

WATER QUALITY MONITORING AND CONTAMINANTS ANALYSIS WITH
COFFEE-RING EFFECT BY MACHINE LEARNING

By

Xiaoyan Li

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Environmental Engineering – Doctor of Philosophy
Computational Mathematics, Science and Engineering – Dual Major

2023

ABSTRACT

In the first stage, a low-cost tap water fingerprinting technique was explored using the coffee ring effect, which produces distinguishable residue patterns after tap water evaporates. This technique was evaluated by photographing tap water droplets from different communities in southern Michigan with a cell phone camera and 30x loupe. A convolutional neural network (CNN) model was then trained using the images to group the tap waters with similar water chemistry, achieving 80% accuracy. Further experiments were conducted to determine the influence of lower concentration species in the tap water "fingerprint". By analyzing the residue patterns from salt mixtures with varying concentrations of sodium, calcium, magnesium, chloride, bicarbonate, and sulfate, it was found that the residue patterns are unique and reproducible, and are associated with the water chemistry of the sample. Principal component analysis (PCA) was also applied to the image files and particle measurements, further highlighting differences in the residue patterns. The results suggest that the residue patterns of tap water, imaged with a cell phone camera and loupe, contain valuable information about the composition of tap water, and the coffee ring effect should be further studied for potential use in low-cost tap water fingerprinting.

The second stage examined the coffee-ring effect for tap water component analysis using synthetic samples with varying concentrations of ions. A custom four-axis autosampler was built using Raspberry Pi, a 3D printer stage, and programmed with Ubuntu and Python 3.7. The experiment was conducted in a controlled temperature and humidity chamber. SEM images, EDS mapping, and particle features extracted from photographs were analyzed using statistical methods. Optimal

conditions were identified as 23-26°C with 45%-50% humidity, 20-23°C with 45%-50% humidity, and 26-29°C with 40%-45% humidity, showcasing the coffee-ring effect as a low-cost, effective technique for tap water analysis. In the third stage, three models were evaluated in this research: the One-stage point estimation model (OnePeM), the Two-stage vision-transformer point estimation model (TwoVtPeM), and the Two-stage vision-transformer multiple output estimation model (TwoVtMoM). The TwoVtPeM technique achieved the best performance of the models tested (OnePeM, TwoVtPeM and TwoVtMoM), with OnePeM also performing well and TwoVtMoM falling short. The TwoVtPeM relative percentage errors were $\pm 17.1\%$ for oxygen, $\pm 4.5\%$ for sulfur, $\pm 19.9\%$ for sodium, $\pm 5.7\%$ for chlorine, $\pm 19.8\%$ for calcium, $\pm 25.8\%$ for magnesium, and $\pm 20.1\%$ for carbon. The R^2 was 0.95 which is higher than OnePeM with 0.90 R^2 and TwoVtMoM which was 0.54. The TwoVtPeM had a higher error mean than OnePeM, but it exhibited lower relative standard deviations of estimation; the TwoVtPeM relative standard deviations values were: 3.9% for oxygen, 3.0% for sulfur, 5.3% for sodium, 3.9% for magnesium, 5.3% for chlorine, 10.0% for calcium, and 5.9% for carbon. Moreover, 79.2% of water samples were correctly classified for hardness based on the estimated element concentrations by TwoVtPeM. Compared to strip test kits, this technology offers advantages such as speed, low cost, and the ability to simultaneously estimate multiple contaminants. However, addressing certain limitations, such as the quality of the substrate used and the size and complexity of the dataset and models, is essential. The TwoVtMoM is underfitting and requires additional training epochs and fine-tuning. Overall, this research demonstrates a promising technique for water quality analysis, providing a low-cost, fast, and relatively accurate method for estimating water contaminant concentrations.

Copyright by
XIAOYAN LI
2023

To Life.

ACKNOWLEDGMENTS

I would like to express my most profound gratitude to Dr. Lahr and Dr. Xie for graciously providing me with the laboratory facilities within the Department of Civil and Environmental Engineering and Computational Mathematics, Science, and Engineering at Michigan State University. Their invaluable guidance, unwavering support, and enthusiasm have greatly contributed to my research experience. I am deeply thankful to all the committee members for sharing their expertise and for fostering a sense of camaraderie throughout my academic journey.

My heartfelt appreciation goes to Dr. Tan for his generosity in sharing his vast knowledge and for extending me the opportunity to attend his group meetings, where I could learn about cutting-edge machine learning models and techniques. I am also grateful to his current and former group members for their willingness to share their valuable insights and advice on my research project.

I would like to express my gratitude to Dr. Tarabara and his research group for providing me with the essential training and access to the contact angle goniometer, which played a crucial role in my research. I am deeply indebted to Dr. Xie for serving as my advisor in the CMSE department, offering guidance in my research, and granting me access to the lab computer and GPU.

I also extend my thanks to Dr. Phanikumar for his insightful advice and expertise on my research problem and techniques. His guidance has been invaluable in helping me navigate the complexities of my work and ensuring that my research is both rigorous and impactful.

Furthermore, I would like to thank Dr. Yunhao Liu for his guidance of my career and for introducing me to the world of volleyball. His enthusiasm for the sport is contagious,

and playing volleyball has become a source of joy and stress relief for me. I appreciate the time he has spent teaching me the skills and techniques, as well as the camaraderie we have shared on the court. Dr. Liu's support has been instrumental in both my professional and personal development, and I am truly grateful for his mentorship.

Each of these individuals has played a significant role in my academic journey, and I am deeply grateful for their contributions to my growth and success. Their guidance, expertise, and support have been invaluable, and I am honored to have had the opportunity to learn from and work with such exceptional mentors and colleagues.

Thank you all for your unwavering commitment to my development and for helping me become the researcher and individual I am today. Your contributions to my success will never be forgotten, and I look forward to carrying the lessons you have taught me into the future.

Furthermore, I would like to extend my gratitude to my friends and all the individuals I had the pleasure of meeting at Michigan State University, Futurewei Technologies, and Lucy-labs Inc. Their support, advice, and friendship have enriched both my research and personal life.

Lastly, I want to express my utmost appreciation to my family for their unwavering love, encouragement, and support throughout my academic journey. Their belief in me has been a constant source of strength and motivation, and I could not have achieved this without them.

PREFACE

This dissertation represents the culmination of my years of dedicated study and research in the interdisciplinary fields of Civil and Environmental Engineering and Computational Mathematics, Science, and Engineering. Since embarking on my PhD journey in 2016, I have been confronted with an array of challenges that society and the built environment face on a daily basis. My research endeavors to contribute to the advancement of these fields by proposing innovative and effective solutions to real-world problems, paving the way for sustainable development.

The primary focus of my dissertation revolves around the coffee ring effect and tap water quality monitoring, which has broad implications for public health and environmental safety. Throughout my PhD program, I have extensively explored the intricacies of machine learning, statistics, computer vision, analytical chemistry, and microfluidics, and have drawn upon these disciplines to enhance my research. By conducting a comprehensive analysis of the current state of the field, I identified key areas for improvement that helped shape my research agenda. My findings were ultimately derived from a synergistic blend of theoretical analysis, numerical modeling, and hands-on laboratory experimentation. I am deeply grateful to everyone who supported me throughout this journey, particularly my advisor, Dr. Lahr, who has been a constant source of guidance, inspiration, and encouragement. I also wish to extend my appreciation to my colleagues and fellow students, whose camaraderie and intellectual stimulation have enriched my overall experience and contributed to my personal growth.

It is my fervent hope that my research efforts will make a significant and lasting impact on the field of Civil and Environmental Engineering, ultimately contributing to

the betterment of our world through innovative approaches to sustainable development and environmental stewardship. As I look back on my PhD journey, I am filled with a sense of accomplishment and a renewed commitment to using my knowledge and skills to create a healthier, more sustainable future for all.

