ACIBADEM MEHMET ALI AYDINLAR UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

DATA ANALYSIS ON LIMIT OF DETECTION FOR VISCOELASTIC PROPORTIES BY QCM-D MEASUREMENTS

MEHMET MERT YÜKSEL

M.Sc. THESIS

DEPARTMENT OF BIOMEDICAL ENGINEERING

SUPERVISOR
Assist. Prof. Dr. Ceyhun Ekrem Kırımlı

SECONDARY SUPERVISOR
Assist. Prof. Dr. Elçim Elgün Kırımlı

ISTANBUL-2025

ACIBADEM MEHMET ALI AYDINLAR UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

DATA ANALYSIS ON LIMIT OF DETECTION FOR VISCOELASTIC PROPORTIES BY QCM-D MEASUREMENTS

MEHMET MERT YÜKSEL

M.Sc. THESIS

DEPARTMENT OF BIOMEDICAL ENGINEERING

SUPERVISOR
Assist. Prof. Dr. Ceyhun Ekrem Kırımlı

SECONDARY SUPERVISOR
Assist. Prof. Dr. Elçim Elgün Kırımlı

ISTANBUL-2025

DECLARATION

I declare that this thesis work is my own work, I had no unethical behavior at any stages from the planning to the writing of the thesis, I obtained all the information in this thesis in accordance with academic and ethicel rules, I cited all the information and comments that were not obtained with this thesis work, and I provided resources in the list of references. I also declare that there was no violation of any patents and copyrights during the study and writing of this thesis.

30/04/2025
Mehmet Mert Yüksel

PREFACE AND ACKNOWLEDGEMENT

First and foremost, Let me express my deepest gratitude to my primary supervisor, Assist. Prof. Ceyhun Ekrem KIRIMLI, and my co-supervisor, Assist. Prof. Elçim Elgün KIRIMLI, for their invaluable contributions to the successful finish of this master's thesis. Their expertise, mentorship, and steadfast support have been my compass throughout this demanding yet fulfilling journey. I am deeply thankful to Assist. Prof. Ceyhun Ekrem KIRIMLI for his profound understanding of the subject and his ability to offer clear, insightful, and innovative perspectives. His constructive critiques and encouragement pushed me to explore new dimensions in my research and helped me grow both academically and personally. I am equally grateful to Assist. Prof. Elçim Elgün KIRIMLI, whose meticulous approach and thoughtful guidance have significantly contributed to refining my work. Her dedication and practical suggestions were instrumental in overcoming obstacles and ensuring that my thesis achieved the highest standards. Additionally, I owe a special thanks for my family and friends for their perpetual encouragement and patience. Their belief in me and their unwavering support have been my greatest source of strength throughout this process. Finally, I want to express my appreciation to everyone who, through their thoughtful discussions, indirectly supported the completion of this thesis, shared resources, and moral support. This thesis is a reflection of the collective efforts, encouragement, and inspiration I received along the way.

TABLE OF CONTENTS

DECLARATION ..... **iii**

PREFACE AND ACKNOWLEDGEMENT ..... **iv**

TABLE OF CONTENTS ..... **v**

LIST OF FIGURES ..... **ix**

LIST OF TABLES ..... **x**

ABSTRACT **1**

ÖZET **2**

1 INTRODUCTION AND AIM **3**

1.1 QCM, QCM-D and Machine Learning Applications **3**
1.2 Aim of the Study **5**
2 BACKGROUND **8**

2.1 Quartz Crystal Microbalance (QCM) **8**
2.2 Quartz Crystal Microbalance with Dissipation (QCM-D) **11**
2.3 Machine Learning (ML) **13**
2.3.1 Types of machine learning method **14**
2.3.1.1 Unsupervised machine learning **14**
2.3.1.2 Supervised machine learning **14**
2.3.1.3 Semi-supervised machine learning **15**
2.3.1.4 Reinforcement machine learning **15**
2.4 Machine Learning Tasks and Algorithms **15**
2.4.1 Clasification analysis **16**
2.4.1.1 Binary clasification **16**
2.4.1.2 Multiclass clasification **16**
2.4.1.3 Multilabel clasification **17**
2.4.2 Regression analysis **18**
2.4.2.1 Simple and multiple linear regression **18**
2.4.2.2 Polynomial regression **19**
2.4.2.3 LASSO and ridge regression **19**
2.4.3 Cluster analysis **20**
2.4.3.1 Partitioning methods **20**
2.4.3.2 Density-based methods **20**
2.4.3.3 Hierarchical-based methods **21**
2.4.3.4 Grid-based methods **21**
2.4.3.5 Model-based methods **21**
2.4.3.6 K-means clustering **21**
2.4.3.7 Mean-shift clustering **22**
2.4.3.8 GMM clustering **22**

2.4.3.9 Agglomerative hierarchical clustering 23
2.4.4 Dimensionality reduction and feature learning 23
2.4.5 Association rule learning 24
2.4.6 Reinforcement learning 24
2.4.7 Artificial neural network and deep learning 25
3 MATERIALS AND METHODS 28

3.1 Impedance Measurements 28
3.2 Exploratory Data Analysis 30
3.3 Regression Models on Standard Deviations of Dissipation Factors 32
3.4 Two Analytical Models on Glycerol and Bovine Serum Albumin LOD Experiments 33
4 RESULTS 36

4.1 Cluster Centers 36
4.2 Prediction of the Regression Models for Centers 37
4.3 Bulk Liquid Viscosity Measurements 38
4.4 Viscosity, Elasticity, Thickness of BSA Deposits Using 3-Layer Model Fit 41
4.5 Bulk Liquid Viscosity Measurements 42
5 DISCUSSION 44

6 CONCLUSION 46

6.1 Limit of Detection with Multidimensional Impedance Data in QCM 48
6.2 Defining LOD in Multidimensional Sensor Data 49
6.3 Multidimensional Impedance Parameters in QCM 51
6.4 Theoretical Approaches to LOD with Multiple Impedance Parameters 53
7 REFERENCES 58

8 CURRICULUM VITAE 67

ABBREVIATIONS AND SYMBOLS

ANN	Artificial Neural Networks
AV	Averaging
BSA	Bovine Serum Albumin
BSA	Bovine Serum Albumin
CCD	Central Composite Design
DNA	Deoxyribonucleic Acid
DoE	Design Of Experiment
DP	Data Points
EM	Expextation-Maximization
FWHM	Full Width At Half Maximum
GMM	Gaussian Mixture Models
IOT	Internet Of Things
KNN	K-Nearest Neighbors
L	Thickness
LASSO	Least Absolute Shrinkage and Selection Operator
LOD	Limit Of Detection
MDP	Markov Decision Process
ML	Machine Learning
MSE	Mean Squared Error
PSB	Phosphate-Buffered Saline
QCM	Quartz Crystal Microbalance
QCM-D	Quartz Crystal Microbalance With Dissipation Monitoring
RBF	Radial Basis Function
RL	Reinforcement Learning
SLB	Supported Lipid Bilayers
SOM	Self-Organizing Maps
SVM	Support Vector Machines
TSM	Thickness Shear Mode
∆M	Mass Change
f_c	Oscillation Frequency With Quartz Crystal And Film Layer Formed On The Surface
f_0	Quartz Crystal Resonance Frequency

ΔD	The Dissipation Factor Shift
ΔD_n	Dissipation Factor Change For Harmonic n
Δf_n	Frequency Shift For Harmonic n, Calculated As Δf = fc - f0
Δm = C*Δf	Sauerbrey equation
ΔΓ	Change In Dissipation Factor
Δf	fc – f0 = Frequency Change
μ_q	2.947x1011 g/cm.s2 AT Shear Modulus Coefficient For Cut Quartz Crystal
fn	Resonance Frequency For Harmonic n, Defined As fn = n * f1
β0	Intercept Term In Regression Analysis
β1	Regression Coefficient For An Independent Variable
β1x	Coefficient Of x In A Regression Equation
Δ f	f_c- f, For
Δm	-K.Δ f,
Δm	-K.(f_o- f_c)
λn	Wavelength For Harmonic n
A	Quartz Crystal Surface Area
p_q	2.648 g/cm3 Quartz Crystal Density

LIST OF FIGURES

Figure 1. (a) stable wave generated between the disk-shaped quartz crystal and its electrodes for n=1 demonstration (b) qcm's oscillation..... 9
Figure 2. Applying an alternating electric field causes vibrational motion of the crystal at its resonant frequency..... 12
Figure 3. (a) A cnc-machined acrylic flow cell. (b) parameters used to calculate the dissipation factor from a conductance sweep at the fundamental resonance frequency. The inset provides a detailed view of the region outlined by the ellipsoid, illustrating the method used to determine full width half maximum (fwhm) from discrete conductance data. (c) two conductance spectra, each containing different numbers of data points (1000 and 100) and levels of averaging (4 and 1, respectively), are plotted. The left inset offers a close-up of the conductance peak, while the right inset shows a graph of Δd versus time, recorded concurrently in deionized water. 28
Figure 4. Machine learning pipeline 31
Figure 5. Regression results' mse values are plotted. Red, black, blue, green, yellow and brown bars represent first, second and third order polynomial 33
Figure 6. (a) The dissipation factor shifts (Δd) for c1 are plotted for the fundamental resonance frequency and overtones up to the 11th in the glycerol experiments. Only the lod experiment for c1 is included, with the remaining data available in the supplementary information. The shaded areas indicate the standard deviation across three experiments conducted at this concentration. (b) modeled Δd values were used to calculate viscosity for each center, which are then plotted against the known viscosities of each concentration. The error bars represent the standard deviations from the viscosity calculations of three lod experiments per center. 39
Figure 7. (a) The dissipation factor shift (Δd) for c1 is plotted for the fundamental resonance frequency and five overtones in the bsa experiments. Only the lod experiment for c1 is included, with the remaining data presented in the supplementary information. The shaded areas indicate the standard deviation from three experiments conducted at this concentration. (b) the frequency shifts (Δf) from the same experiment are also plotted. 40

LIST OF TABLES

Table 1. A table displaying the seven cluster centers utilized in the glycerol and BSA experiments. 'DP' represents data points, while 'AV' refers to the averaged values. The numbers correspond to the overtone number, with 1 signifying the fundamental resonance frequency. 34
Table 2. Table of predicted and measured standard deviations of dissipation factors at centers. Numbers are multiplied by 107 38
Table 3. Table representing model predicted bsa film thicknesses, elasticities and viscosities for each center. Last column represents the calculated sauerbrey thickness calculated solely from the shifts of the fundamental resonance frequency 41

ABSTRACT

An Investigation into the Impact of Measurement Parameters on the Limit of Detection for the Determination of Viscoelastic Parameters via QCM-D Data Analysis using Machine Learning Model Chaining

This thesis explores the application of Quartz Crystal Microbalance (QCM) technology, focusing on its ability to monitor surface interactions and viscoelastic properties in real-time. As a piezoelectric mass sensing technique, QCM demonstrates high sensitivity in detecting frequency shifts caused by minute mass changes, a principle governed by the Sauerbrey equation. For soft and viscoelastic materials, QCM-D (Quartz Crystal Microbalance with Dissipation) extends QCM's functionality by capturing dissipation shifts (ΔD), which indicate energy loss during oscillation. In this research, QCM-D was utilized to examine liquids like glycerol and Bovine Serum Albumin (BSA). Using a modified 3-layer Kelvin-Voigt model, viscoelastic parameters, including shear modulus, viscosity, and film thickness, were derived. Key findings showed higher harmonic frequencies enhanced sensitivity, revealing viscoelastic properties in adsorbed layers. The adsorption of BSA at low concentrations formed flexible, non-rigid layers with partial surface coverage, highlighting QCM-D's potential in bioanalytical applications. To enhance measurement accuracy, machine learning methods like K-means clustering and regression models were employed. These techniques optimized experimental conditions, enabling precise dissipation factor predictions across harmonics. Random forest and polynomial chain models effectively captured harmonic dependencies, improving the reliability of ΔD measurements. The results underscore QCM-D's versatility in biosensing and materials science, offering a comprehensive tool for studying dynamic surface interactions. This work establishes a foundation for integrating advanced computational techniques with QCM-D, paving the way for innovative applications in complex biological and material systems.

Keywords: BSA Adsorption, Machine Learning, Multi-harmonic Analysis, QCM-D, Quartz Crystal Microbalance, Viscoelasticity

ÖZET

Makine Öğrenmesi Model Zincirleme Kullanılarak QCM-D Veri Analizi Yoluyla Viskoelastik Parametrelerin Belirlenmesinde Ölçüm Parametrelerinin Tespit Sınırı Üzerindeki Etkisinin İncelenmesi

Bu tez, Quartz Kristal Mikrobalans (QCM) teknolojisinin uygulanmasını, özellikle yüzey etkileşimlerini ve viskoelastik özellikleri gerçek zamanlı olarak izleme yeteneğini ele almaktadır. Piezoelektrik kütle algılama tekniği olan QCM, küçük kütle değişikliklerinin neden olduğu frekans kaymalarını tespit etmede yüksek hassasiyet gösterir. Yumuşak ve viskoelastik malzemeler için QCM-D (Dissipasyonlu Quartz Kristal Mikrobalans), salınım sırasında enerji kaybını gösteren dissipasyon kaymalarını (ΔD) yakalayarak QCM'nin işlevselliğini genişletir. Bu çalışmada, QCM-D kullanılarak gliserol ve Bovine Serum Albumin (BSA) gibi sıvılar incelenmiştir. Modifiye edilmiş 3 katmanlı Kelvin-Voigt modeli kullanılarak kayma modülü, viskozite ve film kalınlığı gibi viskoelastik parametreler elde edilmiştir. Temel bulgular, daha yüksek harmonik frekansların hassasiyeti artırarak adsorbe olmuş tabakalardaki viskoelastik özellikleri ortaya çıkardığını göstermiştir. Düşük konsantrasyonlarda BSA adsorpsiyonu, yüzeyi kısmen kaplayan esnek ve rijit olmayan tabakalar oluşturmuş ve QCM-D'nin biyoanalitik uygulamalardaki potansiyelini vurgulamıştır. Ölçüm doğruluğunu artırmak için K-means kümeleme ve regresyon modelleri gibi makine öğrenimi yöntemleri uygulanmıştır. Rastgele orman ve polinom zincir modelleri, harmonik bağımlılıkları etkili bir şekilde yakalayarak ΔD ölçümlerinin güvenilirliğini artırmıştır. Sonuçlar, QCM-D'nin biyosensörler ve malzeme bilimi alanlarındaki çok yönlülüğünü vurgulamakta ve dinamik yüzey etkileşimlerini incelemek için kapsamlı bir araç sunduğunu göstermektedir. Bu çalışma, QCM-D ile ileri hesaplama tekniklerinin entegrasyonu için bir temel oluşturarak karmaşık biyolojik ve malzeme sistemlerinde yenilikçi uygulamalara kapı aralamaktadır.

Anahtar Sözcükler: Deney Tasarımı, Empedans Analizi, Makine Öğrenmesi, Quartz Kristal Mikrobalans, QCMD

1 INTRODUCTION AND AIM

1.1 QCM, QCM-D and Machine Learning Applications

QCM is regarded as among the most sensitive real-time chemical sensors, demonstrating its effectiveness as a mass detection platform in vacuum or gas mediums through the detection of shifts in resonance frequency (1). Linked the shift in resonance frequency to the mass added onto the QCM, based on assumptions primarily relevant to applications in vacuum or gas environments. Advanced thermal deposition and sputtering equipment rely on this foundational relationship to calculate the thickness of deposited thin metal films. However, the assumptions made by Sauerbrey could not explain the significant resonance frequency shift observed when a QCM is submerged in a liquid medium, despite the crystal's inertia being preserved (2). Successfully correlated the surrounding medium's viscosity with its effect on QCM, and independently confirmed Borovikov's findings (3). Both models and subsequent observations elucidated the significant resonance frequency shift occurring without any observable mass change, marking a groundbreaking discovery and providing insight into the underlying mechanisms.

