrap Molasses Is Far Superior than Honey ! What is Blackstrap Molasses ?, **Vol. 45**, pp. 1–10.

Vogt, W. (2015), Lods, *Dict. Stat. Methodol.*, **Vol. 18 No. 12**, pp. 112–114.

Wang, S. Bin and Wu, C.F. (2006), Selections of working conditions for creep feed grinding. Part(III): Avoidance of the workpiece burning by using improved BP neural network, *Int. J. Adv. Manuf. Technol.*, **Vol. 28 No. 1–2**, pp. 31–37.

Wang, S., Guo, Q., Wang, L., Lin, L., Shi, H., Cao, H. and Cao, B. (2015), Detection of honey

adulteration with starch syrup by high performance liquid chromatography, *Food Chem.*, Elsevier Ltd, **Vol. 172**, pp. 669–674.

Wang, Y., Han, P., Lul, X., Wu, R., Huang, J. and Xt, N. (2006), The Performance Comparison of Adaboost and SVM Applied to SAR ATR, **Vol. 00**, pp. 1–4.

Wiley, J. (2006), Raman Spectroscopy : Theory G ´ abor Keresztury, *Spectroscopy*, **Vol. 06,** pp. 1–17.

Wu, W. and Massart, D.L. (1996), Artificial neural networks in classification of NIR spectral data: Selection of the input, *Chemom. Intell. Lab. Syst.*, **Vol. 35 No. 1**, pp. 127–135.

Yang, D. and Ying, Y. (2011), Applications of raman spectroscopy in agricultural products and food analysis: A review, *Appl. Spectrosc. Rev.*, **Vol. 46 No. 7**, pp. 539–560.

Yang, H., Irudayaraj, J. and Paradkar, M.M. (2005), Discriminant analysis of edible oils and fats by FTIR, FT-NIR and FT-Raman spectroscopy, *Food Chem.*, **Vol. 93 No. 1**, pp. 25–32.

Yeo, B., Stadler, J., Schmid, T., Zenobi, R. and Zhang, W. (2009), Tip-enhanced Raman Spectroscopy – Its status , challenges and future directions, *Chem. Phys. Lett.*, Elsevier B.V., **Vol. 472 No. 1–3**, pp. 1–13.

Yoshioka, M., Fujinaka, T. and Omatu, S. (2007), SAR image classification by support vector machine, *Image Process. Remote Sens.*, **Vol. 0000**, pp. 341–354.

Zábrodská, B. and Vorlová, L. (2014), Adulteration of honey and available methods for detection – a review, *Acta Vet. Brno*, **Vol. 83**, pp. S85–S102.

Zeitler, J.A., Taday, P.F., Newnham, D.A., Pepper, M., Gordon, K.C. and Rades, T. (2007), Terahertz pulsed spectroscopy and imaging in the pharmaceutical setting - a review, *J. Pharm. Pharmacol.*, **Vol. 59 No. 2**, pp. 209–223.

Zhao, F., Wang, W., Zhong, H., Yang, F., Fu, W., Ling, Y., & Zhang, Z. (2021). Robust quantitative SERS analysis with Relative Raman scattering intensities. *Talanta*, **Vol. *221***, pp. 121465.

Zhao, H., Feng, Y., Chen, W. and Jia, G. (2019), Application of invasive weed optimization and least square support vector machine for prediction of beef adulteration with spoiled beef based on visible Application of invasive weed optimization and least square support vector machine for prediction of be, *Meat Sci.*, Elsevier, **Vol. 151 No. March 2020**, pp. 75–81.

Zheng, M., Bi, J., Chen, Y., Wang, H., & Zhou, M. (2020). Fluorescence-enhanced second harmonic normal Raman scattering in β-carotene. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **Vol. 239**. *239*, pp. 118494.

Zhu, X., Li, S., Shan, Y., Zhang, Z., Li, G., Su, D. and Liu, F. (2010), Detection of adulterants such as sweeteners materials in honey using near-infrared spectroscopy and chemometrics, *J. Food Eng.*, Elsevier Ltd, **Vol. 101 No. 1**, pp. 92–97.

Zupan, J. (2008), Neural networks in chemistry, **Vol 3,** pp. 469–470.