TABLE OF CONTENTS

CHAPTER 1

Introduction	1
1.1 Need for innovation in drinking water monitoring	1
1.2 Coffee-ring effect introduction	1
1.2.1 What is coffee-ring effect?	1
1.2.2 Several factors in pattern formation of crystals in the coffee-ring effect process.....	2
1.2.3 Understanding the mechanism of coffee-ring effect	3
1.2.4 Crystal structure prediction with energy minimization.....	4
1.2.5 Coffee-ring effect applications.....	4
1.3 Machine-Learning Models in water treatment and modeling	8
1.3.1 Image analysis via convolutional neural network (CNN).....	9
1.3.2 Vision Transformer in computer vision.....	14
1.3.3 Machine-Learning Models and Artificial-Intelligence Methods in Water Treatment	15
1.3.4 Applications of AI and ML methods in Water Treatment	27

CHAPTER 2

Tap water fingerprinting using a convolutional neural network built from images of the coffee-ring effect.....	50
2.1 Abstract.....	50
2.2 Introduction	51
2.3 Experimental	55
2.4 Results and discussion	65
2.5 Conclusions and future outlook	79

CHAPTER 3

Optimal environmental condition for contaminants separation by coffee-ring Effect	81
3.1 Abstract.....	81
3.2 Introduction	82
3.3 Experimental Methods.....	93
3.3.1 Materials and instruments.....	93
3.3.2 Four-axis-autosampler	94
3.3.3 Auto temperature humidity control chamber	97
3.3.4 Water samples.....	99
3.3.5 Coffee-ring effect pattern statistical analysis methods	100
3.3.6 Experiment procedure	106
3.4 Results and Discussion	110
3.4.1 Under what environmental conditions are coffee-ring effect fingerprints are consistent	110
3.4.2 What are the optimal environmental conditions that different water samples exhibit mostly different coffee-ring effect residue patterns	113
3.4.3 Under each environmental condition, are the elements deposition locations significantly different from each other.....	127

3.4.4 Do the water sample coffee-ring effect patterns have significant statistical correlation with element composition.....	130
3.5 Conclusion.....	132
CHAPTER 4	
CNN-Vision-transformer model for elements concentration estimation by coffee-ring effect residue patterns	134
4.1 Abstract.....	134
4.2 Introduction	135
4.2.1 Coffee-ring effect residue provides particles structure information.....	136
4.2.2 Applications of AI and ML methods in Water Treatment	137
4.2.3 Model for elements recognition and concentration estimation	141
4.3 Experimental Methods.....	144
4.3.1 Develop a deep learning model to identify corrosion indicators and quantify their concentrations in tap water.....	144
4.4 Results and Discussion	152
4.4.1 Elements correlations between coffee-ring effect subrings.....	152
4.4.2 Elements mapping estimation model analysis.....	153
4.4.3 Two-stage model produces better results than one-stage model	161
4.5 Conclusion.....	174
CHAPTER 5	
Implications	177
BIBLIOGRAPHY	179
APPENDIX	203

CHAPTER 1

Introduction

1.1 Need for innovation in drinking water monitoring

The need for innovation in drinking water monitoring is growing due to increased awareness of the impact of contaminated water on human health and the environment. Current monitoring methods are often expensive, time-consuming, and reliant on manual analysis. As a result, there is a pressing need for more efficient, cost-effective, and reliable methods to monitor drinking water quality. Innovations in technologies, such as sensors and machine learning, have the potential to revolutionize drinking water monitoring by providing real-time data and reducing the need for manual analysis. In addition, incorporating these technologies into drinking water monitoring systems can help to address the current challenges of limited resources and expertise in many communities, leading to better access to safe and clean drinking water for all.

1.2 Coffee-ring effect introduction

1.2.1 What is coffee-ring effect?

The coffee-ring effect is a low-cost method for separating particles in aqueous samples. It occurs when a water droplet shrinks in height and its particles are squished into concentric circles based on size as the droplet dries on a hydrophobic substrate Wong et al. [2011]. This phenomenon is known as "nanochromatography" and has been used to separate particles with resolutions of 100 nm at low particle volume fractions Wong et al. [2011]. The separation is possible due to the differential effects of adhesion and surface tension forces, which move larger particles towards the center of the drop and hold smaller particles in place at the drop edge.

1.2.2 Several factors in pattern formation of crystals in the coffee-ring effect process

Takhistov and Chang and other researchers found coffee-ring effect (CRE) depends on temperature, concentration of particles and substrate hydrophobicity. film and solutal flux dynamics of such small drops at their contact lines can induce macroscopic concentration segregation and produce distinct large-scale stain patterns such as concentric rings on hydrophilic surfaces and latticed crystals on hydrophobic ones. Coupling between these bulk segregation instabilities and the classical Mullins-Sekerka crystallization instability results in a large variety of crystal patterns with interwoven complex structures of two length scales. Furthermore, low density crystals can occupy a larger area than the initial drop, and gravitational drainage on inclined substrates can change the larger length scale. Takhistov and Chang [2002], Shahidzadeh-Bonn et al. [2008], Zhong et al. [2017]. Researchers also found polyelectrolyte concentration and humidity have effects on pattern formation Kaya et al. [2010]. Shin also demonstrated solubility, evaporation rate and mobility of the contact line determines the pattern of formed crystals in the coffee-ring effect Shin et al. [2014]. Lee proved the degree of supersaturation affects the nucleation pathways of potassium dihydrogen phosphate solution droplet Lee et al. [2016]. It is also found in the evaporation process of NaCl, the hydrophobicity (wettability) of substrate has effects on formed crystal pattern. On hydrophilic surface, ringlike crystalline deposit surrounded by a small spreading film formed and on hydrophobic surface, a close-up of the cauliflower-like pattern on the residue border was formed. And degree of saturation has effects on crystals pattern of Na₂SO₄ Shahidzadeh-Bonn et al. [2008]. Researchers found salts concentration and wettability have effects on the formation of crystal pattern Zhong et al. [2017].

1.2.3 Understanding the mechanism of coffee-ring effect

In terms of numerical approaches, a variety of studies have been conducted on the pattern formation of evaporating suspensions containing dissolved nanoparticles, employing Monte Carlo models Kim et al. [2011], Stannard [2011], Robbins et al. [2011], Brownian dynamics Gupta and Peters [1985], Chen and Kim [2004] and physical microfluid mechanism modeling Kang et al. [2016], Fischer [2002], Shmuylovich et al. [2002], Pauchard and Allain [2003], Popov [2005], Heim et al. [2005].

Previous study investigated a computational Monte Carlo method approach for estimating the ring-like deposition of nanoparticles contained in a drying liquid droplet Kim et al. [2011]. The investigation of non-equilibrium dewetting processes in nanoparticle-containing solutions revealed various pattern for example ring-like structures formations and other underlying mechanisms Stannard [2011]. A dynamic density functional theory was developed to replicate branched 'flower-like', labyrinthine, and network structures and this model was used to examine the effects of solvent evaporation, as well as the diffusion of colloidal particles and liquid across the surface. Robbins et al. [2011]. A study demonstrated the formation of coffee stains necessitates specific boundary conditions, such as pinning boundaries Yunker et al. [2011]. A model based on the bulk flow within the drop transporting particles to the interface where they are captured by the receding free surface and subsequently transported along the interface until they are deposited near the contact line was investigated Kang et al. [2016] A review of recent studies can be found in Larson [2014].