In the following decades, the Kelvin-Voigt and Maxwell models were studied to understand the relationship between higher harmonics with varying penetration depths and the thickness and viscoelastic properties of thin films (4, 5, 6). Additionally, changes in the thickness of deposited thin films manifested as variations in viscoelastic properties, challenging Sauerbrey's rigidity assumption. Collectively, these findings showed that similar frequency shift values could indicate different deposited thicknesses on quartz, reflecting distinct mass changes not explainable solely by Sauerbrey's equation (7). As a result, for a viscoelastic protein layer placed on the QCM surface in a liquid environment, relying solely on frequency shifts to determine the mass of the layer proves to be unfeasible. The dissipation factor, representing energy loss per oscillation cycle, complements the frequency response. It is measurable using the ring-down method or impedance analysis.

The ring-down method involves exciting the resonator with a frequency burst until mechanical oscillations reach their maximum amplitude. Once this maximum is achieved, the burst signal is interrupted, allowing the resonator to dissipate energy until oscillations cease. The piezoelectric resonator's electrical current decays exponentially, and the dissipation factor is calculated using the time constant, which denotes the time required for the signal to decay to 1/e of its initial value (8). Impedance analysis focuses on the frequency domain, utilizing the half-band-half-width (HBH width) of the fundamental resonance frequency's conductance and its harmonics to calculate dissipation factors. The analyzed signal is the Fourier transform of the time-domain signal (9). While both methods mathematically illustrate the same phenomenon, they each offer unique benefits and drawbacks. Careful consideration of their respective advantages is necessary to select the most appropriate approach. The ring-down method is energy-efficient and rapid (10, 11, 12). In contrast, impedance measurements are more accurate (13) and better suited for analyzing both impedance parameters (e.g., absolute impedance, reactance, phase angle) and admittance parameters (e.g., conductance, susceptance, absolute admittance) (14, 15, 16, 17, 18, 19). These features make impedance analysis more comprehensive and suitable for data analysis applications, particularly in monitoring QCM biosensors' electrical behavior under varying conditions.

Machine learning (ML) has emerged as a critical tool in the evolving field of data science. Using statistical methods, ML algorithms reveal key insights into data analysis projects. ML methods are categorized into supervised, semi-supervised, and unsupervised algorithms based on data labeling, with reinforcement learning focusing on ordered data (20, 21). In experimental design, ML is applied based on the type and structure of the data. Design of Experiments (DoE) aims to describe and explain information variation under hypothesized conditions. Historically, DoE has relied on statistical techniques, including sequential designs allowing adaptive learning based on results obtained during experimental stages. DoE has been instrumental in optimizing pharmaceutical products, analytical assays, and green chemistry conditions (22). Recently, ML-enhanced DoE has increased experimental process efficiency, enabling optimal condition discovery and important factor identification. Applications

of ML-assisted DoE span biomaterial and tissue engineering (23), chemical synthesis (24), and social sciences (25).

The limit of detection (LOD) is described as the 3-sigma confidence level, representing the lowest analyte concentration causing a resonance frequency shift (Δf) three times the negative control experiment's standard deviation. By definition, LOD depends on biosensor noise during negative control experiments, previously shown to depend on impedance measurement parameters (26). ML-assisted DoE effectively optimizes QCM impedance measurement parameters, enhancing chemical and biosensor LOD. Optimizing QCM sensor measurement parameters in liquid media is traditionally time-intensive, involving numerous experiments. The authors demonstrated that ML algorithms streamline this optimization, reducing required experiments while maintaining data integrity. Combined supervised and unsupervised ML techniques optimized a representative subset, reducing required experiments significantly and achieving a 12-fold LOD improvement for glycerol concentrations. Further improvement was suggested by selecting measurement parameters from the entire set. In this study, impedance measurement parameters, including data points collected during frequency sweeps and averaging points, were analyzed to determine their effect on dissipation factor shifts for QCM biosensor harmonics (fundamental, 3rd, 5th, 7th, 9th, 11th) at a frequency of fundamental resonance of 10 MHz. Using ML-based optimization, (26) improved LOD by extending frequency shift techniques to multivariable dissipation factor measurements, resulting in a 6-fold LOD difference for glycerol solution viscosity and a 3-fold difference for BSA film thickness variations on QCM gold electrodes.

1.2 Aim of the Study

The goal of this study is to examine how measurement parameters influence the LOD in determining viscoelastic properties by analyzing QCM-D data.. Leveraging advanced computational techniques, including machine learning model chaining, the goal of this study is to enhance the precision of, reliability, and applicability of QCM-D measurements in both scientific and industrial contexts.

Quartz Crystal Microbalance (QCM) and its extended version, QCM-D, are pivotal technologies in material science and biosensor applications. While QCM measures minute mass changes on a quartz crystal surface through frequency shifts, QCM-D adds the capability to measure dissipation shifts (ΔD), providing insights into the viscoelastic properties of soft films and adsorbed biomolecules. Despite its utility, the performance of QCM-D is influenced by various experimental and analytical parameters, especially when analyzing complex systems such as biological materials or multi-layered structures. This research aims to systematically explore the influence of these parameters on the LOD for viscoelastic properties and to establish a framework for optimizing experimental setups and data analysis workflows..

Among the study's central objectives is to improve the accuracy of viscoelastic parameter estimations, such as shear modulus, viscosity, and film thickness, for both bulk fluids and adsorbed layers. These estimations are critical for applications in biomedical engineering, material science, and biosensing. By integrating ML techniques like regression models and clustering algorithms, the study's objective is to offer a robust framework for optimizing QCM-D experiments. This includes evaluating harmonic dependencies, predicting dissipation factor standard deviations, and determining measurement setups that yield the highest sensitivity.

Machine learning plays a transformative role in this study by addressing challenges in high-dimensional datasets and improving the interpretability of QCM-D results. Regression models, including polynomial and random forest approaches, are used to predict viscoelastic parameters across multiple harmonics. Additionally, K-means clustering is employed to identify representative experimental conditions, enabling the study to pinpoint optimal measurement parameters. These methodologies aim to streamline the experimental process and maximize the reliability of QCM-D data, particularly for low-concentration samples where the detection of subtle changes is critical.

A key application explored in this study involves the analysis of glycerol and Bovine Serum Albumin (BSA) solutions. Using a modified three-layer Kelvin-Voigt model, the study examines the relationship between experimental parameters and the accuracy of viscoelastic measurements. For glycerol, bulk liquid viscosity is assessed, while for BSA, the focus is on determining film thickness, elasticity, and viscosity. The findings show the limitations of traditional analytical techniques, including the Sauerbrey equation, for non-rigid films in addition to validating the efficacy of higher harmonic frequencies in enhancing sensitivity.

Ultimately, this research contributes to advancing the field of QCM-D by integrating machine learning with traditional sensing technologies. The findings underscore the potential for QCM-D to serve as a versatile tool for studying dynamic surface interactions in complex biological and material systems. By addressing the interplay between measurement parameters and LOD, this study establishes the foundation for creating automated and intelligent sensing systems, enhancing the applicability of QCM-D in various scientific and industrial domains.

2 BACKGROUND

2.1 Quartz Crystal Microbalance (QCM)

QCM used to detect changes in mass, that transforms to mass changes into an electrical signal. It is highly sensitive and provides a simple output signal. The QCM operates as a piezoelectric resonator, meaning it uses a quartz crystal that changes its physical dimensions (such as buckling) when a force or electrical current is applied. The crystal is specially cut to enhance its piezoelectric properties. QCM is widely used in various sensor applications and is extremely responsive to variations in mass on it’s surface. Its first application dates back to the early 18th century, initially for qualitative measurements. However, after Sauerbrey introduced his equation QCM began to be used for quantitative determinations. (1). During electrochemical processes, the QCM measures mass differences by detecting changes in its resonant frequency. This frequency deviation is directly related to the mass sensitivity, which is remarkably high. The device comprises a quartz crystal with piezoelectric properties, a quartz electrode, an oscillator circuit, and a computer-interfaced frequency counter. When subjected to an alternating high current, the quartz crystal vibrates in mechanical resonance mode. The device's sensitivity is influenced by not only the oscillation frequency but also the overall mass of the crystal.

A quartz crystal resonator is a segment sliced from a single quartz crystal. When a changing potential difference is applied to the electrodes of the crystal, it generates an alternating electric field. The applied potential disrupts the crystal lattice's physical properties, leading to the formation of a standing wave with a specific oscillation frequency within the crystal disk (Figure 1). The oscillation direction is influenced by the alignment of the crystal lattice within the electric field. In thickness shear mode (TSM), the oscillation results in displacement along the surface of the quartz substrate. This oscillation occurs only in the region between the electrodes,

making the inter-electrode area the active region for sensing.

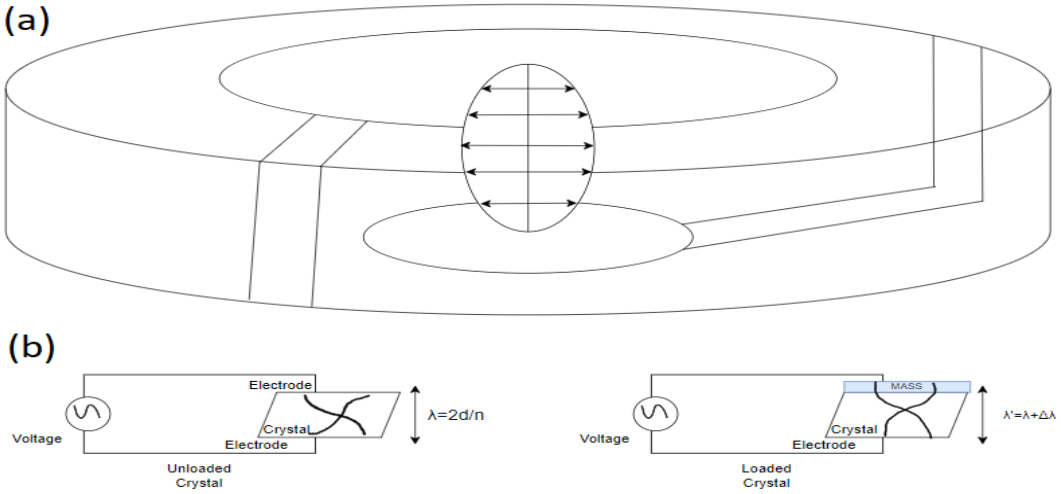

Figure 1. (a) Stable wave generated between the disk-shaped quartz crystal and its electrodes for n=1 demonstration (b) QCM's Oscillation

The standing wave in the inter-electrode region is shown in Figure 1. Gray color electrode on the top surface of the crystal and the electrode shown in blue on the bottom of the crystal surface. In the volume between the two electrodes, the particles are in the +x to -x direction while the wave motion moves in the +y to -y direction, i.e. vertical to the electrode plane. is in line.

$$L = (n.\lambda n/2), (n = 1,2,3, \dots n) \quad (1.1)$$

If λn is subtracted from Equation 1.1;

$$\lambda n = 2L/n \quad (1.2)$$

$$fn = V/\lambda n \ = nV/2L \quad (1.3)$$

Since the fundamental frequency for n=1 is f1, the resonant frequency for n=n is fn = n. f1. Generating a stable wave may not be possible at all frequencies. Generating a standing wave for the vibration frequency to be the same as the frequency of the source. Therefore, the source used for generating waves must be in resonance with the waves to achieve a stable standing wave. You can resonate the vibrator not

only with its fundamental frequency It is also possible to excite with single (1, 3, 5,...) harmonics.

The frequency decreases as the thickness (L) of the crystal increases. For obtain a quartz crystal with high frequency, the crystal must be cut thinner. Therefore, the crystal thickness can be reduced to certain levels. Generally used quartz crystals are 5 MHz, 10 MHz and 30 MHz quartz crystals (27)

Lord Rayleigh was the first to suggest that a change in the mass of a mechanically resonant crystal alters its resonant frequency. As shown in Figure 1, since there is no wave motion outside the region between the electrodes, only the mass changes on the electrode surface affect the vibration of the crystal (28).
The frequency change caused by the mass accumulation on the electrode is obtained from the Sauerbrey equation. This equation was created by G. Sauerbey in 1959 to relate the change of mass bound on piezoelectric crystals to the oscillation frequency (29).

Sauerbrey equation: $\Delta f = f_c - f_0 = \frac{2 f_0^2 \Delta M}{A\sqrt{p_q}\, u_q}$ (1.4)

Sauerbrey demonstrated that the frequency shift of a quartz crystal resonator is directly proportional to the mass added. Sauerbrey's work can be regarded as a novel technique for measuring even the smallest masses. An increase in mass on the quartz crystal surface causes a decrease in the resonant frequency (30). For the Sauerbrey equation to be applicable in quartz crystal sensors, the analyte's mass must be solid, uniformly distributed across the quartz crystal surface, and the frequency change Δf / f should be less than 0.02.

In summary, QCM serves as an analytical technique facilitating in-situ analysis of surface interactions. Its primary advantages in sensing applications include exceptional sensitivity, remarkable stability, fast response times, and cost-effectiveness. Furthermore, it provides label-free detection, making it highly suitable for bio-sensing purposes. Being label-free eliminates the need for time-consuming and

costly labeling steps, as well as any potential interference with the "true" binding process caused by labels. Given the critical importance of bioanalysis, this thesis aims to provide an overview of significant advances and milestones in the field of QCM biosensors based on nanoparticles, highlighting the various approaches reported thus far. Since nanoparticles and biomolecules typically interact at the same nanometer scale, this interdisciplinary approach is expected to contribute to the development of a new field, referred to as immunosensing nanotechnology or nanoimmunosensing. Nanoparticle-based immunosensing systems are proving valuable in advancing novel and versatile protein detection methods (31).

2.2 Quartz Crystal Microbalance with Dissipation (QCM-D)

QCM-D is a label-free technique that has become widely used for studying various biological systems, including cells, vesicles, and proteins, in liquid environments, as well as their interactions with natural and artificial interfaces. This method can qualitatively detect the adsorption and subsequent organization of biomolecules on the sensor surface. (32)

QCM-D has been used to explore several biologically important questions, including enzyme activity, nucleotide strand hybridization, features of supported lipid membranes, DNA-drug interactions, and protein cross-linking, and cell adhesion and spreading. The investigation of vesicle interactions with solid surfaces via QCM-D was pioneered by Kasemo and colleagues in 1998 (33). Since then, vesicle fusion has become a well-established method for creating supported lipid bilayers (SLBs), due to QCM-D's ability to track the entire process in real time, including the quality assessment and kinetic analysis of the formed bilayer (34).

In 1880, Pierre and Jacques Curie discovered that applying mechanical stress to the surface of quartz produces an electrical potential, while applying an electric field to quartz leads to mechanical deformation. This phenomenon, called the piezoelectric effect, takes its name from the Greek words piezein (pressure) and elektro (electric) (35). The term piezoelectricity describes the coupling between the mechanical and

electrical properties of a material. When an alternating voltage near the resonant frequency (f0) of the quartz crystal is applied, mechanical oscillations are induced as demonstrated on Figure 2 (36). Applying a voltage between the electrodes causes the dipoles within the quartz to reorient, resulting in shear deformation of the crystal. This deformation generates an acoustic wave that propagates across the quartz and reflects off its surfaces (37).

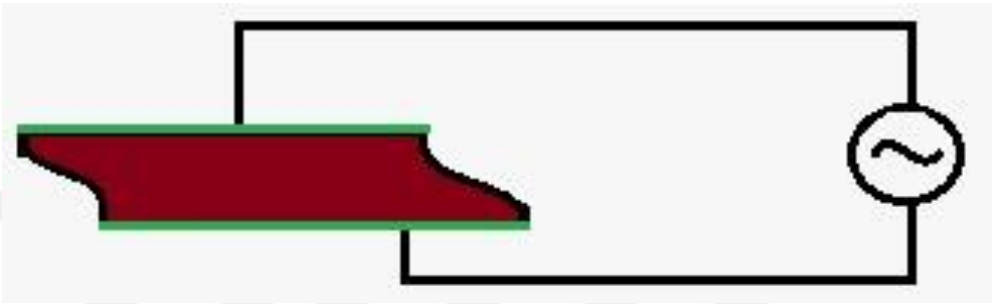

Figure 2. Applying an alternating electric field causes vibrational motion of the crystal at its resonant frequency

Currently, a circular piece of quartz, typically an “AT cut” for its advantageous stability properties, is insert between two metal electrodes in QCMs. In accordance with the fundamental theory of QCM proposed by Sauerbrey in 1959, the frequency change (Δf) of a quartz crystal resonator is linearly related to the adsorbed mass (Δm), as described by the Sauerbrey equation (Equation 1.5) (38).