APPENDICES

Table A: Carbohydrate content comparison between molasses and honey for 100 g sample (Views *et al.*, 2019)

Carbohydrate Contents	Molasses		Honey	
	Amount (g)	RDI (%)	Amount (g)	RDI (%)
Total Carbohydrate	74.73	25 %	84.4	27
Dietary Fiber	0	0%	0.2	1
Starch	-	-	-	-
Sugars	54. 49		82.12	
Sucrose	29.4		0.89	
Glucose	11.92		35.75	
Fructose	12.79		40.94	
Lactose	-		-	
Maltose	-		1.44	
Galactose	-		3.1	

Table B: Mineral content comparison between molasses and honey for 100 g sample (Views *et al.*, 2019)

Mineral content	Molasses		Honey	
	Amount (mg)	RDI (%)	Amount (mg)	RDI (%)
Calcium	205	20	6	1
Iron	4.2	26	0.42	2
Magnesium	242	60	2	0
Phosphorous	32	3	4	0
Potassium	1464	31	52	1
Sodium	37	2	4	0
Zinc	0.29	2	0.22	1
Copper	0.487	24	0.036	2
Manganese	1.53	76	0.08	4
Selenium	17.8 mcg	25	0.8	1
Fluoride	-	-	7 mcg	-

Table C: Water soluble Vitamin comparison between molasses and honey for 100 g sample (Views *et al.*, 2019)

	Molasses		Honey	
Vitamin Content	**Amount(mg)**	**RDI (%)**	**Amount(mg)**	**RDI (%)**
Vitamin A	0 IU	0	0 IU	0
Vitamin B6	0.67	34	0.24	1
Vitamin B12	0 mcg	0	0 mcg	0
Vitamin B12, Added	0 mcg	0	0 mcg	0
Vitamin C	0	0	0.5	0
Vitamin D	0 IU	0	0 IU	0
Vitamin D2	-	-	-	-
Vitamin D3	-	-	-	-
Vitamin D(D2+D3)	0 mcg	0	0 mcg	0
Vitamin E	0	0	0	0
Vitamin E, Added	0	0	0	0
Vitamin K	0 mcg	0	0 mcg	0
Thiamin	0.041	3	0	0
Riboflavin	0.002	0	0.038	2
Niacin	0.93	5	0.121	1
Pantothenic Acid	0.804	8	0.068	1
Folate	0 mcg	0	2 mcg	0
Folate, Food	0 mcg	0	2 mcg	0
Folate DFE	0 mcg DFE	0	2 mcg DFE	0
Choline	13.3	-	2.2	-
Betaine	-	-	1.7	-

Table D: Honey composition in g/100g (Bogdanov, 2017)

|	Blossom honey		Honeydew honey	
	average	min-max	average	min-max
Water content	17.2	15-20	16.3	15-20
Fructose	38.2	30-45	31.8	28-40
Glucose	31.3	24-40	26.1	19-32
Sucrose	0.7	0.1-4.8	0.5	0.1-4.7
Other disaccharides	5.0	28	4.0	16
Melezitose	<0.1		4.0	0.3-22.0
Erlose	0.8	0.56	1.0	0.16
Other oligosaccharides	3.6	0.5-1	13.1	0.1-6
Total sugars	*79.7*		*80.5*	
Minerals	0.2	0.1-0.5	0.9	0.6-2
Amino acids, proteins	0.3	0.2-0.4	0.6	0.4-0.7
Acids	0.5	0.2-0.8	1.1	0.8-1.5
pH	3.9	3.5-4.5	5.2	4.5-6.5
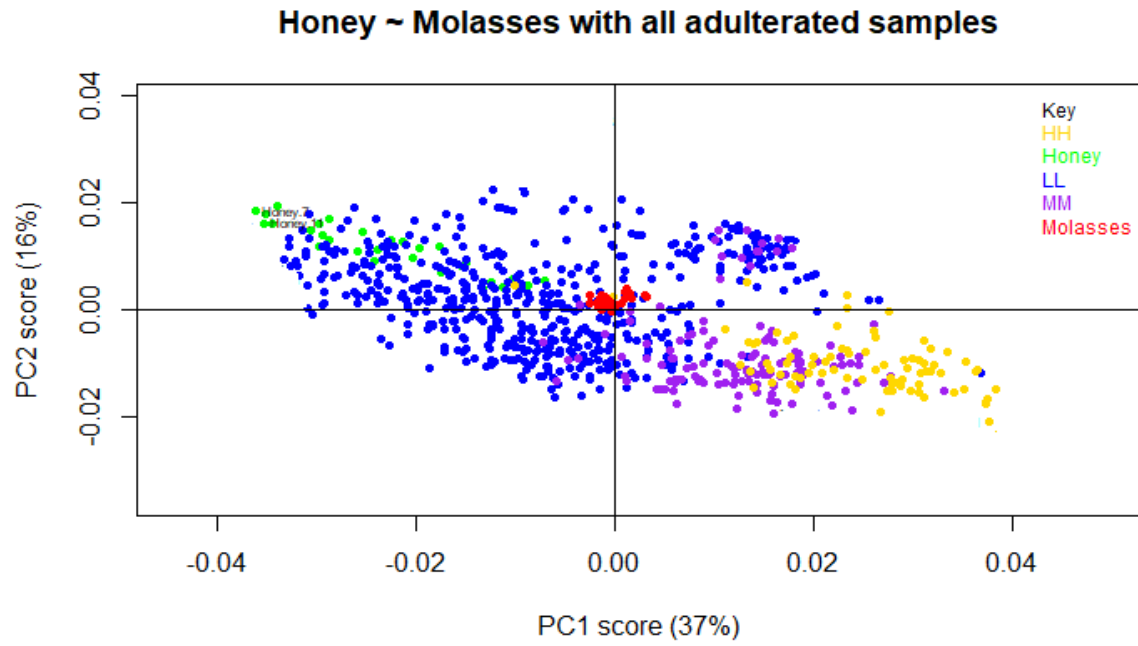
Figure 17: PCA score plot of authentic honey, molasses and all samples adulterated by molasses. LL – Low concentrations, MM – Middle concentrations, HH – High concentrations.