1.2.4 Crystal structure prediction with energy minimization

Material synthesizing is an active area both in research and industry. Once a material is finally synthesized and characterized, its properties can be evaluated in the

engineering design process. However, to synthesize the desired material, most applications require an optimization of multiple properties which may be interrelated. In field of thermoelectrics, materials are compared to one another using a figure of merit. In this equation, S is the Seebeck coefficient, σ is the electrical conductivity, χ is the thermal conductivity, and T is temperature. However, the material properties σ , χ , and S are all interrelated. For example, electrical conductivity is positively related with high carrier concentration, whereas Seebeck coefficient is negatively related with carrier concentration to increase zT . In addition, thermal conductivity also increases with carrier concentration which in turn decreases zT . Therefore, optimization of thermoelectric materials requires a compromise between these properties. Also, the most significant advances in this field have come from identifying new compounds which exhibit a better intrinsic balance in these properties Graser et al. [2018].

1.2.5 Coffee-ring effect applications

Understanding and controlling the process of solute deposition in the presence of coffee-ring effect is important in manufacturing processes involving evaporation on surfaces including printing Park and Moon [2006], Friederich et al. [2013], Kuang et al. [2014], Sun et al. [2015], Huang and Zhu [2019] and fabrication of ordered structures Han and Lin [2012], functional nanomaterials Shao et al. [2014], Zou and Kim [2014] and colloidal crystals Park et al. [2006], Cui et al. [2009]. coffee-ring effect also improves the performance of commercial applications including fluorescent microarrays Blossey and Bosio [2002], Dugas et al. [2005], matrix assisted laser desorption ionization (MALDI) spectrometry Hu et al. [2013], Mampallil et al. [2012], Kudina et al. [2016], Lai et al. [2016], and surface enhanced Raman spectroscopy (SERS) Zhou et al. [2014a], Wang et al. [2014], Garcia-Cordero and Fan [2017]. coffee-ring effect has also implications in plasmonics Li et al. [2016a], solute separation Wong et

al. [2011], diagnostics Brutin et al. [2011], Wen et al. [2013], Gulka et al. [2014] and electronics applications de Gans and Schubert [2004].

Suppression of coffee-ring effect

Coffee-ring effect can be suppressed through one of the three physical strategies (i) preventing the pinning of the contact line; (ii) disturbing the capillary flow towards the contact line and (iii) preventing the particles being transported to the droplet edge by the capillary flows. The coffee-ring effect could be suppressed by preventing contact line pinning using hydrophobic surfaces. Increasing the hydrophobicity of surfaces is often accompanied by decreasing contact angle hysteresis (CAH) Eral et al. [2013]. Lower CAH in essence means reduced contact line pinning which leads to suppression of coffee-ring effect. Lower CAH could be achieved by patterning of controllable surface wettability as reviewed previously by Tial et al. Tian et al. [2013]. These methods include chemical modification Ko et al. [2004], Tian et al. [2013], Li et al. [2018] and physical modification Yunker et al. [2011].

On hydrophobic and partially hydrophobic surfaces, pinning can even occur when the CAH or solute concentration is high. If CAH is high, during the contact angle decreases to the receding angle, typically a few seconds depending upon the rate of evaporation, solutes can accumulate at the contact line. Such accumulation produces ring-like deposits only if the duration of pinning is above a critical value for a given substrate-solute system

Moraila-Martinez et al. [2013]. However if the pinning time is short, even with high initial solute concentration, the coffee-ring effect will just produce smaller inner rings Nguyen et al. [2013]. The nanoparticles are more prominent to form ring like patterns compared with larger particles as they can flow into the microscopic regions of the droplet edge faster. In

the presence of solute particles in the droplet, electrowetting (EW) can reduce the pinned contact line on (partially)-hydrophobic surfaces Mugele and Baret [2005], Li and Mugele [2008]. A droplet is deposited on a dielectric layer covering an electrode. When a voltage is applied between the droplet and the electrode an electric force pulls the contact line outward, overcoming the pinning forces so the contact line pinning is reduced. The coffee-ring effect can also be suppressed by vibration and acoustics, marangoni flow and other factors Mampallil and Eral [2018]. Researchers have also proposed a method that relies on the covalent cross-linking of monodisperse materials, which allows for the formation of thin films with uniform thicknesses and macroscale cohesion. This approach prevents the coffee-ring effect by inducing gelation of the coating materials through a thioacetate-disulfide transition, counterbalancing the capillary forces generated by evaporation Li et al. [2018].

Enhancing coffee-ring effect

Evaporation of droplets can be utilized as a method to concentrate its solutes in it. Evaporation of the solvent can increase the analyte concentration making the reactions more probable Hernandez-Perez et al. [2016], De Angelis et al. [2011]. Concentrating solutes at the rim of the droplet by coffee-ring effect is called the self-ordered ring (SOR) method. It acts as a pre-concentration procedure before other analyses. The deposition of solutes and particles are exploited as a pre-concentration method 1.1. To enhance the coffee-ring effect, hydrophobic surface is usually used as the substrate. Drying process on hydrophobic surfaces forms smaller rings with higher solute density as the contact line is pinned only in the later stages of the evaporation. Liu et al. demonstrated that the SOR method enhanced the fluorescence detection of orally administrated berberine in human urine Liu et al. [2002]. Similarly, fluorescent detection of trace levels of tetracycline

Huang et al. [2004a], quinidine sulfate in serum samples Yang and Huang [2006] and fluorescein Liu et al. [2006] was demonstrated based on the SOR method. Coffee-ring effect could facilitate identifying pathogens which are associated with diseases by isolating the disease markers from body fluids Wong et al. [2011], Chen and Evans [2010].

The coffee-ring effect has been found to have several practical applications in various fields. In particular, it has been utilized to enhance the deposition of gold nanoparticles (AuNPs) on cellulose nanofibers (CNFs) for the purpose of improving surface-enhanced Raman scattering (SERS) as reported in several studies Chen et al. [2017], Wang et al. [2014], Hussain et al. [2019], Juneja and Bhattacharya [2019], Zhou et al. [2014b]. The coffee-ring effect has also been used as a low-cost approach for malaria diagnosis Gulka et al. [2014]. Additionally, the coffee-ring effect has shown potential for monitoring tap water quality with the help of deep neural networks Li et al. [2020].

Furthermore, the coffee-ring effect has the potential to aid in identifying pathogens associated with various diseases by isolating disease markers from body fluids Wong et al. [2011], Chen and Evans [2010]. These findings demonstrate the versatile and practical applications of the coffee-ring effect in various fields.

		Hydrophobic	Electrowetting	electroosmotic flow	Acoustics	Marangoni flow (thermal)	Marangoni flow (surfactant)
Mode of action	De-pinning contact line	✓	✓				
	Disturbing capillary flow		✓	✓	✓	✓	
	Trapping solute			✓			
	Dynamic control		✓	✓			
Advantages	Compatible with wide substrate-wettability		✓	✓	✓	✓	
	Compatible with particle denser than liquid			✓			
	Compatible with biological samples		✓	✓	✓	✓	
	Suitable for MALDI sample preparation	✓	✓	✓			
Limitations	Require substrate modification	✓	✓				
	Depends on particle shape						
	Invasive (electric field)		✓				
	Invasive (chemical)		✓			✓	
	Invasive (heating)		✓	✓	✓		
	Require external forcing	✓	✓	✓			
		Particle-LG interface interactions	Particle-SL interface interactions	Transition of liquid property	Humidity cycling	Porous substrates	
Mode of action	De-pinning contact line						
	Disturbing capillary flow			✓	✓		
	Trapping solute	✓	✓	✓			
	Dynamic control						
Advantages	Compatible with wide substrate-wettability	✓	✓	✓			
	Compatible with particle denser than liquid		✓				
	Compatible with biological samples		✓				
	Suitable for MALDI sample preparation						
Limitations	Require substrate modification				✓		
	Depends on particle shape	✓					
	Invasive (electric field)						
	Invasive (chemical)						
	Invasive (heating)						
	Require external forcing			✓			

Figure 1.1: Suppression and Enhancement of coffee-ring effect. Comparison of different methods. The working principle, advantages and limitations are illustrated.