$\Delta m = C^{*}\Delta f$ (Sauerbrey equation) (1.5)

This relationship applies when the film is uniformly spread over the crystal's active area, is thin relative to the crystal’s thickness, rigid, and is attached to the crystal as though it were an extension of it. (39). The mass sensitivity constant (C) is approximately -17.7 Hz ng/cm² for a 5-MHz crystal, depending on the thickness and density of the quartz (37).

Since QCMs were successfully applied in liquid environments in the 1980s, there has been a significant increase in their use in various biotechnological

applications, particularly in the field of biosensors. However, the presence of viscosity and elasticity effects (damping of the resonator) in liquid phase measurements challenges the assumptions of the Sauerbrey equation. This has prompted researchers to develop new approaches to reinterpret data and characterize mass deposits that exhibit frictional dissipative losses due to their viscoelastic properties. (36).

2.3 Machine Learning (ML)

ML, which is popularly used in every field around us, is encountered in many sectors (banking, entertainment, finance, biotechnology, education, gaming, marketing, etc. ....) ML enables new inferences to be obtained from the proliferation of data and the whole of this data. It creates the basis for producing predictable results with the existing big data. Today's data is complex and voluminous, far exceeding the capacity of human beings to interpret it in a short time. At this stage, machine learning becomes essential, allowing for precise predictive analysis based on complex data (40).

ML involves the creation and utilization of computer systems capable of learning and adjusting without explicit guidance, using algorithms and statistical models to analyze data and recognize patterns. Building upon this foundation, machine learning serves as a critical layer within artificial intelligence, encompassing techniques like neural networks and deep learning to further enhance data-driven decision-making and predictive capabilities.

ML is actually related to deep learning and artificial intelligence. ML is necessary for artificial intelligence to exist. ML provides effective methods for transforming datasets into predictive models, making it a widely used approach in data-driven applications. Machine learning employs statistical techniques to train algorithms for making classifications and predictions. Even as big data expands and grows, it enables machine learning to produce better predictions and more precise results, instead of becoming dysfunctional.

ML was first used in history in 1959 by "Arthur Lee Samuel" while designing the checkers game developed for IBM. He said that machine learning is "an area of research that should enable computers to learn without being explicitly programmed (41).” In 1962, Robert Nealey (Checkers expert) lost the game of checkers against the computer in a match against the IBM7094 computer produced by IBM. Since the program designed for the checkers game had a very small amount of computer memory, a scoring function was designed for the positions of the pieces on the board. The purpose of the scoring function was to try to measure the chances of both sides winning the game. The program used game strategy by using the "minimax algorithm" to choose the next moves. To make the game better, Samuel didn't stop there, he designed a series of mechanisms.

2.3.1 Types of machine learning method

ML algorithms can generally be divided into four categories: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning (42).

2.3.1.1 Unsupervised machine learning

Unsupervised learning focuses on analyzing datasets without labels, functioning autonomously without human input, making it a data-driven approach (43). Unsupervised learning is widely used for extracting generative features, identifying key trends and structures, forming clusters, and conducting exploratory analysis. Key activities in unsupervised learning involve clustering, estimating density, learning features, reducing dimensionality, identifying association rules, and detecting anomalies.

2.3.1.2 Supervised machine learning

This type of machine learning involves developing a function that maps inputs to outputs based on example input-output pairs (43). Supervised learning uses labeled

training data and a set of examples to infer the underlying function. This approach is employed when specific objectives are identified and are to be achieved using a given set of inputs (44), making it a task-driven method. Common tasks in supervised learning encompass "classification," which aims to sort data into specific categories, and "regression," which seeks to fit data to a predictive model. For instance, text classification is a supervised learning application where the objective could be to determine the sentiment or category of a text, such as assigning a class label to a product review.

2.3.1.3 Semi-supervised machine learning

Semi-Supervised learning combines elements of both supervised and unsupervised methods by utilizing both labeled and unlabeled data cPositioned between "learning without supervision" and "learning with supervision," it addresses situations where labeled data are scarce but unlabeled data are abundant (42). The primary objective of a semi-supervised learning model is to achieve better prediction accuracy than a model that relies solely on labeled data.

2.3.1.4 Reinforcement machine learning

Reinforcement learning is a machine learning method inspired by behaviorism, focusing on determining the actions agents should take in an environment to maximize their reward. Because of its broad applicability, this problem is also explored in various fields, including game theory, control theory, operations research, information theory, simulation-based optimization, and statistics (45).

2.4 Machine Learning Tasks and Algorithms

A ML-based predictive model typically involves two phases. In the first phase, historical data is utilized to train the model using ML algorithms, leading to the creation of a predictive model that can be applied for future predictions. In the second

phase, new data is fed into the trained model, which then produces forecast or outcomes based on this data.

2.4.1 Clasification analysis

Classification is a supervised learning technique in ML, commonly used for predictive modeling. In this context, the aim is to predict a class label for a given instance. Mathematically, it involves mapping a function (f) from input variables (X) to output variables (Y), which represent targets, labels, or categories (43). Classification can be used on both structured and unstructured data to identify the category of data points. A typical example is email spam detection, where emails are classified as either "spam" or "not spam." Below is an overview of common classification challenges.

2.4.1.1 Binary clasification

Binary classification refers to tasks that involve two class labels, such as "yes and no" or "true and false" (46). In such tasks, one class often signifies the normal condition, while the other denotes an abnormal condition. For example, in a disease diagnosis system, "healthy" could correspond to the normal state, whereas "diseased" might indicate the abnormal state. Similarly, in spam detection, emails are categorized into two labels: "spam" and "not spam."

2.4.1.2 Multiclass clasification

Multiclass classification involves tasks with more than two class labels(47). In contrast to binary classification, multiclass classification does not adhere to the normal vs. abnormal outcome framework. Rather, examples are sorted into one of multiple predefined classes. For example, in an image recognition task, a model may categorize images into groups like "cat," "dog," "bird," and "fish." Another example is classifying various types of plants in a dataset, where the categories might include "rose," "tulip," "sunflower," and "daisy."

2.4.1.3 Multilabel clasification

Multilabel classification is a key approach where an example can belong to multiple classes or labels. This method extends multiclass classification by allowing hierarchical structuring of classes, where each example may belong to more than one class at different levels, such as in multi-level text classification (48). For example, Google News articles can be categorized under multiple headings like "city name," "technology," or "latest news." Multi-label classification employs sophisticated algorithms that can predict multiple, non-exclusive labels, in contrast to traditional classification tasks where labels are mutually exclusive (49).

Shortly, widely used classification algorithms are tools developed to solve various problems in the field of data analysis and modeling. Each algorithm offers advantages for a specific data set and problem type (50).

- Logistic Regression predicts event probabilities between 0 and 1 for binary classification.
- Support Vector Machines (SVM) separates classes by finding the optimal hyperplane in higher-dimensional space.
- Decision Trees split data through a tree structure to classify it.
- Random Forests use multiple decision trees and majority voting for robust classification.
- Naive Bayes utilizes Bayes' theorem under the assumption of feature independence to enable rapid classification.
- K-Nearest Neighbors (KNN) determines the classification of a data point by identifying the majority class among its k closest neighbors.
- Artificial Neural Networks leverage multi-layered networks to handle complex data and learn deep features.

2.4.2 Regression analysis

Regression analysis is a method employed to examine the relationship between two or more numerical variables. When analyzing just one variable, it is called univariate regression, whereas involving more than one variable is termed multivariate regression analysis. This approach helps determine if a relationship exists between the variables and, if present, the strength of that relationship.

In regression, one variable is considered dependent, while the others are independent. The reasoning is that the variable on the left side of the equation is influenced by the variables on the right, which are not influenced by the other variables. In mathematical terms, this means that these variables have an effect when included in a linear equation, without issues like multicollinearity or sequential dependence.

2.4.2.1 Simple and multiple linear regression

Simple linear regression is a technique used to represent the relationship between two variables. In this method, the linear connection between an independent variable x and a dependent variable y is expressed by the equation below:

$$y = \beta 0 + \beta 1x + \varepsilon \quad (1.6)$$

Here, β0 is the intercept term, β1 is the slope coefficient and ε is the error term. The aim of the model is to determine the values of β0 and β1 that best represent the points in the data set. Simple linear regression is widely used to measure the effect of the independent variable on the dependent variable and to make predictions (51).
Multiple linear regression analyses the effects of multiple independent variables on a dependent variable. This model is expressed by the following equation:

$$y = \beta 0 + \beta 1x1 + \beta 2x2 + \cdots + \beta kxk + \varepsilon \quad (1.7)$$

Here, x1,x2,...,xk are the independent variables, β1,β2,...,βk\ are the coefficients of each of these variables and β0 is the intercept term. ε represents the error term. Multiple linear regression allows to evaluate the effects of multiple factors on the dependent variable simultaneously and provides a more comprehensive analysis. This model is used to understand complex relationships and make predictions in economic, social sciences and other research areas (52).

2.4.2.2 Polynomial regression

Polynomial regression is a type of regression analysis where the relationship between the independent variable x and the dependent variable y is modeled as a polynomial of degree nnn rather than a linear function (49). The polynomial regression equation extends from the linear regression model (which is a polynomial regression of degree 1) and is defined as follows:

$$y = b\,(7)\,0 + b1x + b2x2 + b3x3 + \cdots + bnxn + er^2 \quad (1.8)$$

In this equation, y represents the predicted or target output, b0, b1, ...bn are the regression coefficients, and x denotes the independent or input variable.

2.4.2.3 LASSO and ridge regression

LASSO (Least Absolute Shrinkage and Selection Operator) regression is a technique used in linear regression models for variable selection and controlling model complexity. LASSO adds a penalty to the model's coefficients, encouraging some of them to become exactly zero. This feature helps in automatically removing less important or redundant variables, leading to simpler and more interpretable models. This is especially useful in high-dimensional datasets, where it improves clarity and overall model performance by focusing on the most relevant variables (53).

Ridge regression is a method used in linear regression to address the problem of multicollinearity and improve the model's generalization performance. It works with

adding a penalty to the size of the coefficients, which helps to keep their values small and reduces the risk of overfitting. Ridge regression stabilizes the model by shrinking the coefficients but does not necessarily eliminate any of them. This technique is valuable in situations with many predictors, as it maintains model stability and performance, though it does not perform variable selection as effectively as LASSO (53).

2.4.3 Cluster analysis

Cluster analysis or clustering is the problem of grouping a set of objects. In this problem, objects must be more similar to each other in some way than to elements in other clusters in order to be in the same cluster (cluster). It is one of the main problems of data mining and a widely used technique in statistical data analysis. It is also used in ML, pattern recognition, image analysis, information retrieval, bioinformatics, data compression and computer graphics (54)

2.4.3.1 Partitioning methods

Based on the characteristics and similarities in the data, this clustering technique groups the data into various clusters. Data scientists or analysts usually decide the number of clusters either dynamically or statically, depending on the specific requirements of the application, to implement the clustering methods (55).

2.4.3.2 Density-based methods

These methods identify clusters by finding contiguous regions of high point density separated by regions of low point density. Points not part of a cluster are considered noise. Popular density-based algorithms include DBSCAN and OPTICS. These methods typically struggle with clusters of similar density and high-dimensional data (56).

2.4.3.3 Hierarchical-based methods

Hierarchical clustering constructs a hierarchy of clusters, forming a tree structure. There are two main strategies: (i) Agglomerative, a "bottom-up" approach where each observation starts in its own cluster and pairs of clusters are combined as one moves up the hierarchy, and (ii) Divisive, a "top-down" approach where all observations start in one cluster and are split recursively, moving down the hierarchy (57)

2.4.3.4 Grid-based methods

Suitable for massive datasets, grid-based clustering first summarizes the dataset with a grid representation and then combines grid cells to obtain clusters. Standard algorithms include STING and CLIQUE (58)

2.4.3.5 Model-based methods

These methods can be divided into two types: one using statistical learning and the other based on neural network learning (59). For instance, Gaussian Mixture Models (GMM) exemplify statistical learning (60), while Self-Organizing Maps (SOM) represent neural network learning (61).

2.4.3.6 K-means clustering

K-means clustering is a fast, robust, and simple algorithm providing reliable results when datasets are well-separated. It allocates data points to clusters to minimize the squared distance between data points and the centroid. The K-means algorithm identifies the k number of centroids and assigns each data point to the nearest cluster. However, it is sensitive to outliers due to the impact of extreme values on the mean (62).

2.4.3.7 Mean-shift clustering

Mean-shift clustering is a non-parametric technique that does not require prior knowledge of the number of clusters or constraints on cluster shape. It aims to discover "blobs" in a smooth distribution of samples by updating centroid candidates to be the mean of the points in a given region. Mean-shift clustering is computationally expensive and performs poorly in high-dimensional data (63).

2.4.3.8 GMM clustering

Gaussian Mixture Models (GMMs) are a probabilistic model used for clustering. GMMs assume that all data points are generated by a mixture of a finite number of Gaussian distributions with unknown parameters. The Expectation-Maximization (EM) algorithm is used to estimate these parameters. GMM clustering accounts for uncertainty and returns the likelihood that a data point belongs to a cluster, making it more robust than K-means for non-linear data distributions (64).

Unsupervised learning is a ML method designed to discover patterns and relationships within unlabeled data, which lacks predefined outcomes (65). In other words, unsupervised learning is a technique that identifies underlying structures in data by clustering or grouping similar data points. This method is particularly advantageous in cases where labeling data is impractical or costly. One key application of unsupervised learning is clustering, where the algorithm organizes data by identifying inherent similarities or common features.

GMM is a commonly used method within unsupervised learning, often applied in clustering problems by modeling data as a mixture of multiple Gaussian distributions. GMM assumes that each data point originates from one of several Gaussian distributions, each defined by specific mean and variance values (66). This allows GMM to assign probabilities for data points, which determine their likelihood of belonging to a particular Gaussian component. Such a probabilistic approach makes GMM suitable for data that do not fit neatly into discrete clusters and may overlap.

In constructing a GMM, parameters like the number of Gaussian components are initially specified, and the Expectation-Maximization (EM) algorithm is frequently employed to optimize the model. This algorithm iteratively calculates the probability of each data point belonging to a Gaussian component and updates the parameters to maximize the likelihood (67). GMM has shown notable effectiveness in fields such as image processing, bioinformatics, and speech recognition, where data structures are often complex and difficult to analyze through simpler methods.

For instance, GMM has been used to segment images by identifying regions with similar intensities or textures, a method widely applied in medical imaging to assist in tumor identification and other diagnostic tasks (68). In bioinformatics, GMM aids in clustering genetic data, facilitating the identification of distinct populations or gene expressions. The probabilistic nature of GMM also makes it valuable for capturing uncertainties, providing a robust framework for complex data analysis and aiding in new discoveries and insights.

2.4.3.9 Agglomerative hierarchical clustering

This method organizes objects based on their similarity using a bottom-up approach. Initially, each object is treated as a singleton cluster, and pairs of clusters are gradually merged until all objects form a single large cluster. The outcome is a dendrogram, a tree-like representation of the elements. Agglomerative hierarchical clustering provides a more informative tree-structure hierarchy compared to the unstructured collection of flat clusters returned by K-means, aiding in better decision-making for relevant applications (69).

2.4.4 Dimensionality reduction and feature learning

In ML and data science, dealing with high-dimensional data poses considerable challenges for researchers and application developers. As a result, dimensionality reduction, an unsupervised learning technique, becomes essential. It improves human interpretability, lowers computational costs, and helps prevent overfitting and

redundancy by simplifying models. Dimensionality reduction can be achieved through two main processes: feature selection and feature extraction. The key difference between these two approaches is that feature selection retains a subset of the original features, while feature extraction generates entirely new features (70).

2.4.5 Association rule learning

Association rule learning is a rule-based ML technique used to identify interesting relationships in large datasets through "IF-THEN" statements involving variables. For instance, a common example might state that if a customer purchases a computer or laptop, they are likely to also buy antivirus software simultaneously. This method is widely applied across various fields, including Internet of Things (IoT) services, medical diagnosis, usage behavior analytics, web usage mining, smartphone applications, cybersecurity, and bioinformatics. Unlike sequence mining, association rule learning typically does not consider the order of items within or across transactions (71). The effectiveness of association rules is often assessed using parameters such as 'support' and 'confidence.'