Appendix 1: ANN Model Script in R

```
# Read the Data
#data <- ANNF3_Scores.csv
data <- read.csv("ANNF2_Scores.csv")
str(data)
############################## data partitioning ##############################
#honey adulterated samples Data Partition
set.seed(123)
ind <- sample(2, nrow(data), replace = TRUE, prob = c(0.7, 0.3))
training <- data[ind==1,]
testing <- data[ind==2,]
trainX <- training[,2:6]
trainY <- training$conc
testX <- testing[,2:6]
testY <- testing$conc
# cluster <- makeCluster(detectCores() - 1) # convention to leave 1 core for OS
# registerDoParallel(cluster)
# set.seed(1234)
library(caret)
set.seed(1222)
NN1 <- train(conc~ Comp.1+Comp.2+Comp.3+Comp.4+Comp.5,
######+Comp.6+Comp.7+Comp.8+Comp.9+Comp.10,

        data=training,

        method = "neuralnet",

        trControl = trainControl(method = "CV",

                    number = 2,

                    returnResamp = "final",
```

```r
                    savePredictions = "final"),
        tuneGrid = data.frame (layer1 = 2:8,
                        layer2 = 16,
                        layer3 = 3),
        rep = 2,
        algorithm = "rprop+",
        threshold = 0.1,
        stepmax = 1e+05,
        preProc = c("center", "scale"))
print(NN1)
plot(NN1)
##################### training #####################
library(dplyr)
predicted<-NN1$pred$pred
actual<-NN1$pred$obs
training_model <- bind_cols(predicted=predicted, actual=actual)
write.csv(training_model,"ANN_model_results_training.csv")
####prediction##########
ANN_pred <- predict(NN1,testX)
postResample(pred = ANN_pred,obs = testY)
ANN_model_testing<-bind_cols(predicted=ANN_pred,Actual=testY)
write.csv(ANN_model_testing,"Ann_model_testing_results.csv")
## plot  observed vs predicted values for  test set
plot(y = NN1_pred, x = testY,
    xlab = 'prediction',
    ylab = 'observed')
abline(a = 0, b = 1, lty = 2, col = 2)
######################### end of ANN #################################
```

Appendix 2: PCA Script in R

```
library(ChemoSpec)
# Reading a matrix data file stored in the working directory
rawspec <- matrix2SpectraObject(gr.crit = c("Honey","LL","MM","HH"),
                    gr.cols = c("green","blue","purple","red"),
                    freq.unit = "Wavelength (/cm)",
                    int.unit = "Intensity",
                    descrip = "honey_molasses_Study",
                    in.file = "honey_adulterated_samples.csv",
                    out.file = "honey_aduletration",
                    chk = TRUE,
                    sep = ",",
                    dec = ".")
# Summarizing the data
sumSpectra(rawspec)
#creating a title
title<-expression(bolditalic(Raman)~bold(Spectra))
# Plotting all the spectra
plotSpectra(rawspec,
        main = title,
        which = c(1:800),
        yrange = c(0,2500),
        xlim =c(249,1810),
        offset = 100,
        showGrid = FALSE,
        lab.pos = 1780)
#abline(h=600,v=30)
# Plotting the spectra
```

```
plotSpectra(rawspec,
        main = title,
        which = c(2:800),
        yrange = c(0,2500),
        xlim =c(248,1810),
        offset = 100,
        lab.pos = 1280)
##########################################
spec3<-normSpectra(rawspec)
plotSpectra(spec3,
        main = title,
        which = c(2,5,8),
        yrange = c(0,250),
        xlim =c(249,1300),
        offset = 100,
        lab.pos = 1280)
##################################################
pca<-r_pcaSpectra(spec3, choice = "noscale")
plotScores(spec3, pca,
        main ="Honey ~ Molasses", pcs = c(1,2), tol = 0.01)
abline(h =0,v=0)
###############################################################
# getting the PC from PCA attributes and saving it as a csv
attributes(pca)
pca_scores <-pca[["x"]]
pca_scores
#write.csv(pca_scores,'All0.5_50_scores.csv')
###############################################################
```

```
plotScores(spec3, pca,

       main ="Adulterated_honey_samples", pcs = c(1,2), ellipse = "rob", tol = 0.01)
#potential outliers
diagnostics<-pcaDiag(spec3,

             pca,

             pcs = 2,

             #quantile = 0,95,

             plot = c("OD","SD"))
#number of pcs measured
plotScree(pca,

       style = "alt")
plotLoadings(spec3, pca, loads = c(1,2))
```