1.3 Machine-Learning Models in water treatment and modeling

The table referred to as Table 1.1 provides a summary of AI and ML models and methods used in water treatment and modeling applications. It highlights their general and specific uses, as well as the advantages and disadvantages of each method. The final column includes references to peer-reviewed textbook sources that offer comprehensive and in-depth explanations of these models and methods. Although the table may not cover every aspect of water treatment and modeling, the applications selected are based on a well-defined

methodology. It is worth noting that the majority of the ML methods listed in the table fall under the "black-box" category, which is generally considered a drawback for most models. However, the exception to this are Genetic Algorithms (GA) and Gaussian Processes (GPs).

1.3.1 Image analysis via convolutional neural network (CNN)

The basic ideas underlying the use of convolutional neural networks (CNNs, also known as ConvNets) for inverse problems are not innovative. For more historical perspective, see Schmidhuber [2015], Li et al. [2016b], and for an accessible introduction to deep neural networks and a summary of their recent research, see LeCun et al. [2015], Schwendicke et al. [2019], Brinker et al. [2018]. The CNN architecture was proposed in 1986 in RUMBERT [1986] and were developed for solving inverse imaging problems as early as 1988 Zhou et al. [1988]. These approaches, which used networks with a few parameters and did not always include learning, were largely superseded by compressed sensing (or, broadly, convex optimization with regularization) approaches in the 2000s. As computer hardware improved, it became feasible to train larger and larger neural networks, until, in 2012, Krizhevsky et al. [2017] achieved a significant improvement over the state of the art on the ImageNet classification challenge by using a GPU to train a CNN with 5 convolutional layers and 60 million parameters on a set of 1.3 million images. This work spurred a resurgence of interest in neural networks, and specifically CNNs, for not only computer vision tasks, but also inverse problems and more. With the development of CNN models, both accuracy and operation have increased dramatically.

Basic CNN components

There are numerous variants of CNN architectures in the literature. However, their basic components are the same. They all consist of three types of main layers, namely

convolutional, pooling, and fully-connected layers. The convolutional layer aims to learn feature representations of the inputs, for example human eyes features, nose features or objects. As shown in Figure. 1.2 Convolution layer is composed of several convolution kernels which are used to compute different feature maps. Specifically, each neuron of a feature map is connected to a region of neighbouring neurons in the previous layer. This neighbourhood is referred to as the neuron's receptive field in the previous layer. The new feature map can be obtained by first convolving the input with a learn-able kernel and then applying an element-wise nonlinear activation function on the convolved results. After the activation function, a pooling layer is normally applied to the feature map to filter the high frequency noise. The complete feature maps are obtained by using several different kernels with the same or different activation and pooling functions Gu et al. [2018]. Mathematically, the feature value at location (i, j) in the kth feature map of lth layer, $z_{i,j,k}^l$ is calculated by the equation:

$$z_{i,j,k}^l = w_k^l x_{i,j}^l + b_k^l \quad (1.2)$$

Where w_k^l and b_k^l are the weight vector and bias term of the k th filter of the l th layer respectively, and $x_{i,j}^l$ is the input patch centered at location (i, j) in the previous layer, the l th layer. It worth to know that the kernel w_k^l that generates the feature map $z_{i,j,k}^l$ is shared but there are several different kernels generated and learned in the model building process

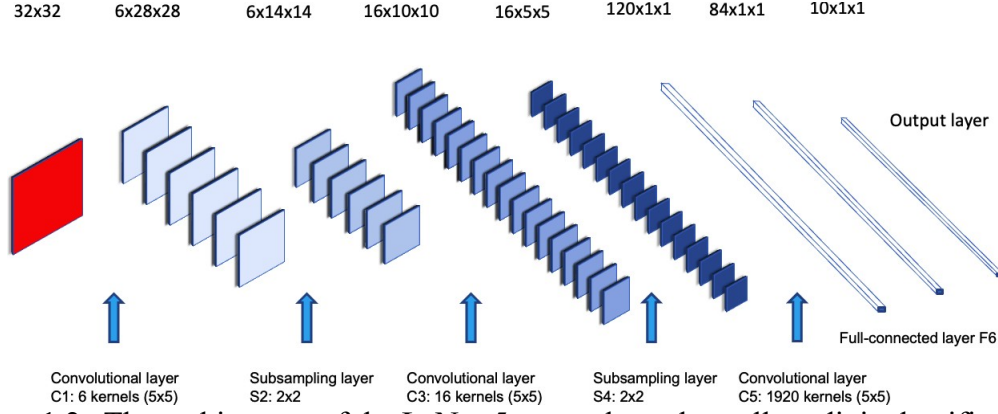


Figure 1.2: The architecture of the LeNet-5 network works well on digit classification task.

Such a weight sharing mechanism has several advantages such as it can reduce the model complexity and make the network easier to train. At the same time, to not lose generality and information, several kernels are trained and implemented in the model structure. The activation function introduces nonlinearities to CNN, which are desirable for multi-layer networks to detect nonlinear features Gu et al. [2018]. The activation functions are normally Sigmoid function, ReLU function, Tanh function and their derivatives LeCun et al. [2012], Hinton [2010]. Let $a(\cdot)$ denote the nonlinear activation function. The activation value $a(i, j, k)$ of convolutional feature z^l can be computed as

$$a_{i,j,k}^l = a(z_{i,j,k}^l) \quad (1.2)$$

The pooling layer aims to achieve shift-invariance and information aggregation by reducing the dimension of the feature maps in the previous layer. It is usually placed between two convolutional layers. Each feature map of a pooling layer is connected to its corresponding feature map of the preceding convolutional layer. Denoting the pooling function as $\text{pool}(\cdot)$, each feature map a^l could be denoted as:

$$Y = \text{pool}\left(a_{\{m,n,k\}}^{\{l\}}\right) \quad (1.3)$$

In this equation, where R_{ij} is a local neighbourhood around location (i, j) . The typical pooling operations are average pooling Wang et al. [2012] and max pooling Boureau et al.

[2010], Murray and Perronnin [2014]. The kernels in the lower convolutional layers are designed to detect low-level features such as edges and curves, while the kernels in higher layers are learned to detect more abstract features. By stacking several convolutional, activation and pooling layers, the model could gradually extract higher-level feature representations.

After the convolutional and pooling layers, there may be one or more fully-connected layers which aim to perform high-level reasoning Simonyan and Zisserman [2014], Zeiler and Fergus [2014], Hinton et al. [2012]. They take all neurons in the previous layer and connect them to every single neuron of current layer to generate global semantic information. Note that fully-connected layer not always necessary as it can be replaced by a 1×1 convolution layer Lin et al. [2013], Saxena and Verbeek [2016]. The last layer of CNNs is an output layer. Softmax operator is commonly used for classification tasks Russakovsky et al. [2015]. Another commonly used method is SVM, which can be combined with CNN features to solve different classification tasks Tang [2013], Madjarov et al. [2012]. Let θ denote all the parameters of a CNN (e.g., the weight vectors and bias terms). The optimum parameters for a specific task can be obtained by minimizing an appropriate loss function defined on that task. Suppose we have N desired input-output relations $(\mathbf{x}^n, \mathbf{y}^n)$; $n \in [1, \dots, N]$, where $\mathbf{x}^n$ is the n -th input data, $\mathbf{y}^n$ is its corresponding target label and o^n is the output of CNN.

The aim of training CNN is a problem of global optimization. However, in practice, it is often a local minimum problems and by minimizing the loss function. Stochastic gradient descent is a common solution for optimizing to find the best fitting set of parameters.