2.4.6 Reinforcement learning

Reinforcement learning (RL) is a type of machine learning that enables an agent to learn through trial and error while interacting with an environment. Unlike supervised learning, which uses fixed examples, RL learns from the agent's experiences and decisions over time. In RL, the problem is often framed as a Markov Decision Process (MDP), which consists of four main components: the Agent, the Environment, Rewards, and the Policy (72). RL techniques are divided into model-based, which use a model to predict the best actions based on outcomes, and model-free, which learn directly from experiences without a model (73). Examples of model-based RL include AlphaZero and AlphaGo (74), while model-free methods include Q-learning and Deep Q Networks (75). The key difference between the two is that model-based RL requires a policy network, whereas model-free RL does not.

2.4.7 Artificial neural network and deep learning

Deep learning belongs to a broader category of machine learning methods that rely on artificial neural networks (ANN) and incorporate representation learning. By utilizing multiple processing layers—such as input, hidden, and output layers—deep learning establishes a computational framework that enables learning directly from data (43). A key benefit of deep learning compared to traditional machine learning techniques is its superior performance in various situations, especially when working with large datasets (76), (77).

Supervised learning is a machine learning method that aims to obtain a function that covers all of these data and results by making use of previously observed and labeled data with known results (78). In short, Supervised Learning is a technique that generates a function from training data. The algorithm of this learning technique works to produce the best matching function between inputs (unlabeled data) and outputs (labeled data). The training data consists of input objects and the desired outputs.
When training the system, inputs and outputs are given for each sample in the data set. In classification studies, the content of the text is used as input data, while the category of the text is used for output. A test data set is used to validate the system. In the validation phase, the learning algorithm assigns a suitable output from the training data to a test data whose category it does not know (79).

ANNs are used as a model framework in supervised learning scenarios, tasked with understanding the relationships between input data and corresponding outputs. These networks are structured with an input layer for data intake, at least one hidden layer for processing information, and an output layer to generate predictions.

The first studies on ANNs started in 1950 when Frank Rosenblatt proposed a computational model based on a simple neuron structure. Later, the first single layer ANN model called perceptron was developed (80).

ANN is a method with a working principle that models the nervous system of the human brain. It is designed to realize the features of the human brain such as creating new information with learning, making discoveries, making interpretations

about events with existing information, making decisions, and establishing relationships between events (81).

ANNs are formed by combining artificial neural cells with each other in various ways and are usually organized in layers. They can be designed with electronic circuits as hardware. ANNs can gather data following a learning process and retain this information through connection weights between neurons. It also includes learning algorithms that allow the ANN weights to be updated in order to achieve the desired goal through the learning process (82).

An ANN is built for a specific purpose and learns from examples, such as humans. Learning takes place in Artificial Neural Networks by updating its structure and weights with repeated inputs (83). There are two types of artificial neural networks: Single Layer (Perceptron) and Multilayer. The Multilayer Perceptron Model consists of an input layer, one or more hidden layers and an output layer. The general structure of the Multilayer Perceptron Model is given below (81).

A study has demonstrated that combining machine learning with QCM can lead to significantly positive outcomes. Using machine learning to optimize impedance measurement parameters in QCM not only improves the LOD, but also streamlines the experimental process by providing a systematic method for parameter selection. This approach demonstrates the potential of integrating advanced computational techniques with traditional sensing technologies to achieve superior performance in chemical and biosensor applications (26).

Moreover, a different study demonstrated that integrating machine learning with QCM technology produces highly positive outcomes. Researchers developed an affordable, Arduino-based QCM system that uses advanced data analysis techniques. This setup performed nearly as well as more expensive commercial systems and proved effective in various applications such as analyzing gas interactions, identifying materials, and monitoring environmental conditions. This shows the promise of combining modern computational methods with traditional sensing technology (84).

Moreover, machine learning algorithms can uncover patterns and correlations in complex datasets that may be missed by traditional analysis methods, resulting in new insights and discoveries. By continuously learning and adapting from new data, these algorithms can improve the robustness and reliability of QCM applications over time.

In addition, combining machine learning with QCM can facilitate automatised and intelligent sensing systems, minimizing the need for manual intervention and reducing operational costs.

3 MATERIALS AND METHODS

3.1 Impedance Measurements

The impedance measurements were performed using a plano-convex AT-cut quartz crystal with a fundamental resonance frequency of approximately 10 MHz. A custom-made acrylic flow cell, fabricated via CNC cutting, with a closed chamber volume of roughly 50 µl, was employed in this study (Figure 3a). Detailed information on the flow cell's fabrication is provided in the supplemental information, Section S.1. For the QCM-D experiments, conductance and phase angle were monitored by connecting the flow cell to the test fixture (16047E) of the impedance analyzer (Keysight, E4990 A).

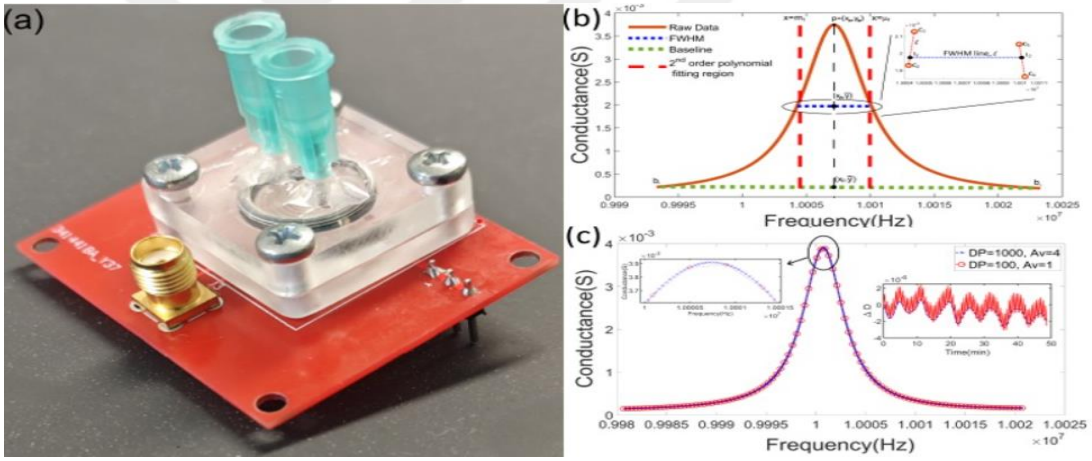

Figure 3. (a) A CNC-machined acrylic flow cell. (b) Parameters used to calculate the dissipation factor from a conductance sweep at the fundamental resonance frequency. The inset provides a detailed view of the region outlined by the ellipsoid, illustrating the method used to determine Full Width Half Maximum (FWHM) from discrete conductance data. (c) Two conductance spectra, each containing different numbers of data points (1000 and 100) and levels of averaging (4 and 1, respectively), are plotted. The left inset offers a close-up of the conductance peak, while the right inset shows a graph of ΔD versus time, recorded concurrently in deionized water.

Two key impedance parameters were optimized in this study: Data Points (DP), representing the number of evenly distributed frequencies for each sweep, and Averaging (Av), which refers to the number of averaged points per frequency. The window size, or the frequency range between the start and stop frequencies of each sweep, was only utilized for Δf computations, as described previously (26), for each harmonic under consideration. However, the focus here will be on ΔD calculations, as follows: 10 different data points, ranging from 100 to 1000 with increments of 100, were selected along with 16 averaging points, starting from 1 to 16, resulting in a total of 160 distinct experimental measurements across all harmonics (n=1, 3, 5, 7, 9, 11).

To calculate the dissipation factor, conductance sweeps were performed with a constant window size for each harmonic, ensuring that both flanks of the peak were included (Figure 3b). Each of the 160 experiments was repeated for the fundamental frequency and five harmonics, with data recorded over a 10-minute period in deionized water at a flow rate of 1 ml/min, maintained by a peristaltic pump (Masterflex L/S HV-77925-10) at room temperature. The baseline conductance across the frequency range was established by defining b1 and b2 as the average of the first and last 3% of the frequency values, respectively (Figure 3b). These points were then used to calculate the baseline equation in the form $y = mbx + kb$, accounting for the slight inclination of the baseline as observed in the supplemental information, Section S.2.

Next, the peak position $p = (xp, yp)$ of the conductance was determined using a peak-finding method previously described (85). This method focused on a specific region of the data, covering conductance values above the average of the maximum and minimum measured conductance values, as defined by the vertical dashed red lines in Figure 3b. The full width at half maximum (FWHM) was then calculated based on the peak position as follows: for $p = (xp, yp)$, the baseline conductance value $\tilde{y} = mbxp + kb$ was computed. The vertical midpoint of yp and $\tilde{y}$, $y = (\tilde{y} + yp)/2$, was then used to define the FWHM line, assumed to be parallel to the baseline (with slope mb). The point-slope form of the FWHM line equation was given by $y = mb(x - xp) + y$. Two points on either side of the FWHM, denoted as c1, c2 (left side) and c3, c4 (right side),

were selected based on the closest measured conductance values. The equations of the lines connecting these points, l1 and l2, were then determined as y = m1x + k1 and y = m2x + k2, respectively. The FWHM value was calculated as the Euclidean distance between the intersection points ι1 and ι2, where l1 and l2 intersect the FWHM line.

Finally, the FWHM was used to determine the dissipation factor D for the harmonic by dividing it by the peak frequency, D = FWHM/fr, where Γ = FWHM/2 is referred to as the HBH-width, and fr is the peak frequency. Standard deviations of the dissipation factors for each harmonic (fundamental and first 5 odd overtones: 3rd through 11th) were computed using this algorithm and recorded for each 10-minute experiment in deionized water. Imperfect conductance peak windows were defined as described in Section S.2 of the supplemental information.

3.2 Exploratory Data Analysis

The purpose of EDA is to design experimental parameters for viscoelastic modeling when at least one harmonic is at LOD. The machine learning (ML) assisted design of experiment (DoE) steps are summarized in Figure 4. The initial data set used in this study consists of 6 tables of size 160x3 corresponding to each odd harmonic. Each table's 160 rows represent experiments for each harmonic, where 3 columns contain 2 impedance measurement parameters (DP and Av) and calculated standard deviations of D. Since viscoelastic modeling requires ΔD measurements for each harmonic simultaneously, any combination of rows from these 6 tables forms a possible experiment setup for LOD experiments (multiplying up to 160^6 experiments).

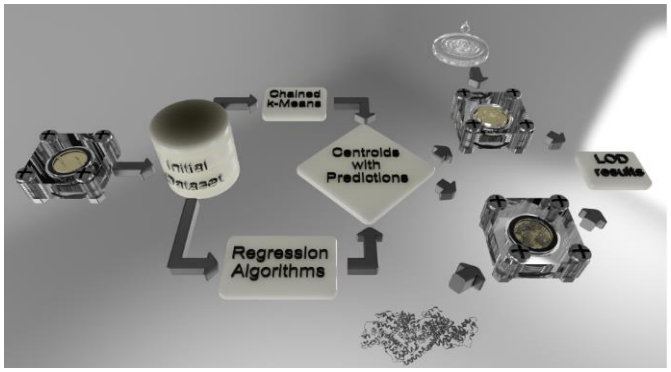

Figure 4. Machine Learning Pipeline

Practically, it is impossible to perform LOD experiments for all such combinations of measurement parameters. Furthermore, even fitting machine learning models to this data set exceeds the computational power available to the researchers (an HPC with dual Intel Xeon Gold 6252 CPUs, dual Nvidia Quadro RTX 6000 GPUs, and 192 GB of RAM was used in this study). To analyze all harmonics together, the measurements were sampled using 6 data points, namely 100, 200, 400, 600, 800, and 1000, along with 5 different averaging values (1, 2, 4, 8, and 16), forming the Reduced Data Set, consisting of a reduced 729 million combinations (30^6).

To determine a representative subset of impedance parameter sets for each harmonic, two distinct methods were utilized for implementing the K-means clustering algorithm, the output of which is represented by cluster centers or centroids. In the first approach, all harmonics were treated simultaneously, leading to the application of the K-Means algorithm once to the Reduced Data Set using the GPU-accelerated faiss module of Python.

In the second approach, two steps of K-Means algorithms were employed. In the initial approach, the K-Means algorithm was applied individually to the six tables, each corresponding to a separate harmonic., as described above, resulting in centroid points for each harmonic. In the second step, combinations of these centroid points were considered as the New Data Set of Centers, and one final K-Means algorithm was applied to this New Data Set.

3.3 Regression Models on Standard Deviations of Dissipation Factors

To continue with the limit of detection experiments on the determined centroid points, the first step of experiments is modeled to obtain values of standard deviations of dissipation factors. Each harmonic is treated separately, using the 6 tables from Section 2.2. Predictions of standard deviations of dissipation factors for each center were performed using 6 different regression models: 1st, 2nd, and 3rd degree polynomials, linear and radial basis function (RBF) kernels of support vector regression, and random forest regression. The best-fitting model was determined by calculating the MSE on the test data set with a test/train ratio of 30/70. Models were trained over 100 epochs for each regression algorithm, and MSEs were averaged to determine the model to be used for predictions.

Two different approaches were taken in these training processes to study the dependency relationship between dissipation factors of each harmonic. First, all regression models were fitted separately for each overtone. Subsequently, chain-type model fitting was applied to explore whether the overtones depend on each other as the overtone number increases. Generally, in multi-output regression analysis, the regression operates independently for each variable. However, in chain models, the output predictions for one variable are used as inputs for predicting the next variable, under the assumption that there might be dependencies between the parameters.

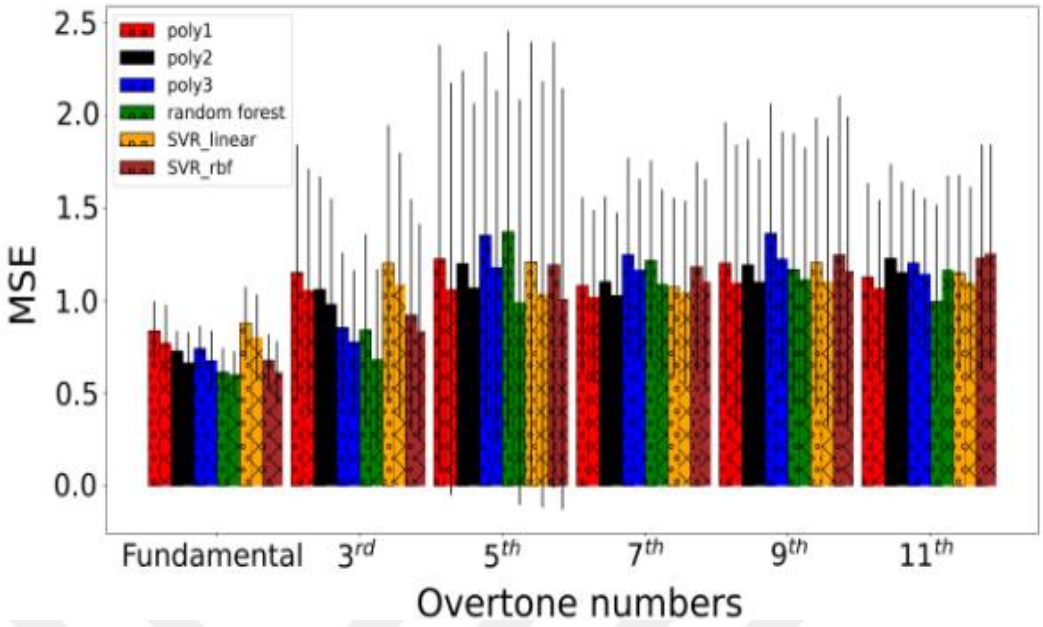

Figure 5. Regression results' MSE values are plotted. Red, black, blue, green, yellow and brown bars represent first, second and third order polynomial

In this study, the regressor chain was applied iteratively. Predictions for the fundamental resonance frequency dissipation factors were used as inputs to predict the 3rd overtone. The cumulative predictions for the fundamental and 3rd overtones were then employed to predict the 5th overtone. This cumulative chain algorithm continued until predictions for the 11th overtone were reached.