######################### **end of Principal Components Analysis** ####################

Appendix 3: Random Forest Script in R

```
# Read Data
#change p=1, m=2, y=3 and Z=4 in rawspec then transpose in excel and paste
#the 1,2 to$samples in Raman_pca_scores1
data <- read.csv(file.choose(), header = T)
#data <- Raman_pca_scores1
str(data)
#data2 <- read.csv("~/Desktop/CTG.csv", header = TRUE)
data$samples <- as.factor(data$samples)
str(data)
#select the first 10 PCs
#data2 <-data[,1:11]
#str(data2)
table(data$samples)
## Data Partition
set.seed(123)
ind <- sample(2, nrow(data), replace = TRUE, prob = c(0.7, 0.3))
train <- data[ind==1,]
test <- data[ind==2,]


############################# Random Forest #############################
library(randomForest)
set.seed(222)
Set1 <- randomForest(samples~., data=train,
            ntree = 15,
            mtry = 4,
            importance = TRUE,
```

```
          proximity = TRUE)

print(Set5)

attributes(Set5)

####  Prediction & Confusion Matrix - train data

library(caret)

p1 <- predict(Set5, train)

confusionMatrix(p1, train$samples)

##### Prediction & Confusion Matrix - test data

p2 <- predict(Set5, test)

confusionMatrix(p2, test$samples)

####### Error rate of Random Forest

plot(Set5)

# Tune mtry

t <- tuneRF(train[,-1], train[,1],

        stepFactor = 0.5,

        plot = TRUE,

        ntreeTry = 50,

        trace = TRUE,

        improve = 0.05)

####### No. of nodes for the trees

hist(treesize(Set5),

    main = "No. of Nodes for the Trees",

    col = "green")

####### Variable Importance

varImpPlot(Set5,

        sort = T,

        n.var = 10,

        main = "Top 10 - Variable Importance")
```

importance(Set5)

varUsed(Set5)

################################# **end of Random Forest** ########################

Appendix 4: Support Vector Machine Script in R

######################### **Support Vector Machine** #############################

```
# Read Data
data <- read.csv("Set1_Scores22.csv")
#data <- Raman_pca_scores1
str(data)
#data2 <- read.csv("~/Desktop/Set1_scores.csv", header = TRUE)
data$samples <- as.factor(data$samples)
str(data)
# Plot data
library(ggplot2)
ggplot(data = data, aes(x = Comp.1, y = Comp.2,
                 color = samples, shape = samples)) +
  geom_point(size = 2) +
  scale_color_manual(values=c("#000000", "#FF0000")) +
  theme(legend.position = "none")
#str(data2)
table(data$samples)
# Data Partition
set.seed(123)
ind <- sample(2, nrow(data), replace = TRUE, prob = c(0.6, 0.4))
train <- data[ind==1,]
str(train)
test <- data[ind==2,]
#support vector machines
library(e1071)
mymodel <- svm(samples~.,
```

```
        data = data,

        kernel="linear",

        cost=5, gamma=1,scale = FALSE)

summary(mymodel)

plot(mymodel,data=data,

    Comp.2~Comp.1,

    slice = list(Comp.3=3,Comp.4=4,

          Comp.5=5, Comp.6=6,

          Comp.7=7, Comp.8=8,

          Comp.9=9, Comp.10=10))

#confusion matrix and misclassification error

pred <- predict(mymodel,train)

tab <- table(predicted=pred,actual=train$samples)

tab

#accuracy

1-sum(diag(tab))/sum(tab)

mean(pred== train$samples)
```

#################### end of Support Vector Machine ###############

Mr.