The loss of CNN can be calculated as follows:

$$L = \frac{1}{N} \sum_{n=1}^N l(\theta, y^n, o^n) \quad (1.4)$$

Recent advances in convolutional neural networks

Since 2006, many methods have been developed to overcome the difficulties encountered in training deep CNNs Niu and Suen [2012], Russakovsky et al. [2015], Simonyan and Zisserman [2014], Szegedy et al. [2015]. For example, the CNN model proposed by Krizhevsky et al. showed significant improvements upon previous methods on the image classification task. The overall architecture of their method, i.e., AlexNet Russakovsky et al. [2015], is similar to LeNet-5 but with a deeper structure. With the success of Krizhevsky's work, many works have been proposed to improve its performance. Among all these works, there are four models which are most representative. These models are ZFNet Zeiler and Fergus [2014], VGGNet Simonyan and Zisserman [2014], GoogleNet Szegedy et al. [2015] and ResNet He et al. [2016]. From the evolution of the model architectures, a typical trend is that researchers are building deeper networks, e.g., ResNet, which won the champion of ILSVRC 2015, is about 20 times deeper than AlexNet. Theoretically, By increasing depth, the network can achieve better feature extraction and representation which could approximate the target function better. However, deeper model architecture also increases the complexity of the network, which makes the network be more difficult to optimize and easier to get overfitting and suffer th curse of dimensionality problem. Along this way, various methods have been proposed to deal with these problems in various aspects.

1.3.2 Vision Transformer in computer vision

Deep neural networks (DNNs) form the core of AI systems. Different types of networks are designed for different tasks. The multi-layer perceptron (MLP) or fully connected (FC) network, made up of multiple linear layers and nonlinear activations, is a

classic neural network Rosenblatt [1957]. Convolutional neural networks (CNNs), consisting of convolutional and pooling layers, are used to process images and other shift-invariant data LeCun et al. [1998], Krizhevsky et al. [2017]. Recurrent neural networks (RNNs) use recurrent cells to process sequential or time series data Hochreiter and Schmidhuber [1997]. The transformer is a novel neural network that uses self-attention mechanisms Bahdanau et al. [2014], Parikh et al. [2016] to extract intrinsic features Vaswani et al. [2017]. It has shown potential for a wide range of AI applications, especially in NLP. For example, Vaswani et al. [2017] proposed the transformer for machine translation and English constituency parsing tasks, and BERT (Bidirectional Encoder Representations from Transformers) was introduced by Devlin et al. [2018], a language representation model that pre-trains the transformer on unlabeled text, considering the context of each word in a bidirectional manner. BERT achieved state-of-the-art results on 11 NLP tasks. Brown et al. [2020] pre-trained the massive transformer-based model, GPT-3 (Generative Pre-trained Transformer 3), using 45 TB of compressed plaintext data and 175 billion parameters, and it performed well on various downstream NLP tasks without fine-tuning. These transformer-based models have brought significant advances to NLP.

Inspired by the success of transformer architectures in NLP, researchers have recently applied them to computer vision (CV) tasks. Although CNNs have been traditionally considered the foundation of CV He et al. [2016], Ren et al. [2015], the transformer is emerging as a potential alternative. Chen et al. [2020] trained a sequence transformer to auto-regressively predict pixels, achieving results comparable to CNNs in image classification tasks. Dosovitskiy et al. [2020] proposed the vision transformer model, ViT, which directly applies a pure transformer to sequences

of image patches to classify the full image, and it has achieved state-of-the-art performance on multiple image recognition benchmarks. Transformer has also been used to solve various other CV problems, such as object detection Carion et al. [2020], Zhu et al. [2020], semantic segmentation Zheng et al. [2021], image processing Chen et al. [2021], and video understanding Zhou et al. [2018]. Its exceptional performance has attracted more researchers to propose transformer-based models for a wide range of visual tasks.

1.3.3 Machine-Learning Models and Artificial-Intelligence Methods in Water Treatment

Table 1.1 summarizes AI and ML models and methods, highlighting their general and specific usages in water treatment and modeling applications, as well as their advantages and disadvantages. The final column includes peer-reviewed textbook sources that provide foundational and in-depth explanations of these models and methods. While not all-encompassing, the selected water treatment and monitoring applications are based on a specified methodology. The majority of the included ML methods fall under the "black-box" archetype, which is generally considered a disadvantage for most models, with the exception of GA/GPs.

Table 1.1: A summary of AI methods and ML models used in water treatment and monitoring.

Learning and Modeling Technique	General Applications	Reviewed Water Treatment and Monitoring Applications	Advantages	Disadvantages
Support Vector Machines, Regressions	Classification Regression, Classification, Pattern Analysis Cortes and Vapnik [1995], Chua [2003], Noble [2006], Caie et al. [2021], Goodfellow et al. [2016]	Models for disinfection by-product (DBP) modeling Models for membrane process parameter modeling Models for biological oxygen demand (BOD) and chemical oxygen demand (COD) modeling Models for dissolved oxygen modeling of rivers Models for aquaponics growth rate modeling Models for aquaponics growth stage classification	Developing models capable of handling high dimensional datasets (i.e., datasets with a high number of inputs vs. a lower number of outputs) Developing models that can handle small changes in the dataset Developing models that are functional with both linear and nonlinear data.	Kernel selection is initially difficult and time consuming when using SVM/SVR modeling SVM/SVR modeling requires high computational power, making it mostly unsuitable for larger datasets SVM/SVR modeling is susceptible to noise in datasets SVM/SVR modeling has relatively long training times.

Table 1.1: (cont'd)

Leaning and Modeling Technique	General Applications	Reviewed Water Treatment and Monitoring Applications	Advantages	Disadvantages
Random Forest (RF)	Supervised machine learning Regression, Classification Maimon and Rokach [2005], Ceri et al. [2003], Singh et al. [2016], Liu et al. [2012], Hastie et al. [2009]	Modeling adsorption process parameters and percent removal using ML Developing simple and hybrid models for dissolved oxygen prediction and modeling	Intuitive model architecture for efficient and effective ML modeling Models capable of handling continuous and categorical inputs, even with missing values or data Models that are relatively stable and have less impact due to noise and outliers Bagging algorithms to reduce overfitting and variance in the model	Accuracy and robustness of the model are determined by the density of decision trees Increasing the density of decision trees results in significant increases in model complexity, training period, and required computational power

Table 1.1: (cont'd)

Learning and Modeling Technique	General Applications	Reviewed Water Treatment and Monitoring Applications	Advantages	Disadvantages
k-Nearest Neighbor (k-NN)	Supervised machine learning Classification Gaya et al. [2017], Zhu [2002], Abba et al. [2020], Wills et al. [2013], Allafi et al. [2017]	Classification of aquaponics growth stage	Requires minimal training and can be easily implemented Capable of handling new data additions without requiring significant modifications to the model	Poor performance with large datasets or those with high dimensionality Susceptible to noise and missing data, which can result in decreased accuracy

Table 1.1: (cont'd)

Learning and Modeling Technique	General Applications	Reviewed Treatment and Monitoring Applications	Advantages	Disadvantages
Fuzzy Inference System (FIS)	Decision making, system control Moraga et al. [2003], Afroozeh et al. [2018], Moon et al. [2011], Kaynak et al. [1998], Zadeh [1998]	Models for chlorine dosage set-point control Developing models for hydroponics system and environmental control	Utilizing fuzzy logic rather than binary logic to better model the human experience of decision making Developing models with easily interpretable outputs and decisions with a well defined system	The applicability of models developed with fuzzy logic is dependent on operator defined parameters and experience, which makes them prone to human error.