3.4 Two Analytical Models on Glycerol and Bovine Serum Albumin LOD Experiments

The viscosity of bulk fluid was determined by fitting the HBH width changes $\Delta\Gamma$ to the generalized form of Kanazawa and Borovikov's equation (86).

$$\frac{\Delta f}{f_f} = \frac{1}{\pi Z_q}\frac{-1+i}{\sqrt{2}}\sqrt{2\pi n f_f p_2}(n' - in'') \quad (2.1)$$

In this equation f_f represents the fundamental resonance frequency Z_q is the acoustic impedance of AT-cut quartz, n is the overtone number, p_2 s the density of the semi-infinite viscoelastic medium, and η' and η" correspond to the storage and loss

moduli, respectively. For Newtonian solutions of glycerol (where η' is constant and η" = 0), $\Delta\Gamma$ and Δf are equal in magnitude but opposite in sign, leading the equation to simplify to

$$\Delta\Gamma n = \frac{1}{Z_q}\sqrt{\frac{nf_f^3 p_2 n'}{\pi}} \tag{2.2}$$

The densities of various concentrations of glycerol were calculated as previously described, and viscosities were determined by fitting $\Delta\Gamma$ alues for each overtone. Glycerol solutions were prepared, and LOD experiments were conducted as described earlier. (26)

Bovine Serum Albumin (BSA) concentrations were prepared by dissolving BSA (Thermo Fisher) in 1xPBS (phosphate-buffered saline) solution. The solutions were made using deionized water (Milli-Q system, Millipore). A modified version of the three-layer Kelvin-Voigt model was employed (87) to predict the viscosity, elasticity, and thickness of BSA films formed on gold electrodes using the following equations ;

$$\Delta \boldsymbol{fn} = \frac{-1}{2\pi p_q h_q}\left(\boldsymbol{h_1 p_1 w_n} - \left(\frac{n_2}{\delta_{2,n}}\right)^2 \frac{2h_1 n_{1,n} w_n^2}{u_{1,n}^2 + (w_n n_{1,n})^2}\right) \tag{2.3}$$

Table 1. A table displaying the seven cluster centers utilized in the glycerol and BSA experiments. 'DP' represents data points, while 'AV' refers to the averaged values. The numbers correspond to the overtone number, with 1 signifying the fundamental resonance frequency.

Cntr	**Dp1**	**Av1**	**Dp3**	**Av3**	**Dp5**	**Av5**	**Dp7**	**Av7**	**Dp9**	**Av9**	**Dp11**	**Av11**
C1	586	8	100	1	435	6	530	9	486	7	489	8
C2	595	6	587	8	547	8	530	9	582	8	554	9
C3	585	9	587	8	436	6	530	9	554	8	100	1
C4	560	14	587	8	547	8	530	9	582	8	554	9

Cntr	Dp1	Av1	Dp3	Av3	Dp5	Av5	Dp7	Av7	Dp9	Av9	Dp11	Av11
C5	586	8	526	7	436	6	100	1	486	7	489	8
C6	587	8	587	8	436	6	530	9	100	1	503	8
C7	586	8	587	8	100	1	530	9	582	8	554	9

The density and thickness of quartz, $p_q \; and \; h_q$, and those of BSA, $h_1 \; and \; p_1$, are used. w_n represents the angular frequency of the nth overtone ($2\pi fn$), η_2 η_2 is the shear viscosity of the bulk liquid, and μ_1,n and η_1,n are the shear modulus and shear viscosity of the BSA layer at the n^{th} overtone. The coupling constants α and β (previously described) are used to fit the calculated (87) and measured ΔD_n and Δf_n. A simplex search method, as explained (88) earlier, was applied using a MATLAB routine to determine the thickness, viscosity, and elasticity of the BSA films (89). The density of BSA in PBS solutions was assumed to be 1.35g/cm^3 and the viscosity of 1xPBS solutions (90) was taken as 1.02mPa·s.

4 RESULTS

4.1 Cluster Centers

On two distinct datasets, the original consisting of 6 tables and a Reduced dataset derived from impedance measurements in deionized water, two variations of the unsupervised K-Means clustering algorithm were applied. The results from the first approach, applied to the Reduced dataset and summarized in Supplemental Information Section S.3, revealed a misrepresentation of very small and very large standard deviations under various conditions. This discrepancy stems from the inherent multinomial distribution nature within the combination set. The resulting cluster centers in this approach were noted to share multiple common data points, with each overtone's averaging pairs analyzed separately, since the entire dataset originates from 729 million distinct conditions, based on a combination of 30 different conditions per overtone.

The cluster centers obtained through the second approach of the K-Means algorithm are displayed in Table I. This approach, too, reflects the inherent repeating structure within the dataset, resulting in similar behavior of the identified cluster centers. Notably, a set of 100 data points, averaged once, which generates the highest standard deviation for each overtone, appears exactly once for 5 harmonics. This indicates that the condition exerts a similar influence across all overtones, except for the fundamental resonance frequency. This particular set was chosen as the cluster center exactly once among the 7 cluster centers. The value k = 7 was selected using the elbow method, which assesses the inertia of the K-Means technique (details are provided in Supplemental Information Section S.3). Interestingly, the computed centroids resemble the axial points in central composite designs (CCD), chosen to maximize variation when constructing statistical models. In statistical terms, a central composite design is used to create a second-order (quadratic) model for a response variable, without the need for a full factorial experiment. The axial points of a CCD are used to examine the extremes of input parameters. The centroids resulting from the K-Means clustering displayed in Table I exhibit similar characteristics to the axial

points of CCD, highlighting that the clustering algorithms successfully captured the variation within the Reduced dataset (91).

4.2 Prediction of the Regression Models for Centers

As illustrated in Figure 5, in 34 out of 36 regression models, the regressor chain models outperformed the separate regressor models, which were fitted to the overtones individually. Although the mean differences of 100 MSEs are smaller than the standard deviations, the histograms of MSEs for most overtones (details, including the Student's t-test results and histograms, can be found in Supplemental Information Section S.4) showed that the regressor chain performed slightly better in predicting the standard deviations of the ΔD values. This outcome aligns with expectations, as Equation (2) suggests that the changes in ΔΓ—and consequently ΔD—increase in proportion to the square root of the harmonic number.

However, any errors originating from uncontrollable factors that have an inertial component, such as temporary local mass or viscoelastic changes in the bulk fluid near the electrode surface (presumed to be within the penetration depth of each harmonic), could contribute to the standard deviation dictated by this relationship. This may introduce a certain degree of independence between the harmonics. By definition, regressor chain models assume an inherent relationship between overtones, which aligns with the underlying theory.

Regarding the best-performing models for each overtone, the random forest regressor chain models were used for predicting the fundamental resonance frequency, the 3rd and 5th overtones. For the 7th overtone, a first-order polynomial regressor chain model was employed, while a second-order polynomial regressor chain model was used for the 9th overtone. For the 11th overtone, separate random forest models were selected as the best-performing models, using MSEs as the evaluation metric. The predicted and measured standard deviations are presented in Table 2. The mean percentage error and standard deviation of the percentage errors in Table 2 is reported

as 10.32 ∓ 7.54%. It's possible that the predicted values are identical for centers with the same or very similar data points and averaging pairs.

4.3 Bulk Liquid Viscosity Measurements

Figure 6a shows a bulk liquid viscosity measurement at the limit of detection (LOD) measured with the 3-sigma method satisfying all the overtones for C1. It should be noted that the time resolution of each harmonic is different as the time resolution is inversely proportional with both the data points and averaging as discussed before (26).

Table 2. Table of predicted and measured standard deviations of dissipation factors at centers. Numbers are multiplied by 10^7

Cntr	**Fund**	**a**	**b**	**c**	**d**	**e**
C1	20.3/19.9	16.0/12.1	2.4/2.1	2.9/2.8	1.9/1.6	1.2/1.3
C2	16.6/19.3	3.6/3.4	2.2/2.0	2.9/2.6	1.9/1.3	2.4/2.2
C3	23.8/25.5	3.6/3.5	2.4/2.6	2.9/3.0	1.9/2.0	5.2/6.9
C4	4.1/4.4	3.6/3.4	2.2/2.3	2.9/2.8	1.9/2.3	2.4/2.2
C5	20.3/19.5	3.3/3.1	2.4/2.7	3.7/3.3	1.9/1.6	1.2/1.3
C6	20.3/22.2	3.6/3.9	2.4/2.3	2.9/2.7	3.9/3.4	1.2/1.0
C7	20.3/21.2	3.6/3.3	4.7/4.7	2.9/2.5	1.9/2.3	2.4/2.2

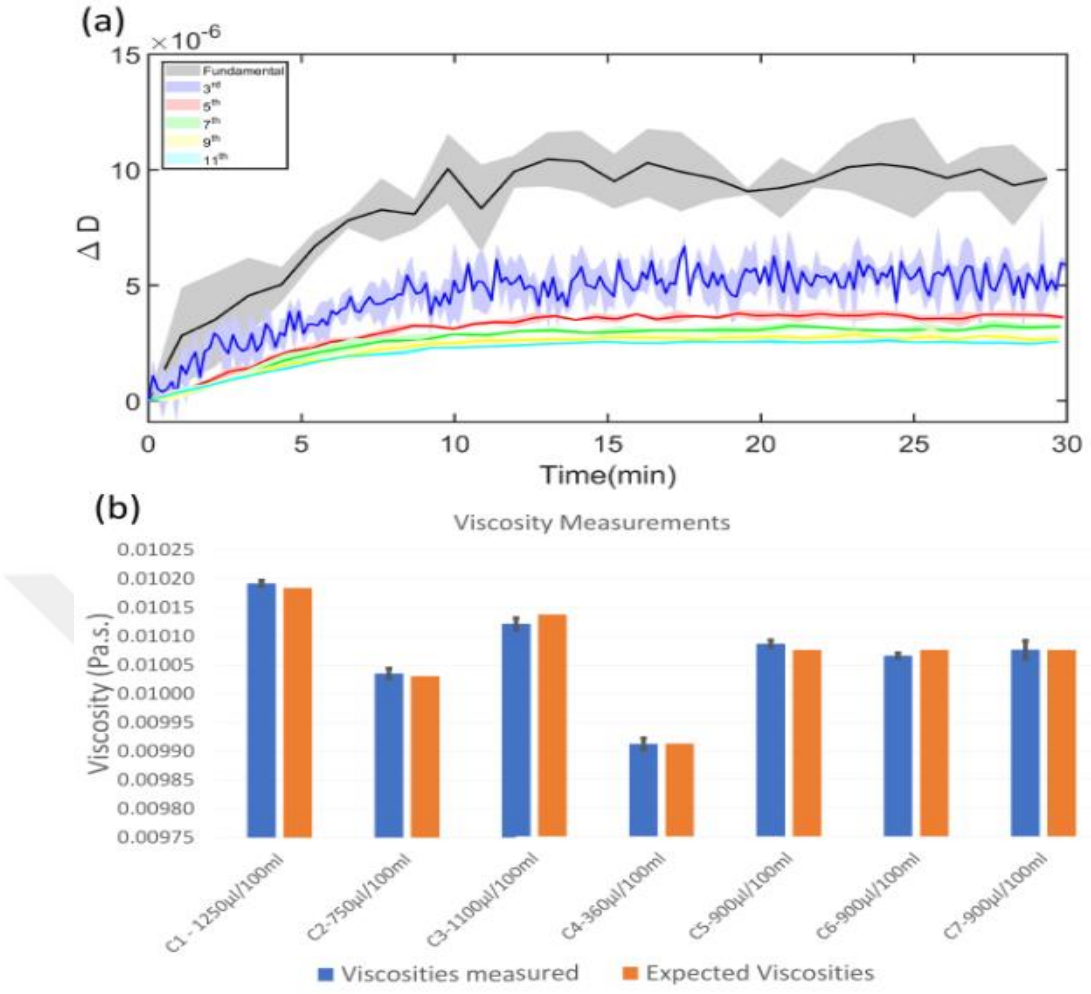

Figure 6. (a) The dissipation factor shifts (ΔD) for C1 are plotted for the fundamental resonance frequency and overtones up to the 11th in the glycerol experiments. Only the LOD experiment for C1 is included, with the remaining data available in the supplementary information. The shaded areas indicate the standard deviation across three experiments conducted at this concentration. (b) Modeled ΔD values were used to calculate viscosity for each center, which are then plotted against the known viscosities of each concentration. The error bars represent the standard deviations from the viscosity calculations of three LOD experiments per center.

Each LOD experiment was repeated 3 times and this was plotted as the shaded region around harmonics data on Figure 6a and as error bars in Figure 6b, which represents the concentration of glycerol in each LOD experiment for each center and predicted concentration calculated from the measured viscosity as explained before

(26). Mean absolute error percentage and standard deviation thereof is ≈ 0.07 ∓ 0.06 which is similar to error measured from Kanazawa's formula using only the fundamental resonance frequency, and Δf instead of ΔD as previously shown (26). This indicated ΔΓ increased with the square root of overtone number and model fitting displayed a similar relative error. However, although the accuracy is increased, the LOD is still affected as seen in Figure 6b, as 3-sigma was satisfied for all the overtones in an experiment, which resulted in ΔD shifts up to more than 21 times for some overtones in worst cases as every overtone had different standard deviations and shifts in ΔΓ followed a proportionality increasing with the square root of the overtone number. This can be seen in Figure 6a as in higher overtones, the shaded error is negligible with respect to the shift in those overtones. (Bulk liquid viscosity measurement at the limit of detection (LOD) measured with the 3-sigma method satisfying all the overtones for the remaining centroids can be found in supplemental information Section S.5.)

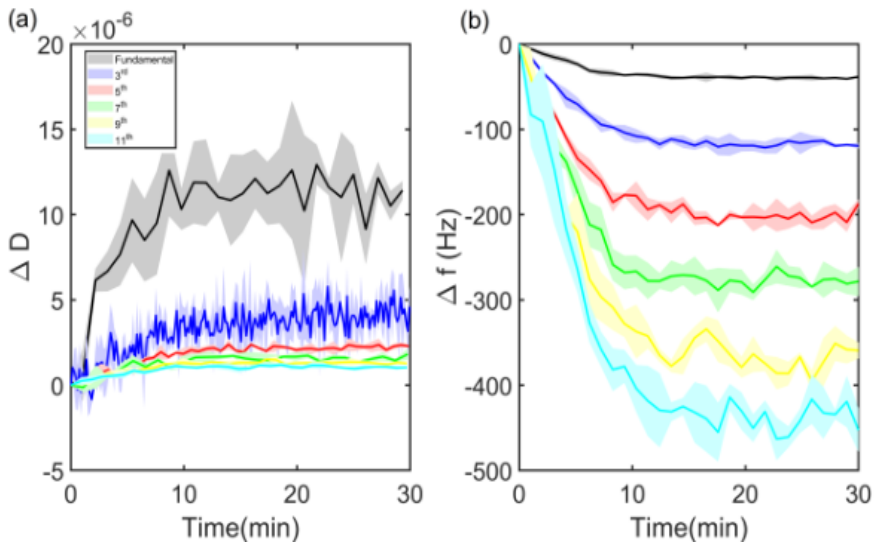

Figure 7. (a) The dissipation factor shift (ΔD) for C1 is plotted for the fundamental resonance frequency and five overtones in the BSA experiments. Only the LOD experiment for C1 is included, with the remaining data presented in the supplementary information. The shaded areas indicate the standard deviation from three experiments conducted at this concentration. (b) The frequency shifts (Δf) from the same experiment are also plotted.

4.4 Viscosity, Elasticity, Thickness of BSA Deposits Using 3-Layer Model Fit

Figure 7a and 7b show the ΔD and Δf vs. time graph of BSA absorption on the gold-coated surface of QCM, for C1, respectively. Shaded regions around each curve shows the standard deviation of 3 experiments for the same concentration. Concentrations of BSA solution for each center are shown in Table 3. Notice that in order to satisfy the 3-sigma LOD condition for the fundamental resonance frequency's ΔD shift, a higher concentration had to be added to the PBS solution, which resulted in a significantly high shift in Δf shifts of all overtones, and even for ΔD shifts of higher overtones.

Table 3. Table representing model predicted bsa film thicknesses, elasticities and viscosities for each center. Last column represents the calculated sauerbrey thickness calculated solely from the shifts of the fundamental resonance frequency

Cntr	BSA thickness (nm)	Elasticity (Pa.)	Viscosity (μPa.s.)	Sauerbrey thickness (nm)	[BSA] (μg/ml)
C1	1.3	223	30.4	2.9	30
C2	0.9	294	34.9	2.0	21
C3	2.6	308	35.8	5.9	62
C4	0.9	284	34.3	2.0	21
C5	0.9	323	36.6	2.0	21
C6	1.1	283	34.3	2.5	26
C7	0.9	241	31.7	2.0	21

Table 3 represents the fitted model results for BSA thickness, elasticity, and viscosity values of BSA layers at the LOD for each center. The ratio of the Sauerbrey thicknesses to fitted thicknesses is 2.26 ∓ 0.04, which indicates very close elasticity and viscosity predictions, as can be seen. Elasticity and viscosity were predicted in the range of 279 ∓ 35 Pa and 34 ∓ 2.2 μPa.s. The viscosity was identical to that estimated by (92). As the measurements were calculated from LODs of ΔD values, the estimated thicknesses from the 3-layer model were in the range of 1.2 ∓ 0.6 nm, which is below the 7.1 nm size of the BSA molecule (93). This indicates that the surface coverage of BSA at such concentrations is below 100%. The difference between the thickness of the BSA layer calculated from the 3-layer model fit and that of Sauerbrey's equation

is mainly due to the fact that the assumptions of Sauerbrey are not met in this case, as even at a sub-molecular size average thickness, the rigidity of the BSA layer resulted in dissipation of the energy of acoustic vibrations, and also an uneven spread of BSA molecules on the surface at these low concentrations. Also, it has to be noted that the LOD of detectable thickness difference between the centers is 3-fold, which was 4 times less than that of the viscosity of the bulk fluid calculated from the fundamental resonance frequency shift only (26). (ΔD and Δf vs. time graph of BSA absorption on the gold-coated surface of QCM, for the remaining centroids can be found in supplemental information Section S.6.)

4.5 Bulk Liquid Viscosity Measurements

The dose-response analysis was conducted using QCM-D measurements under varying concentrations ranging from 0.1 μM to X μM. A total of 12 dose-response graphs were generated to illustrate the sensor's response to the incremental addition of each concentration.

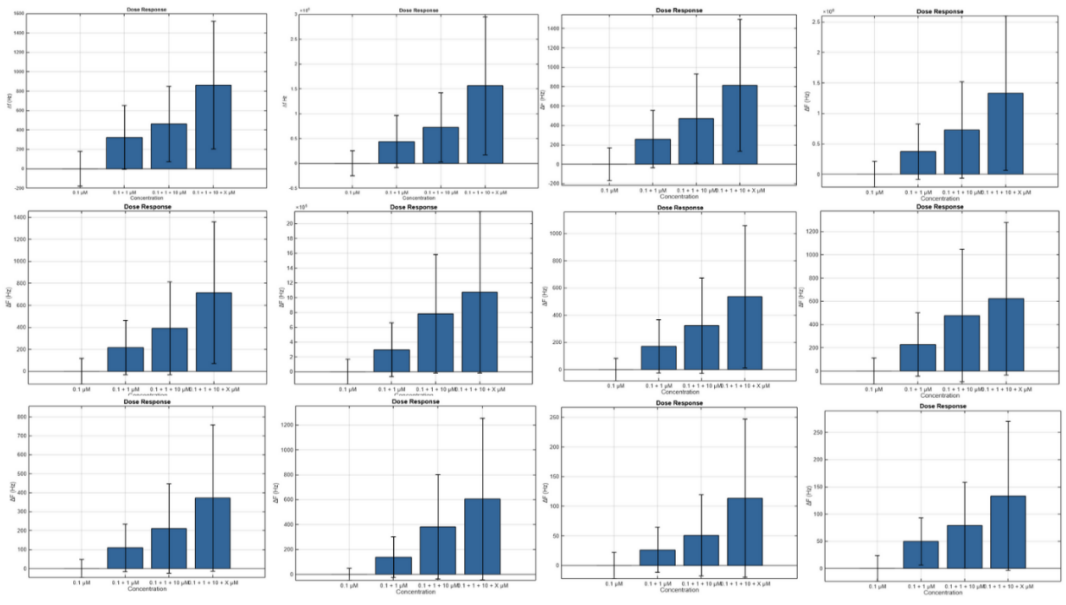

Figure 8. Cumulative Dose-Response Analysis – Frequency Shifts Across Multiple Concentrations

The graphical representation was structured to provide a comprehensive overview of how the frequency shifts (ΔF) and their corresponding standard deviations evolved as the concentration increased cumulatively.

In each graph, the concentration labels were presented in a cumulative manner, such as 0.1 μM, 0.1 + 1 μM, 0.1 + 1 + 10 μM, and 0.1 + 1 + 10 + X μM. This cumulative representation was employed to effectively depict the additive effects of each concentration on the overall frequency response, enabling a more accurate understanding of the sensor's sensitivity and detection capabilities.

The obtained data demonstrate a noticeable increase in frequency shift as higher concentrations were introduced. Additionally, the standard deviations, represented as error bars, indicate increasing variability at higher concentrations, suggesting potential limitations in sensor stability and consistency under elevated concentration levels. This trend provides valuable insights into the sensor's operational range and potential detection limits, highlighting areas for further optimization in sensor calibration and data analysis (26).

Overall, the dose-response analysis offers a comprehensive view of the sensor's behavior under varying concentration levels, allowing for a more informed assessment of its sensitivity and applicability in quantitative detection scenarios. The consolidated graphical representation serves as a visual summary of the observed trends and statistical variations, contributing to the broader interpretation of QCM-D measurement data within the context of this study.

5 DISCUSSION

The findings presented in the results section reveal significant insights into the application of K-Means clustering and regression models for analyzing complex datasets. The clustering outcomes demonstrated the power of the K-Means algorithm in capturing variations, particularly within the Reduced dataset. The resemblance of the cluster centers to the axial points of central composite design (CCD) underscores the algorithm's capability to identify statistical model extremes. This observation aligns with the multinomial distribution nature of the dataset, stemming from overlapping data points and repetitive structures within harmonic combinations. Such behavior emphasizes the need for customized approaches when dealing with datasets featuring strong structural dependencies.

The superiority of regressor chain models over individual regressor models highlights the fundamental relationships between harmonics. The observed increase in dissipation factor (ΔD) with the square root of harmonic numbers, consistent with theoretical expectations, validates the accuracy of the selected regression models. Deviations in standard deviation predictions for certain harmonics likely result from uncontrollable factors such as local mass changes, underscoring the challenges inherent in modeling such systems. However, the use of random forest and polynomial regression models tailored to specific harmonics effectively captured these differences, demonstrating the importance of adaptive model selection.

The application of the 3-sigma method for detecting changes in dissipation factors (ΔD) provided insights into the precision and limitations of bulk liquid viscosity measurements. The proportional increase in ΔΓ changes with the square root of harmonic numbers, particularly evident in higher harmonics, highlighted the trade-off between sensitivity and detection limits. These findings underscore the utility of higher harmonics for detecting subtle changes in bulk liquid properties, despite the challenges posed by variability in standard deviations.

The analysis of BSA adsorption on QCM surfaces offered valuable insights into the interplay between viscosity, elasticity, and thickness at molecular scales. The discrepancies between thicknesses derived from Sauerbrey's equation and those obtained from the 3-layer model fit were attributed to deviations from Sauerbrey's assumptions. The results indicated incomplete surface coverage at low concentrations and uneven molecular distribution. The 3-layer model effectively captured these deviations, providing more accurate predictions of viscoelastic properties even at sub-molecular thickness levels.

Overall, the combined findings of this study emphasize the importance of specialized analytical techniques when working with datasets featuring high dimensionality and inherent distributions. The clustering results and regression models pave the way for improved predictive tools, particularly in applications involving harmonics and material characterization. Future studies could investigate the incorporation of additional machine learning methods or experimental configurations to improve the accuracy of dissipation factor predictions and expand the use of QCM measurements in various fields.

6 CONCLUSION

This study utilized both QCM and QCM-D to monitor mass variations and the viscoelastic properties of adsorbed layers, offering a comprehensive analysis of surface interactions. QCM was effective in detecting mass changes through frequency shifts, while QCM-D provided additional insights into energy dissipation, allowing the characterization of both rigid and viscoelastic layers on the sensor surface.

Initially, QCM was used to measure mass buildup on the quartz surface. As the Sauerbrey equation predicts, an increase in mass led to a corresponding decrease in resonant frequency, demonstrating QCM's high sensitivity to small mass changes. This makes it especially useful in biosensor and electrochemical applications, where even nanogram-level mass changes can significantly impact sensor performance. Impedance measurements using a plano-convex AT-cut quartz crystal supported the theoretical mass-frequency relationship. A custom acrylic flow cell was used to facilitate these measurements, and key parameters such as data points and averaging were optimized across different harmonics.

However, QCM's limitation lies in its assumption that the adsorbed layer is both rigid and uniform. When adsorbed the material is soft or viscoelastic, like proteins or hydrated layers, this assumption no longer holds. To address this, QCM-D was employed. Unlike standard QCM, QCM-D measures not only the frequency shift (Δf) but also dissipation shift (ΔD), providing a more detailed picture of the mechanical properties of the adsorbed layers. QCM-D is particularly valuable for studying soft or hydrated films where energy loss (dissipation) occurs during oscillation. The ΔD measurements indicated that materials like proteins formed flexible, viscoelastic layers, resulting in significant energy dissipation.

The dissipation factor (ΔD) calculated from QCM-D allowed for a more detailed understanding the material properties on the sensor surface. As mass accumulated, ΔD shifts were observed alongside frequency shifts, indicating that the adsorbed layers were not purely rigid but had viscoelastic characteristics. This finding

is crucial for applications in biosensing, where the structural flexibility of adsorbed molecules, such as proteins or lipids, can affect binding dynamics and sensor performance. The use of higher harmonics in QCM-D further increased sensitivity, as higher harmonics are more sensitive to surface interactions and viscoelastic changes, making QCM-D an indispensable tool for complex biological systems.

To improve the precision and reliability of QCM-D measurements, machine learning techniques were employed. Two variations of K-means clustering were applied to optimize experimental parameters, helping to identify conditions that maximized sensitivity and accuracy. The clustering algorithms identified optimal settings for data acquisition, ensuring that both Δf and ΔD shifts were measured with high precision across all harmonics. Regression models, such as random forest and polynomial chain models, were then used to predict dissipation factor changes. These models helped to elucidate the relationships between harmonics, revealing that predictions for lower harmonics could improve the accuracy of measurements at higher harmonics. This dependency is particularly important for multi-harmonic QCM-D measurements, where each harmonic offers unique insights into the material properties being studied.

The viscosity and elasticity of bulk liquids, such as glycerol and Bovine Serum Albumin (BSA), were also explored using QCM-D. The QCM-D system accurately measured the viscosity of glycerol solutions, with results aligning closely with known viscosity values. In the case of BSA adsorption, QCM-D offered in-depth insights into the viscoelastic properties of the protein layers. By using a modified three-layer Kelvin-Voigt model, the thickness, elasticity, and viscosity of the BSA layers were calculated. The ΔD data from QCM-D revealed that BSA forms soft, viscoelastic layers at low concentrations, which did not reach full monolayer formation. This behavior highlights QCM-D's ability to capture the complex structural dynamics of biological molecules on surfaces, offering a more complete analysis than mass measurement alone.

In conclusion, combining QCM with QCM-D significantly enhanced the scope of this study by not only detecting minute mass changes but also revealing the viscoelastic properties of the materials involved. The incorporation of ML techniques further refined the experimental parameters, enhancing the accuracy and consistency of the measurements. QCM-D's exceptional sensitivity and real-time monitoring capabilities, particularly when combined with machine learning, position it as a potent tool for applications in biosensing, material science, and environmental sensing. Future studies could aim on expanding the applications of QCM-D, particularly in dynamic biological systems where viscoelasticity plays a crucial role.

6.1 Limit of Detection with Multidimensional Impedance Data in QCM

The limit of detection (LOD) is a fundamental performance metric for sensors, indicating the smallest analyte quantity that can be reliably distinguished from a blank. In classical (univariate) sensing, LOD is often defined as a signal change equal to three times the root-mean-square (rms) noise of the baseline. In practice, achieving a low LOD requires high signal sensitivity and low noise (and drift) in the measurement. For quartz crystal microbalance (QCM) biosensors – which detect mass changes via shifts in resonant frequency – the LOD can reach the sub-monolayer level (on the order of ~0.05 nm of film thickness for a ~90 mHz frequency shift) under good conditions. However, conventional QCM LOD in liquid biosensing is often limited by noise and baseline drift, making it “good, but not strictly fantastic” compared to some optical methods. For example, surface plasmon resonance (SPR) typically achieves lower detection limits in solution-phase assays than QCM, and surface acoustic wave (SAW) devices can outperform QCM in gas-phase sensitivity.

A key advantage of QCM and related piezoelectric sensors is their ability to provide multidimensional impedance data beyond a single frequency shift. Modern QCM instruments (often termed QCM-D, for QCM with dissipation monitoring) measure energy dissipation or damping alongside the frequency change. Impedance analysis of the QCM’s electrical response (using equivalent-circuit models like the Butterworth–Van Dyke) can extract multiple parameters – e.g. motional resistance

(R□), motional capacitance (C□), motional inductance (L□), and phase or conductance spectra – yielding a multi-parameter output for each sensing event. These additional channels of information can help distinguish between different physical effects (mass loading vs. viscoelastic changes) and potentially improve the confidence in detection. This report reviews how LOD is defined and theoretically modeled using such multi-parameter impedance data, with an emphasis on QCM-based biosensors. We survey relevant theoretical frameworks (noise models, equivalent circuits, multivariate analysis) and highlight case studies (including biomedical diagnostics) from both the scientific literature and patents.

6.2 Defining LOD in Multidimensional Sensor Data

LOD Criteria in Univariate vs. Multivariate Systems: In traditional univariate measurements, the LOD is often determined by the 3σ criterion – i.e. a signal change equal to three times the standard deviation of noise in the blank signal. This ensures a 99% confidence that the signal is real. Formally, IUPAC and other bodies define LOD as the analyte concentration yielding a signal significantly different from zero with a given confidence (often 95% or 99%). The simplest formula (for a linear calibration) is:

$$\mathrm{LOD}_{\mathrm{univariate}} = \frac{3 \cdot s_{\mathrm{noise}}}{\mathrm{slope}} \tag{6.1}$$

snoise is the noise (standard deviation)
slope is the sensitivity (signal change per analyte unit).

However, when a sensor provides multiple signals or a sensor array produces a vector output, defining LOD becomes more complex. There is currently no universally accepted standard for LOD in multivariate sensor systems. Researchers have proposed various approaches, such as computing the LOD from the residual error of multivariate calibration models or examining when the sensor response becomes repeatably distinguishable from noise in all channels. For instance, Oleneva et al. (2019) suggested a visually intuitive method for electronic tongue sensor arrays:

incrementally increase analyte concentration and identify the point at which the mean relative error (MRE) of predictions stabilizes, indicating the signal is reliably above the noise/background. This concentration is taken as the LOD for the multi-sensor system. Their procedure provided a consistent way to compare multi-sensor performance.

In a multivariate context, one can also consider the multivariate noise (covariance of baseline signals) and define LOD based on a combined response. One theoretical approach is to use concepts like the Mahalanobis distance or principal component projections of the multi-parameter signal. Indeed, applying principal component analysis (PCA) or other data fusion techniques can enhance detection: by extracting the relevant signal components from multiple channels, the effective signal-to-noise ratio can improve. A PCA-based analysis of multi-overtone QCM data was shown to improve sensitivity and LOD compared to analyzing frequency shifts alone. In effect, the multivariate LOD might be defined by a threshold in a combined response space (for example, a PCA score or a discriminant function) that corresponds to the smallest detectable analyte presence with 95% confidence.

Theoretical modeling of LOD inevitably requires modeling of noise in the sensor signals. For impedance-based sensors like QCM, noise can arise from electronic circuits, thermal fluctuations, and mechanical instabilities. Notably, frequency stability of the resonator is crucial. Techniques from frequency metrology (e.g. Allan variance and Hadamard variance) have been employed to characterize QCM frequency noise and drift. The Allan deviation $\sigma(\tau)$ describes the frequency fluctuation as a function of averaging time τ, and is related to the resonator's quality factor (Q-factor). One patent suggests an approximate relation $\sigma \approx 10^{(-7)}/Q$ for AT-cut quartz at 1 s averaging. Using this, the detection limit in frequency can be estimated as:

$$\Delta f_{\min}(\tau) \approx \sigma(\tau) \cdot f_0 \tag{6.2}$$

where f_0 is the resonant frequency. For example, a QCM with Q ~ 50,000 at 5 MHz has $\sigma \sim 2\times10^{(-12)}$ (fractional) over 1 s, giving a frequency noise on the order of a few mHz. By converting this frequency resolution to mass via the Sauerbrey

coefficient, one can obtain the mass LOD. In the cited patent, the authors use the product of Allan deviation and resonant frequency to get Δf, then divide by the Sauerbrey mass sensitivity constant K to get a mass resolution. Such theoretical calculations predict femtogram-level mass detection capability when Q is extremely high and noise is correspondingly low. In practice, achieving that requires careful control of the system (temperature, vibrations, etc.), or innovative designs as discussed.