Table 1.1: (cont'd)

Learning and Modeling Technique	General Applications	Reviewed Treatment and Monitoring Applications	Advantages	Disadvantages
Artificial Neural Network	Supervised machine learning Regression, Classification Goodfellow et al. [2016], Shahmansouri et al. [2021]	DBP (disinfection byproduct) formation modeling Adsorption process parameter modeling Membrane process parameter modeling Chlorine dosage/set-point Dissolved oxygen concentration modeling	Capable of handling high dimensional datasets Modeling/prediction results obtained in a reasonable amount of time Forward propagation capable of cheap and fast computation	High computational power associated with backward propagation stage Some models and architecture themselves are difficult to interpret See below for specific ANN model disadvantages

Table 1.1: (cont'd)

Leaning and Modeling Technique	General Applications	Reviewed Water Treatment and Monitoring Applications	Advantages	Disadvantages
Convolutional Neural Network (CNN)	Regression, Classification, Segmentation LeCun et al. [2015], Kim and Kim [2017], Acharya et al. [2017], Gu et al. [2018]	Disinfection by-product formation modeling	CNNs have been shown to produce highly accurate results on a wide range of image and video recognition tasks Operations run in parallel and results are obtained quickly	Data must be in fixed dimensions Requires high computational power: Training and processing CNNs can be computationally intensive, requiring significant computational power and resources

Table 1.1: (cont'd)

Learning and Modeling Technique	General Applications	Reviewed Treatment and Monitoring Applications	Advantages	Disadvantages
Recurrent Neural Network (RNN)/Long Short Term Memory (LSTM)	Regression, Classification LeCun et al. [2015], Zhou et al. [2019], Zhang et al. [2020], Hochreiter and Schmidhuber [1997], Smagulova and James [2020]	Parameter modeling of membrane process Modeling of dissolve oxygen concentration modeling	Suitable for sequential datasets especially time series datasets and modeling Suitable for varying lengths of sequence datasets	Training and processing RNNs requires high computational power Prone to gradient exploding and vanishing
Hammerstein Wiener (HW)	Regression	Dissolved oxygen concentration modeling	Capturing nonlinear effects and simultaneously being computationally less complex than fully nonlinear dynamic models	Limited model structure

Table 1.1: (cont'd)

Learning and Modeling Technique	General Applications	Reviewed Water Treatment and Monitoring Applications	Advantages	Disadvantages
Genetic Algorithm	Evolutionary, stochastic algorithm Regression, Classification Agrawal and Mathew [2004] , Yang [2020] , Katoch et al. [2021]	DBP formation modeling	Parallelism: Genetic algorithms can explore multiple solutions simultaneously, allowing for faster convergence to an optimal solution Applicability: GAs are applicable to a wide range of problems, including those with discrete, continuous, or mixed variable types, and those with multiple objectives or constraints	Slow convergence: GAs can sometimes take a long time to converge to the optimal solution, especially for large or complex problems Premature convergence: GAs can converge prematurely to suboptimal solutions if the population diversity is lost Computational cost: Genetic algorithms can be computationally expensive, particularly for large scale or high dimensional problems

Table 1.1: (cont'd)

Learning and Modeling Technique	General Applications	Reviewed Water Treatment and Monitoring Applications	Advantages	Disadvantages
Radial Basis Function (RBF) Kernel	Regression, Classification LeCun et al. [2015], Karimi et al. [2020], Powell et al. [1981], Baddari et al. [2009]	Modeling of DBP formation Prediction of adsorption process removal efficiency Modeling of membrane process parameters	RBF networks are capable of approximating any continuous function, given a sufficient number of hidden neurons and appropriate basis functions RBF networks can be trained more quickly than other types of neural networks RBF networks are generally more robust to noise than other types of neural networks	RBF networks are hard to scale to large datasets and high dimensional datasets The model may become overly complex or overfit the data if the basis functions not chosen correctly Susceptible to local minima The choice of radial basis functions is fixed which limits its flexibility

Table 1.1: (cont'd)

Learning and Modeling Technique	General Applications	Reviewed Water Treatment and Monitoring Applications	Advantages	Disadvantages
Adaptive Neuro-Fuzzy Inference Systems (ANFIS)	Regression, Classification Farhoudi et al. [2010], Karaboga and Kaya [2019], Adedeji et al. [2019]	DBP formation modeling Adsorption process removal efficiency modeling Membrane process parameters modeling Dissolved oxygen concentration modeling BOD/COD modeling	Fuzzy logic components of ANFIS allow for greater interpretability of the model ANFIS is capable of modeling complex nonlinear relationships between inputs and outputs, making it suitable for a wide range of applications ANFIS models are generally robust to noise and uncertainties in the data	ANFIS models can be complex, with many parameters to tune The training process of ANFIS can be computationally intensive and time consuming ANFIS model is prone to overfitting the data ANFIS may not scale well to large or high dimensional datasets The performance of ANFIS can be sensitive to the initial settings of the membership functions and rule base

Table 1.1: (cont'd)

Learning and Modeling Technique	General Applications	Reviewed Water Treatment and Monitoring Applications	Advantages	Disadvantages
Extreme Learning Machine (ELM)	Regression, Classification Zhu et al. [2005], Huang et al. [2004b]	Dissolved oxygen concentration modeling	Relatively short training times Suitable for pattern classifications	Often faces over fitting or under fitting if too many/few hidden nodes are utilized
Boltzmann Machines	Unsupervised learning Optimization, system control Demertzis et al. [2022], Harrou et al. [2018]	Wastewater treatment process modeling water treatment automated anomaly detection	Capable capture complex dependencies between variables Provide a measure of uncertainty for the learned representations Flexible architecture: Boltzmann machines can be adapted and extended to various architectures, such as Restricted Boltzmann Machines (RBMs) and Deep Belief Networks (DBNs)	Learning is slow and computationally intensive Challenge to scale to large datasets and high dimensional problems Learning algorithm can get stuck in local optima Difficult to interpret Outperformed by modern techniques, such as deep learning models

1.3.4 Applications of AI and ML methods in Water Treatment

Chlorination control has been effectively managed using AI methods, while ML models have shown efficacy in modeling DBP concentrations and significant parameters for adsorption and membrane-filtration processes. Commonly used statistical measures for evaluating results include the coefficient of correlation (R), coefficient of determination (R^2), mean average error (MAE), mean square error (MSE), root mean square error (RMSE), and relative error (RE). The following sections provide a brief overview of the applications of AI and ML methods in water treatment.

Chlorination and Disinfection By-Product Estimation

In water and wastewater treatment plants, disinfection is crucial for killing or inactivating microorganisms and viruses, often with chlorine-based disinfectants Li et al. [2017], Xu et al. [2015, 2013]. However, chlorine poses human health hazards and can react with bromide and organic matter to create disinfection by-products (DBPs), which are suspected carcinogens and reproductive disruptors Sedlak and von Gunten [2011], Bull et al. [1995]. DBPs are divided into two subcategories, trihalomethanes (THMs) and haloacetic acids (HAAs), with THMs being the most common form. ML technologies are well-suited for predicting and mitigating DBP formation. AI methods can be used for controlling chlorination. The studies often tested models on surface waters treated with chlorine as the primary disinfectant and noted success in modeling DBP concentrations in treated water distribution networks and at consumer taps Librantz et al. [2018], Godo-Pla et al. [2021], Singh and Gupta [2012], Mahato and Gupta [2022], Park et al. [2018], Lin et al. [2020], Xu et al. [2022], Peleato [2022], Okoji et al. [2022], Cordero et al. [2021]. Common model inputs include water temperature, pH, chlorine concentration, contact time, and TOC/DOC concentrations, as well as other markers such as bromine concentration, UV_{254} , algae and

chlorophyll-a concentrations, and DBP-precursor chemicals.

The most commonly tested ML model for chlorination and DBP prediction is the Artificial Neural Network (ANN), although other models such as support vector machines, fuzzy inference systems, and genetic algorithms have also been used. In comparative studies, ANNs generally outperform GAs and SVMs, although in some cases, SVMs have provided a slight advantage when using R^2 as a comparison metric Wortmann and Flüchter [2015], Imo et al. [2007]. Researchers have modeled and predicted common DBPs, such as total trihalomethanes (TTHM) and total haloacetic acids (THAA), as well as specific DBP compounds including dichloroacetic acid (DCAA), trichloroacetic acid (TCAA), bromochloroacetic acid (BCAA), HAA5, HAA9, trichloromethane (TCM), bromodichloromethane (BDCM), and dibromochloromethane (DBCM). Statistical model validation numbers did not show significant differences in predictions for TTHMs or THAAs versus their individual compounds.