6.3 Multidimensional Impedance Parameters in QCM

QCM and the Butterworth–Van Dyke (BVD) Model: A QCM can be modeled electrically by an equivalent circuit (the BVD model), which includes a motional branch (L□, C□, R□) representing the oscillating crystal and any coupled load, in parallel with a static capacitance C_0 of the electrodes. When a QCM is used in a fluid for biosensing, multiple physical properties influence these circuit elements. Added mass (e.g. an adsorbed biomolecular layer) primarily shifts the resonant frequency (related to L□ and C□), while changes in viscoelastic or dissipative loading (e.g. a soft film or viscous adlayer) increase the motional resistance R□ and bandwidth (inverse of the Q-factor). Modern impedance analysis instruments can sweep a range of frequencies around resonance and fit the response to the BVD model, thus extracting multi-dimensional data per measurement: Δf (frequency shift), ΔR□ (change in motional resistance, related to dissipation), ΔC□, ΔL□, and even ΔC_0 . In practice, not all of these parameters are equally informative for biosensing. A study by Su et al. on a QCM immunosensor illustrates that frequency shift and motional resistance are the dominant signals, whereas changes in C□, L□, or C_0 were negligible for their biorecognition scenario. The resonant frequency (or its shift Δf) is directly tied to the mass loading via the Sauerbrey equation (for rigid masses) or more complex viscoelastic models, and the motional resistance ΔR□ (or related dissipation factor D = 1/Q) captures energy losses due to the viscoelasticity of the film or medium.

Dissipation and Multiharmonic Data: QCM-D instruments measure the frequency and dissipation at multiple harmonic overtones of the crystal (e.g. 1st, 3rd,

5th overtone frequencies). These provide an even richer multidimensional dataset: each overtone may respond slightly differently to the same surface perturbation, which can be used to model the viscoelastic properties of the film. For instance, a purely elastic (rigid) mass loading will cause proportional frequency shifts on all overtones and minimal dissipation change, whereas a soft viscoelastic layer will cause frequency shifts that depend on overtone number and significant dissipation increases. Using subsets vs. full sets of parameters is a key consideration. In some applications, monitoring just Δf is sufficient for LOD (especially if the film is rigid and noise is low). In others, including ΔD or ΔR□ can greatly enhance detection. Notably, if an analyte causes a very tiny mass change but a disproportionate change in viscoelastic damping, the frequency signal might be under the noise while the dissipation signal is clear. In such cases, dissipation can extend the effective LOD. A classic example is detection of whole bacterial cells: these are large, soft particles that, when captured on a QCM surface, may not produce a large frequency drop (because a single or few cells have small mass relative to the crystal) but can induce a measurable increase in energy dissipation (due to viscoelastic binding and interfacial slip). Su and Li (2005) demonstrated this with a Salmonella immunosensor: using magnetic beads to concentrate the bacteria on the QCM, they achieved a LOD of ~10^2 cells/mL based on the motional resistance change (ΔR□), whereas the frequency shift at that low concentration was not detectable (Δf did not correlate with cell count in that range). This underscores how multi-parameter impedance data (here, R□ vs. f) can push the detection limit lower than one parameter alone.

Beyond frequency and dissipation, researchers have explored alternative impedance metrics. The phase angle of the impedance at or near resonance can serve as a sensitive indicator of shifts. For example, measuring the phase change at a fixed drive frequency (instead of tracking frequency in an oscillator circuit) is another way to transduce a small resonance shift. Arnau and coworkers introduced a phase-shift measurement technique on high-frequency QCMs, showing it could dramatically improve sensitivity. The peak conductance (G) or admittance magnitude at resonance is another parameter: as a film loads the crystal, not only does the resonant frequency shift, but the peak conductance typically drops (due to increased damping R□).

Tracking the drop in conductance or increase in R□ could be used as a detection metric on its own. In effect, all these parameters are interrelated (they stem from the same resonance phenomenon), but using them in combination or choosing the most sensitive one for a given application is crucial for optimizing LOD.

6.4 Theoretical Approaches to LOD with Multiple Impedance Parameters

Noise Modeling and Resolution Limits: The theoretical minimum detectable change in each QCM parameter can be estimated by noise analysis. For frequency, as discussed, one can use Allan or Hadamard variances to account for noise and drift. For dissipation or R□, similar analysis applies: the noise floor on R□ (or bandwidth) is related to electronic noise and mechanical loss fluctuations. In an impedance measurement, R□ is obtained from the width of the resonance curve – theoretically, a high Q (narrow) resonance allows finer resolution in determining R□ changes, but also a high Q implies low baseline dissipation (which might be near the noise floor of the instrument). When multiple parameters are considered together, one simple theoretical criterion for detection is to require that at least one of the parameters exceeds its 3σ noise threshold. A more rigorous approach is to combine them: for example, one could form a joint detection statistic that is a weighted sum of the normalized changes in each parameter, and require that this multi-parameter statistic exceed a certain confidence limit. This would be analogous to multivariate hypothesis testing in statistics. While a general formula for multivariate LOD is not standardized, recent work in chemometrics suggests using the concept of "multivariate signal-to-noise" or performing inverse modeling (multivariate calibration) then converting the univariate LOD through the model's error propagation.

LOD Enhancement Strategies: Several theoretical and experimental strategies have emerged to improve LOD using multi-dimensional data. Optimizing Measurement Conditions: By adjusting how impedance data is collected (frequency range, number of points, overtone selection, etc.), one can enhance sensitivity to the target analyte. In a previous work, our team employed a machine-learning optimization of QCM impedance measurement parameters and found that the apparent LOD in

QCM-D experiments could be improved by an order of magnitude (~12×) by selecting optimal measurement settings. In their work, different choices of impedance spectral range and data features led to variations in the derived viscoelastic parameters; the bulk fluid viscosity measurement's LOD varied up to 6-fold, and a biomolecular film's thickness LOD varied ~3-fold depending on the impedance sampling strategy used. This indicates that the way we use the multi-dimensional data (which frequencies or which combination of parameters) can strongly affect detection limits.

High Fundamental Frequencies: Using a higher frequency resonator increases mass sensitivity (Sauerbrey's equation shows $\Delta f \propto -f_0$ for a given Δm). The trade-off is often increased damping in liquids, but if manageable, a higher f_0 can lower LOD. Ogi et al. demonstrated a 170 MHz electrodeless QCM which achieved a LOD of 0.5 pM for a protein (IgG) – about three orders of magnitude better mass sensitivity than a conventional 5 MHz crystalpubmed.ncbi.nlm.nih.gov. Theoretical modeling in that study predicted a fundamental limit: at sufficiently high frequency, the QCM response would eventually "break down" when the oscillation frequency approaches the intrinsic mechanical relaxation frequency of the binding interactionspubmed.ncbi.nlm.nih.gov. In other words, when the analyte-binding behaves like a spring–mass system with its own resonance (e.g., an antibody–antigen bond having a certain compliance and mass), driving the sensor much faster (one order above that resonance) decouples the mass from the sensor. This insight sets a theoretical frequency limit beyond which increasing f_0 no longer improves (and can degrade) the detection of that particular analytepubmed.ncbi.nlm.nih.gov.

Phase/Delay Monitoring: As noted, measuring phase shift at a fixed frequency is a form of multi-parameter usage (the fixed frequency is offset from the natural resonance, so phase is a sensitive function of resonance frequency and damping). In a 2015 study, March et al. used a 100 MHz QCM with phase monitoring to detect pesticides in a competitive immunoassay. They reported a two-order-of-magnitude improvement in LOD compared to a standard 9 MHz QCM oscillator: for carbaryl pesticide, the LOD improved from 11 μg/L at 9 MHz to 0.14 μg/L at 100 MHzawsensors.com. Notably, the combination of higher frequency and the new phase

measurement scheme led to this dramatic enhancement. The high-frequency QCM outperformed even SPR in that case, approaching the sensitivity of specialized lab assaysawsensors.com. The theoretical basis is that at a fixed frequency on the steep slope of the resonance phase curve, a tiny resonance shift produces a relatively large phase change, which can be measured with low noise – effectively boosting the signal-to-noise.

Multi-mode or Multi-sensor Data Fusion: Some advanced sensors use more than one sensing mode in tandem (for example, a coupled acoustic-electrochemical sensor). An electrochemical QCM-D (EQCM-D) can measure both the acoustic signals (f & D) and electrochemical impedance or current, providing orthogonal data about an analyte (mass change and charge transfer). Combining these via multivariate analysis could improve detection of complex analytes (like differentiating binding vs. electrocatalytic signals). Even within pure QCM, one can have an array of crystals coated with different receptors (like an "electronic nose/tongue" using QCMs). In such cases, multivariate analysis of the array's frequency shifts can greatly improve specificity and potentially the collective LOD (by reducing false positives). While this strays from a single QCM device, it aligns with the idea of multidimensional data. The general point is that multidimensional impedance measurements open up opportunities for algorithmic improvements (machine learning, PCA, etc.) to tease out smaller signals than a single-channel measurement could.

Patent Innovations: Patents in this space often target ultra-sensitive detection for medical diagnostics. One notable example is a patented high-Q QCM femtogram sensor for detecting biomarkers like troponin. The device uses an array of five quartz disks with varying electrode sizes to modulate their Q-factors and responses. The system measures multiple parameters from each disk – including the resonant frequency shift, Q-factor (from which dissipation is derived), and impedance – and then applies an Allan variance analysis to isolate true frequency changes. By leveraging the redundancy and consistency across multiple sensors and parameters, the design minimizes the impact of noise and drift, theoretically allowing detection of masses on the order of 10^(-15) g (femtograms). While primarily an engineering

advance, the underlying theory treats the ensemble of readings as a combined indicator of a true signal, filtering out uncorrelated noise. Another patent approach is the use of novel interrogation methods (e.g., a wireless QCM-D excitation and transient response measurement) to get richer data from each sensor oscillation, potentially improving LOD by capturing the full transient decay (which contains frequency and dissipation information in one shot). These innovations illustrate how multi-parameter data and theoretical noise models guide the development of next-generation biosensors with ever lower detection limits.

Multidimensional impedance readouts from biosensors like QCM have enabled more robust and sensitive definitions of detection limits than were possible with single-parameter measurements. Theoretically, the LOD in such systems can be modeled by extending classical noise analysis to multiple degrees of freedom – whether by combining noise contributions for a joint detection criterion or by identifying which parameter best signals the presence of analyte. Frequency shift (Δf) remains the primary indicator of mass changes, but dissipation and related parameters (R□, Q-factor, phase lag) provide invaluable complementary information. They can reveal subtle interactions (e.g. a viscoelastic binding or a sparse large particle attachment) that would be missed by frequency alone. Consequently, a subset or full set of impedance parameters can be used to define detection: for example, one might define that an analyte is "detected" if either Δf or ΔR□ exceeds its noise threshold, or more elegantly, if a combination of Δf and ΔR□ falls outside the distribution of baseline fluctuations with a given confidence.

From a design perspective, theoretical frameworks such as equivalent-circuit modeling, noise spectral density analysis (Allan/Hadamard variances), and multivariate calibration models guide how we utilize the rich data. They help predict how changes in one parameter affect the others and where the fundamental limits lie (e.g. the point of diminishing returns with higher frequency or higher Q). We have seen in case studies that incorporating more information (multiple harmonics, multiple parameters, or multiple sensors) generally pushes LOD lower – for instance, reaching picogram mass or sub-picomolar concentration detection. However, each added

dimension also requires careful calibration and sometimes complex data processing (such as machine learning or PCA) to fully exploit it.

In summary, defining LOD in multidimensional impedance biosensors is an evolving practice that merges classical detection theory with modern data analytics. Peer-reviewed studies have demonstrated clear benefits of multi-parameter approaches in biosensing applications ranging from pathogen detection to immunoassays. Meanwhile, patents reveal a drive toward practical implementations that achieve ultra-low LODs by combining high-performance hardware with multi-channel data strategies. As the field advances, we expect more standardized methods to emerge for quantifying LOD in multivariate sensor systems, enabling fair comparison and aiding the design of the next generation of highly sensitive biosensors.

7 REFERENCES

1. Sauerbrey G. Verwendung von schwingquarzen zur wägung dünner schichten und zur mikrowägung. Zeitschrift für Physik 1959; 155(2): 206–222.
2. Borovikov A. Measurement of viscosity of media by means of shear vibration of plane piezoresonators. Instrum Exp Tech 1976; 19(1): 223-224.
3. Kanazawa KK, Gordon JG. The oscillation frequency of a quartz resonator in contact with liquid. Analytica Chimica Acta [Internet]. 1985 Jan 1;175:99–105. Available from: https://doi.org/10.1016/s0003-2670(00)82721-x
4. Voinova MV, Rodahl M, Jonson M, Kasemo B. Viscoelastic acoustic response of layered polymer films at Fluid-Solid Interfaces: Continuum Mechanics Approach Physica Scripta [Internet]. 1999 May1 ;59(5):391–6. Available from: https://doi.org/10.1238/physica.regular.059a00391
5. Voinova MV, Jonson M, Kasemo B. Missing mass' effect in biosensor's qcm applications. Biosensors and Bioelectronics, 2002; 17(10);835–841. Available from: https://www.sciencedirect. com/science/article/pii/S0956566302000507
6. Van Der Westen R, Sharma PK, De Raedt H, Vermue I, Van Der Mei HC, Busscher HJ. Elastic and viscous bond components in the adhesion of colloidal particles and fibrillated streptococci to QCM-D crystal surfaces with different hydrophobicities using Kelvin–Voigt and Maxwell models. Physical Chemistry Chemical Physics [Internet]. 2017 Jan 1;19(37):25391–400. Available from: https://doi.org/10.1039/c7cp04676f
7. Rodahl M, Kasemo B. On the measurement of thin liquid overlayers with the quartz-crystal microbalance. Sensors and Actuators a Physical [Internet]. 1996 Jun 1;54(1–3):448–56. Available from: https://doi.org/10.1016/s0924-4247(97)80002-7
8. Banica FG, Urban GA. Chemical sensors and biosensors: Fundamentals and applications. Analytical and Bioanalytical Chemistry [Internet]. 2013 Mar 28;405(16):5365–6. Available from: https://doi.org/10.1007/s00216-013-6870-9
9. Johannsmann D. The quartz crystal microbalance in soft matter research [Internet]. Soft and biological matter. 2014. Available from: https://doi.org/10.1007/978-3-319-07836-6

10. Arnau A. A review of interface electronic systems for AT-cut quartz crystal microbalance applications in liquids. Sensors. 2008: 8(1); 370-411.
11. Zeng Z, Pertijs MP, Karabacak DM. An energy-efficient readout circuit for resonant sensors based on ring-down measurement. Review of Scientific Instruments [Internet]. 2013 Feb 1;84(2). Available from: https://doi.org/10.1063/1.4792396
12. Alassi A, Benammar M, Brett D. Quartz Crystal Microbalance Electronic Interfacing Systems: A review. Sensors [Internet]. 2017 Dec 5;17(12):2799. Available from: https://www.mdpi.com/1424-8220/17/12/2799
13. Park JY, Choi JW. Review—Electronic Circuit Systems for piezoelectric resonance Sensors. Journal of the Electrochemical Society [Internet]. 2020 Jan 2;167(3):037560. Available from: https://doi.org/10.1149/1945-7111/ab6cf7
14. Su XL, Li Y. A QCM immunosensor for Salmonella detection with simultaneous measurements of resonant frequency and motional resistance. Biosensors and Bioelectronics [Internet]. 2005 Mar 19;21(6):840–8. Available from: https://doi.org/10.1016/j.bios.2005.01.021
15. Itoh A, Ichihashi M. Separate measurement of the density and viscosity of a liquid using a quartz crystal microbalance based on admittance analysis (QCM-A). Measurement Science and Technology [Internet]. 2010 Dec 13;22(1):015402. Available from: https://doi.org/10.1088/0957-0233/22/1/015402
16. Kasper M, Traxler L, Salopek J, Grabmayr H, Ebner A, Kienberger F. Broadband 120 MHz Impedance Quartz Crystal Microbalance (QCM) with Calibrated Resistance and Quantitative Dissipation for Biosensing Measurements at Higher Harmonic Frequencies. Biosensors [Internet]. 2016 May 25;6(2):23. Available from: https://doi.org/10.3390/bios6020023
17. Dai J, Zhao H, Lin X, Liu S, Fei T, Zhang T. Design strategy for ultrafast-response humidity sensors based on gel polymer electrolytes and application for detecting respiration. Sensors and Actuators B Chemical [Internet]. 2019 Nov 2;304:127270. Available from: https://doi.org/10.1016/j.snb.2019.127270
18. Li D, Xie Z, Qu M, Zhang Q, Fu Y, Xie J. Virtual sensor array based on Butterworth–Van Dyke equivalent model of QCM for selective detection of volatile organic compounds. ACS Applied Materials & Interfaces [Internet]. 2021 Sep 21;13(39):47043–51. Available from: https://doi.org/10.1021/acsami.1c13046

19. Yu X, Chen X, Yu X, Chen X, Ding X. A Quartz Crystal Microbalance (QCM) humidity sensor based on a Pencil-Drawn method with high quality factor. IEEE Transactions on Electron Devices [Internet]. 2021 Aug 18;68(10):5149–54. Available from: https://doi.org/10.1109/ted.2021.3104266
20. Chapelle O, Scholkopf B, Zien A. SemiSupervised Learning, ser. Adaptive Computation and Machine Learning series. UK: MIT Press, 2010.
21. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition [Internet]. O'Reilly Media; 2019. Available from: https://www.oreilly.com/library/view/hands-on-machine learning/9781492032632/
22. Weissman SA, Anderson NG. Design of Experiments and process optimization. A review of recent publications. Organic Process Research & Development [Internet]. 2014 Jun 27;19(11):1605–33. Available from: https://doi.org/10.1021/op500169m
23. Al-Kharusi G, Dunne NJ, Little S, Levingstone TJ. The role of machine learning and design of experiments in the advancement of biomaterial and tissue engineering research. Bioengineering [Internet]. 2022 Oct 17;9(10):561. Available from: https://www.mdpi.com/2306-5354/9/10/561
24. Zhao Y, Schiffmann N, Koeppe A, Brandt N, Bucharsky EC, Schell KG, et al. Machine learning assisted design of experiments for solid state electrolyte lithium aluminum titanium phosphate. Frontiers in Materials [Internet]. 2022 Feb 3;9:817-821. Available from: https://doi.org/10.3389/fmats.2022.821817
25. Green J, White M. Machine Learning for Experiments in the Social Sciences, ser. Elements in Experimental Political Science. UK: Cambridge University Press;2023. [Online]. Available: https://books.google. com.tr/books?id=B2-7EAAAQBAJ
26. Kirimli CE, Elgun E, Unal U. Machine learning approach to optimization of parameters for impedance measurements of quartz crystal microbalance to improve limit of detection. Biosensors and Bioelectronics: X. 2022; 10:100121. Available: https://www.sciencedirect.com/science/article/pii/S2590137022000139
27. Karanfil G. Kafeik Asit için QCM Nanasensör Hazırlanması, Selçuk Üniversitesi Fen Bilimleri Enstitüsü Kimya Mühensiliği Ana Bilim Dalı, Yüksek Lisans Tezi, Konya, 2013.

28. Rodoplu D. Biyosensör Uygulamalarına Yönelik Olarak Kuvars Kristal Mikroterazi Yüzeylerin Elektro-eğirme Yöntemi ile Modifikasyonu. Hacettepe Üniversitesi Nanoteknoloji ve Nanotıp Anabilim Dalı, Ankara, 2011.
29. Srivastava AK, Sakthivel P. Quartz-crystal Microbalance Study for Characterizing Atomic Oxygen in Plasma Ash Tools. Vac. Sci. Technol 2001; 19:97- 100.
30. Benes E. Improved quartz crystal microbalance technique. Journal of Applied Physics 1984; 56(3):608-626.
31. Heydari S, Haghayegh G. Application of Nanoparticles in Quartz Crystal Microbalance Biosensors. Journal of Sensor Technology 2014;4:81-100.
32. Gomide AB. QCM Analysis of Liposomes and Cells. 2011 Retrieved from http://www.ipeuropa.com/petra/es/actividades/petra2/becarios/Andreza_Barbosa/Project_Description_Andreza_Barbosa.pdf
33. Keller CA, Kasemo B. Surface Specific Kinetics of Lipid Vesicle Adsorption Measured with a Quartz Crystal Microbalance. B Biophys J. 1998;75:1397−1402.
34. Kanazawa K, Nam-Joon C. Quartz Crystal Microbalance as a Sensor to Characterize Macromolecular Assembly Dynamics. Journal of Sensors. 2009;2009.
35. Nicu L, Leichle T. Biosensors and tools for surface functionalization from the macro to the nanoscale: The way forward. J App Phys. 2008;104:1-16.
36. Dixon M. Quartz Crystal Microbalance with Dissipation Monitoring: Enabling Real-Time Characterization of Biological Materials and Their Interactions. J Biomol Tech. 2008;19(3):151–8.
37. Viitala T. Modeling the mechanical properties of adsorbed/deposited layers at air-solid and liquid-solid interfaces with the dissipative KSV QCM-Z500. Helsinki: KSV Instruments Ltd.; 2005; 1-33.
38. Matthew A, Victoria T. A survey of the 2001 to 2005 quartz crystal microbalance biosensor literature: applications of acoustic physics to the analysis of bio molecular interactions. J Mol Recognit. 2007;20(3):154–84.
39. Höök F, Rudh M. Quartz Crystal Microbalance in biomacromolecular recognition. 2005;17:8-13 Available from: https://www.researchgate.net/publication/316618926_Quartz_crystal_microbalances_QCM_in_biomacromolecular_recognition

40. Türkmenoğlu C. Sentiment analysis in Turkish media. Istanbultek [Internet]. 2014; Available from: https://www.academia.edu/9432570/Sentiment_Analysis_in_Turkish_Media.
41. Mockenhaupt A. Digitalisierung und Künstliche Intelligenz in der Produktion. Germany: Springer Vieweg;2021.
42. Mohammed M, Khan MB, Bashier EBM. Machine Learning [Internet]. CRC Press eBooks. 2016. Available from: https://doi.org/10.1201/9781315371658.
43. Han J, Pei J, Kamber M. Data mining: concepts and techniques. 3rd ed. Amsterdam: Elsevier; 2011.
44. Sarker IH, Kayes ASM, Badsha S, Alqahtani H, Watters P, Ng A. Cybersecurity data science: an overview from machine learning perspective. J Big Data. 2020;7(1):1–29.
45. Van Otterlo M., Wiering, M. Reinforcement Learning and Markov Decision Processes. In: Reinforcement Learning. Adaptation, Learning, and Optimization. Springer 2012;12.
46. Kumari R., Srivastava S.K. Machine Learning: A Review on Binary Classification. International Journal of Computer Applications 2017; (160): 11-15.
47. Alsafy BM, Aydam ZM, Mutlag WK. Multiclass classification methods: a review. Int. J. of Advanced Engineering Technology and Innovative Science, 2019;5(3);1-10.
48. Herrera F, Charte F, Rivera AJ, Del Jesus MJ. Multilabel Classification: problem analysis, metrics and techniques [Internet]. 2016. Available from: https://simidat.ujaen.es/node/303.
49. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. Scikit-learn: machine learning in python. J Mach Learn Res. 2011;12:2825–30.
50. Witten IH, Frank E, Hall MA. Data Mining: Practical machine learningtools and techniques. 3rd ed. US: Morgan Kaufmann; 2005.
51. Bangdiwala SI. Regression: simple linear. International journal of injury control and safety promotion. 2018;25(1): 113-115.
52. Uyanık GK, Güler N. A study on multiple linear regression analysis. Procedia-Social and Behavioral Sciences, 2013;106; 234-240.

53. Owen AB. A robust hybrid of lasso and ridge regression. Contemporary Mathematics, 2007;443(7); 59-72.
54. Blashfield RK, Aldenderfer MS. The literature on cluster analysis. Multivariate behavioral research. 1978;13(3); 271-295.
55. Maimon OZ, Rokach L. Data mining and knowledge discovery handbook. Choice Reviews Online [Internet]. 2011 Jun 1;48(10): 321-352. Available from: https://doi.org/10.5860/choice.48-5729.
56. Kanagala HK, Krishnaiah VVJR. A comparative study of K-Means, DBSCAN and OPTICS. 2022 International Conference on Computer Communication and Informatics (ICCCI) [Internet]. 2016 Jan 1;1–6. Available from: https://doi.org/10.1109/iccci.2016.7479923.
57. Hassan SI, Samad A, Ahmad O, Alam A. Partitioning and hierarchical based clustering: a comparative empirical assessment on internal and external indices, accuracy, and time. International Journal of Information Technology [Internet]. 2019 Nov 29;12(4):1377–84. Available from: https://doi.org/10.1007/s41870-019-00406-7.
58. Sarang P. Thinking data science [Internet]. The Springer series in applied machine learning. 2023. Available from: https://doi.org/10.1007/978-3-031-02363-7.
59. Xu D, Yingjie TA. Comprehensive survey of clustering algorithms. Ann Data Sci. 2015;2(2):165–93.
60. Rasmussen C. The infnite gaussian mixture model. Adv Neural Inform Process Syst. 1999;12:554–60.
61. Carpenter GA, Grossberg SA. Massively parallel architecture for a self-organizing neural pattern recognition machine. Comput Vis Graph Image Process. 1987;37(1):54–115.
62. Kodinariya TM, Makwana PR. Review on determining number of Cluster in K-Means Clustering. International Journal. 2013;1(6):90-95.
63. Derpanis KG. Mean shift clustering. Lecture Notes. 2005;32(1-4):16.
64. Wang Z, Da Cunha C, Ritou M, Furet B. Comparison of K-means and GMM methods for contextual clustering in HSM. Procedia Manufacturing [Internet]. 2019 Jan 1;28:154–9. Available from: https://doi.org/10.1016/j.promfg.2018.12.025.
65. Goodfellow I, Bengio Y, Courville A. Deep Learning. The MIT Press; 2016.

66. Bishop CM. Pattern recognition and machine learning. Journal of Electronic Imaging 2007 Jan 1;16(4):049901. Available from: https://doi.org/10.1117/1.2819119.
67. Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data Via the EM Algorithm. Journal of the Royal Statistical Society Series B (Statistical Methodology) [Internet]. 1977 Sep 1;39(1):1–22. Available from: https://doi.org/10.1111/j.2517-6161.1977.tb01600.x.
68. Pham DL, Xu C, Prince JL. Current Methods in Medical Image Segmentation. Annual Review of Biomedical Engineering [Internet]. 2000 Aug 1;2(1):315–37. Available from: https://pubmed.ncbi.nlm.nih.gov/11701515/.
69. Bouguettaya A, Yu Q, Liu X, Zhou X, Song A. Efficient agglomerative hierarchical clustering. Expert Systems with Applications. 2015;42(5): 2785-2797.
70. Sarker IH, Abushark YB, Alsolami F, Khan A. Intrudtree: a machine learning based cyber security intrusion detection model. Symmetry. 2020;12(5):754. doi:10.3390/sym12050754.
71. Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. In: ACM SIGMOD Record. ACM. 1993;22: 207–216.
72. Puterman ML. Markov decision processes: discrete stochastic dynamic programming. Canada: John Wiley & Sons; 2014.
73. Quinlan JR. Induction of decision trees. Machine Learning [Internet]. 1986 Mar 1;1(1):81–106. Available from: https://doi.org/10.1007/bf00116251.
74. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Van Den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, et al. Mastering the game of go with deep neural networks and tree search. nature. 2016;529(7587):484–9.
75. Kaelbling LP, Littman ML, Moore AW. Reinforcement learning: a survey. J Artif Intell Res. 1996;4:237–85. doi: 10.1613/jair.301.
76. Sarker IH, Abushark YB, Khan A. Contextpca: predicting context-aware smartphone apps usage based on machine learning techniques. Symmetry. 2020;12(4):499. doi: 10.3390/sym12040499.
77. Xin Y, Kong L, Liu Z, Chen Y, Li Y, Zhu H, Gao M, Hou H, Wang C. Machine learning and deep learning methods for cybersecurity. IEEE Access. 2018;6:35365–81. doi: 10.1109/ACCESS.2018.2836950.

78. Nizam H, Akın S. Sosyal Medyada Makine Öğrenmesi ile Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması. XIX. Türkiye'de İnternet Konferansı; 2014 Kasım 27; İzmir, Türkiye.
79. Kotsiantis SB. Supervised machine learning: A review of classification techniques. Informatica. 2007;31:249-268.
80. Rosenblatt F. Principles of Neuradynamics: Perceptrons and The Theory of Brain Mechanisms. Washington D.C.: Spartan Boks; 1962.
81. Tekeli K, Aşlıyan R, Adnan Menderes Üniversitesi. Spam E-mail Detection Using Multilayer Perceptron, K-Nearest Neighbor and C4.5 methods [Internet]. 2021. Available from: https://ab.org.tr/ab16/bildiri/216.pdf.
82. Haykin S. Neural Networks and Learning Machines. 3rd Edition. USA: Macmillan College Publishing Company; 1994.
83. Wu CH, Tsai CH. Robust classification for spam filtering by backpropagation neural networks using behavior-based features. Applied Intelligence. 2009;31(2):107-121.
84. Muckley ES, Anazagasty C, Jacobs CB, Hianik T, Ivanov IN. Low-cost scalable quartz crystal microbalance array for environmental sensing. Proceedings of SPIE. 2016 Sep 27;9944:99440Y. Available from: https://ui.adsabs.harvard.edu/abs/2016SPIE.9944E..0YM/abstract.
85. Kirimli CE, Shih WH, Shih WY. DNA hybridization detection with 100 zM sensitivity using piezoelectric plate sensors with an improved noise-reduction algorithm. The Analyst. 2014;139(11):2754–63.
86. Johannsmann D. Studies of Viscoelasticity with the QCM. In: Springer eBooks [Internet]. 2006;49–109. Available from: https://doi.org/10.1007/5346_024.
87. Hovgaard MB, Dong M, Otzen DE, Besenbacher F. Quartz Crystal Microbalance Studies of Multilayer Glucagon Fibrillation at the Solid-Liquid Interface. Biophysical Journal. 2007;93(6):2162–9. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC1959527/.
88. Lagarias JC, Reeds JA, Wright MH, Wright PE. Convergence Properties of the Nelder--Mead Simplex Method in Low Dimensions. SIAM Journal on the Optimization. Available from: https://doi.org/10.1137/s1052623496303470.

89. Landry JP, Sun YS, Guo XW, Zhu XD. Protein reactions with surface-bound molecular targets detected by oblique-incidence reflectivity difference microscopes. Applied Optics. 2008;47(18):3275. Available from: https://opg.optica.org/ao/abstract.cfm?uri=ao-47-18-3275.
90. Ahmed SH, Espinoza-Sánchez NA, El-Damen A, Fahim SA, Badawy MA, Greve B, et al. Small extracellular vesicle-encapsulated miR-181b-5p, miR-222-3p and let-7a-5p: Next generation plasma biopsy-based diagnostic biomarkers for inflammatory breast cancer. PLoS ONE. 2021;16(4):e0250642. Available from: https://doi.org/10.1371/journal.pone.0250642.
91. Montgomery DC, Runger GC, Hubele NF. Engineering statistics. 4th Edition. John Wiley & Sons; 2011.
92. Figueira V, Jones J. Viscoelastic study of the adsorption of bovine serum albumin on gold and its dependence on ph. Journal of Colloid and Interface Science. 2008;325(1):107–113.
93. Koizumi H, Whitten WB, Reilly PTA. Trapping of intact, singly-charged, bovine serum albumin ions injected from the atmosphere with a 10-cm diameter, frequency-adjusted linear quadrupole ion trap. Journal of the American Society for Mass Spectrometry. 2008;19(12):1942–7.

8 CURRICULUM VITAE