Table 1.2: Disinfection by-products (DBP) formation prediction by ML models.

Target Compounds	Water Source	Disinfectants	AI/ML Technique Used	Input Variables	Output	Year
Total trihalomethanes (TTHMs)	Surface water	Chlorine	Artificial neural network (ANN), support vector machine (SVM), and gene expression programming (GEP) modeling	Dissolved organic carbon normalized chlorine dose, water pH, temperature, bromide concentration, and contact time	TTHM effluent concentration Singh and Gupta [2012]	2012
TTHM	Tap water	Chlorine	Artificial neural network and support vector machine	Temperature, pH, residual chlorine, TOC, UV_{254}	TTHM effluent concentration Mahato and Gupta [2022]	2022
Haloacetic acids (HAAs)	Tap water	Chlorine	RBF-ANN, linear/log linear regression (MLR) models	Dissolved organic carbon (DOC), UVA ₂₅₄ , bromine concentration, temperature, pH, Cl_2 concentration, NO_2^- concentration, NH_4^+ concentration, N concentration	DBP concentration tap Lin et al. [2020]	2020

Table 1.2: (cont'd)

Target Compounds	Water Source	Disinfectants	AI/ML Technique Used	Input Variables	Output	Year
TTHM	Tap water	Chlorine	Ion artificial neural network (RBF ANN), Hybrid method of RBF ANN and grey relational analysis (GRA)	Temperature, pH, UV absorbance at 254 (UVA ₂₅₄), dissolved organic carbon, bromide, residual free chlorine, nitrite and ammonia	Trichloromethane (TCM), bromodichloromethane (BDCM) and total-THMs (T-THMs) Hong et al. [2020]	2020
TTHMs, Sum of trichloromethane (TCM), BDCM	Tap water	Chlorine	Linear/log linear regression models (LRM) and radial basis function artificial neural network (RBF ANN)	pH, temperature, $UV A_{254}$, Cl_2 concentration	DBP concentration tap Xu et al. [2022]	2020
TTHMs	Tap water	Chlorine	Classification trees	Fluorescence spectra	Dichloroacetonitrile (DCAN), trichloropropanone (TCP), trichloronitromethane (TCNM) Bergman et al. [2016]	2016
TTHMs, HAAs	Tap water	Peroxide (Ozone), Chlorine	CNN	Fluorescence spectra	DBP effluent concentration Peleato [2022]	2022

Table 1.2: (cont'd)

Target Compounds	Water Source	Disinfectants	AI/ML Technique Used	Input Variables	Output	Year
TTHMs, TCM, BDCM, DBCM	Tap water	Chlorine	Adaptive neuro-fuzzy inference system (ANFIS)	Temperature, pH, UVA254, residual chlorine concentration, dissolved organic carbon	DBP effluent concentration Okoji et al. [2022]	
Trihalomethanes (THMs)	Tap water	Chlorine	Least-square Boost (LSBoost), XGBoost, and Random forest	Chlorine dose/DOC, reaction time, pH, bromide concentration, and temperature	THM concentration Sikder et al. [2023]	2023
DCAN, TCP, TCNM	Tap water	Chlorine	Generalized regression neural network (GRNN)	Tempaerature, total residual chlorine, dissolve organic chlorine, turbidity, pH, conductivity, absorbance, TCM, BDCM, DBCM, DCAA, TCAA	DCAN, TCP, TCNM Mian et al. [2021]	2021

Table 1.2: (cont'd)

Target Compounds	Water Source	Disinfectants	AI/ML Technique Used	Input Variables	Output	Year
Chlorine dose and free residual chlorine (FRC) set point	Surface water	Chlorine	ANN	Reservoir set-point output, FRC of treated water tank, FRC output of WTP (mg/L), WTP production flow rate, compensating system flow rate, dosage error	Chlorine dosage, WTP FRC set point Librantz et al. [2018]	2018
DCAN, chloropicrin, and TCP	Small water distribution networks (SWDNs)	Chlorine	Multivariate linear regression-based model, regression tree-based model, neural networks-based model and advanced non-parametric regression model	Water quality parameters measured in the samples include water temperature, total residual chlorine, dissolved organic carbon, turbidity, pH, conductivity, and ultraviolet absorbance at 254 nm (UV254)	DCAN, chloropicrin (CPK) and TCP Hu et al. [2023]	2023

Table 1.2: (cont'd)

Target Compounds	Water Source	Disinfectants	AI/ML Technique Used	Input Variables	Output	Year
Chlorine	Surface water	Chlorine	FIS	Inflow rate, Raw water total organic carbon (TOC), Raw turbidity, conductivity, temperature, Raw water UV absorbance	Free chlorine and chlorine dioxide dose Godo-Pla et al. [2021]	2021
Haloacetic acids (HAAs), trichloroacetic acid (TCAA), dichloroacetic acid (DCAA)	Lab synthesized	Chlorine	Support vector regressor, random forest regressor, and multilayer perceptron regressor	Number of aromatic bonds, hydrophilicity, electrotopological descriptors related to electrostatic interactions, and atomic distribution of electronegativity, geometry, ionization potential, steric effects, and acid-base interactions et al.	DBP effluent concentration Cordero et al. [2021]	2021

Adsorption Processes

Adsorption processes are a crucial physical and chemical treatment option for removing various contaminants in the water and wastewater treatment industries. These processes transfer target molecules from fluids to solid surfaces, known as adsorbents or sorptive media. Due to the complex interactions involved in the process, it can be challenging to determine the adsorption parameters and ultimate removals accurately Karri et al. [2020], Vinayagam et al. [2022]. Predictive models using ML can optimize the adsorption process and extend the media's life, increasing the plant's effectiveness and confidence in meeting applicable regulations. Studies have modeled adsorption processes with water streams contaminated with metals, industrial dyes, and organic compounds using various adsorbent media, including carbonaceous materials and metal-based nanocomposites Bhagat et al. [2021], Mazloom et al. [2020], Mesellem et al. [2021a], Al-Yaari et al. [2022], Mazaheri et al. [2017], Ahmad et al. [2020], Fawzy et al. [2016], Ullah et al. [2020], Mahmoud et al. [2019], Mesellem et al. [2021b]. Common inputs for modeling adsorption processes include pH, water temperature, adsorbent dose, contact time, and initial adsorbate concentration. Other models have used parameters such as adsorbent particle size, system flow rate, agitation speed, bed height, and BET surface area, among others. The published studies mostly focused on adsorbate percentage removal, while some models predicted adsorption capacity, non-dimensional effluent concentrations, and the relative importance of input water-quality parameters. These models have the potential to support operator decisions and improve the efficiency of the adsorption process.

ANN was the most commonly used ML model in studies involving metal, organic, and industrial-dye contaminants, while ANFIS, SVM, and RF were also studied with notable success. These models generally achieved R_2 values greater than 0.9 and

sometimes greater than 0.99 Bhagat et al. [2021], Mazloom et al. [2020], Mohammadi et al. [2019]. SVM models performed slightly better than ANN models in most cases, producing R_2 and RMSE values with better statistical value. However, in one case, the optimized ANFIS model performed poorly compared to other successful models, with an $R = 0.813$, and was noted as the worst performing model in a comparison between ANN, ANFIS, and SVM models Mesellem et al. [2021a]. In another case, the ANFIS model achieved adequate performance with an R_2 of 0.9333 Al-Yaari et al. [2022].

Table 1.3: Adsorption processes and removal rates prediction by ML models.

Adsorbate	Adsorbent	ML Technique Used	Input Variables	Output	Year
As (III)	Nanosized iron-oxide-immobilized graphene oxide gadolinium oxide (Fe-GO-Gd)	Artificial neural network (ANN)	Initial concentration, adsorbent dosage, pH, and residence time	As percent removal Maurya et al. [2022]	2022
As (III)	A variety of absorbents or biosorbents	Adaptive network-based fuzzy inference system (ANFIS)	pH, As initial concentration, contact time, adsorbent dosage, inoculum size, and temperature, agitation speed, flow rate	Adsorbate percent removal Al-Yaari et al. [2022]	2022
Copper ions	Attapulgate clay	Grid optimization-based random forest (Grid-RF), artificial neural network (ANN) and support vector machine (SVM)	Initial concentration of Cu (IC), the dosage of Attapulgate clay (Dose), contact time (CT), pH, and addition of NaNO ₃	Adsorbate percent removal Bhagat et al. [2021]	2021

Table 1.3: (cont'd)

Adsorbate	Adsorbent	ML Technique Used	Input Variables	Output	Year
As (III, IV)	Biochar	Random forest algorithm	Contents of ash, carbon, hydrogen, oxygen, nitrogen, sulfur, and iron, H/C atomic ratio, O/C atomic ratio, (O + N)/C atomic ratio, and specific surface area (SBET), As species (arsenite or arsenate), initial concentration (CAs), adsorption conditions, reaction temperature, solution pH, adsorbent dosage	As adsorption capacity Liu et al. [2023]	2021
Asphaltenes	Nickle(II) Oxide Nanocomposites	Group Method of Data Handling (GMDH), ANN, Least Squares Support Vector Machine (LSSVM)	BET surface area and volume of micropores of nanocomposite, pH, amount of nanocomposites over asphaltenes initial concentration (D/C_0), temperature	Adsorbate percent removal Mazloom et al. [2020]	2020
Various organic pollutants	Activated carbon	Artificial Neural Networks (ANNs), Support Vector Machines (SVMs) and Adaptive Neuro-Fuzzy Inference System (ANFIS)	Molar mass, initial concentration, flow rate, bed height, BET surface area, time and concentration of non-dimensional effluents	Non-dimensional effluent concentration Mesellem et al. [2021a]	2021

Table 1.3: (cont'd)

Adsorbate	Adsorbent	ML Technique Used	Input Variables	Output	Year
Methylene blue (MB), Cd(II)	Natural walnut activated carbon	Boosted regression trees (BRTs), artificial neural network (ANN) and response surface methodology (RSM)	Stirring time, pH, adsorbent mass, MB concentration, Cd(II) concentration,	Adsorbate percent removal Mazaheri et al. [2017]	2017
Methylene blue (MB)	Graphite oxide (GO) nano	ANN	Solution pH, initial dye concentration, contact time and adsorbent dosage	Methylene blue removal efficiency Ghaedi et al. [2014]	2014
Sunset yellow (SY)	Neodymium(III) chloride modified order mesoporous carbon (OMC)	ANN	Initial concentration, reaction time, and adsorbent dosage	SY removal efficiency Ahmad et al. [2020]	2020
Ni(II), Cd(II)	Typha domingensis (Cattail) biomass	Adaptive neuro-fuzzy inference system (ANFIS)	Initial pH, bioadsorbent dosage, initial metal-ions concentration, contact time, biosorbent particle size	Metal-ions removal efficiency Fawzy et al. [2016]	2016
Zn(II)	Low-cost adsorbents produced from rice husks	ANN	Contact time, initial concentration and the applied temperature	Adsorption capacity Ullah et al. [2020]	2020

Table 1.3: (cont'd)

Adsorbate	Adsorbent	ML Technique Used	Input Variables	Output	Year
Phosphate	Encapsulated nanoscale zero-valent iron	ANN	Initial pH, initial PO_4^{3-} concentration, adsorbent dose, contact time, stirring rate	Adsorbate percent removal Mahmoud et al. [2019]	2018
Systems organic pollutants	Activated carbon	ANN	Molar mass of target contaminant, initial concentration, flow rate, bed height, particle diameter, BET surface area, average pore diameter, time, concentration of dimensionless effluents	Non-dimensional effluent concentration Mesellem et al. [2021b]	2021
Pb (II)	Magnetic ash/graphene oxide (GO) nanocomposites	ANN	Initial Pb ion concentration, temperature	Adsorption capacity Zeng et al. [2022]	2021
Pb (II), Cd (II)	Composite of metal organic framework and layered double hydroxide	ANN	Type of ions (Pb, Cd) and time	Adsorption capacity Wei et al. [2021]	2021
As (III), Cr(VI)	Fibrous zirconium oxide ethylenediamine adipate (ZEDA) hybrid material	Adaptive neuro-fuzzy inference system (ANFIS)	Dose, pH, time, temperature and initial concentration, bed height and flow rate	Removal efficiency Mandal et al. [2015a]	2021

Table 1.3: (cont'd)

Adsorbate	Adsorbent	ML Technique Used	Input Variables	Output	Year
As (III)	Cerium hydroxylamine hydrochloride (Ce-HAHCl) hybrid material	ANN	Adsorbent dose, pH, contact time, initial concentration and contact temperature	Removal efficiency Mandal et al. [2015b]	2015
Cr (IV)	Cerium oxide polyaniline (CeO ₂ /PANI) composite	ANN	Adsorbent dose, time, pH, temperature and initial concentration	Removal efficiency Mandal et al. [2015c]	2015

Membrane-Filtration Processes

Membrane processes separate contaminants in water and wastewater treatment by passing the water through a barrier or filter using high-pressure differentials. These processes are typically used for contaminants that are difficult or costly to remove by chemical or physical means or require a high level of removal that cannot be achieved by other means. Microfiltration, ultrafiltration, nanofiltration, and reverse osmosis are the most commonly used membrane processes Hube et al. [2020], Pronk et al. [2019]. These models have been used with microfiltration, ultrafiltration, nanofiltration, reverse osmosis, and submerged membrane bioreactors to treat various water sources contaminated with pollutants and natural compounds such as petroleum, natural organic matter, industrial and pharmaceutical wastes, and saltwater Zoubeik et al. [2019], Fetanat et al. [2021], Khan et al. [2022], Yusof et al. [2020], Nazif et al. [2020], Shim et al. [2021], Ammi et al. [2021a]. ANN is the most dominant model used, although ANFIS, SVM, and specific forms of ANNs, including RNNs that utilize LSTM, have also been used for membrane-filtration-process modeling.

ML techniques for modeling membrane-filtration processes aim to output several variables, such as transmembrane pressure, permeate flux, and solute rejection. Inputs in published studies include pH, temperature, contact/filtration time, transmembrane pressure, and flux rate, among others. Due to the wide range of models testing for different parameters, it is difficult to make a full statistical comparison of the values obtained in these studies. However, ANN, RNN, and SVM models consistently performed well, achieving R^2 values greater than 0.9 and often greater than 0.99 Zoubeik et al. [2019], Khan et al. [2022], Yangali-Quintanilla et al. [2009] (Table 1.4).

Table 1.4: Membrane-filtration parameters prediction by ML models.

Membrane Type	Water Source	ML Technique Used	Input Variables	Output	Year
Titanium-based ceramic ultrafiltration	Petroleum production wastewater	ANN, ANFIS, RBF-ANN	Transmembrane pressure (TMP), crossflow velocity (CFV), temperature, pH and time	Permeate flux Zoubeik et al. [2019]	2019
Aluminum oxide microfiltration (MF) membrane	Various water types	Hermia model, ANN	Temperature, pH, crossflow velocity (CFV), and transmembrane pressure (TMP)	Permeate flux Zoubeik et al. [2022]	2022
Nanolayered double hydroxide decorated thin-film nanocomposite membrane	Various water types	ANN-GA	Nanolayered double hydroxide (NLDH), polyvinylpyrrolidone (PVP, MW = 29 000 g/mol) and polymer concentrations.	Pure water flux, protein flux and flux recovery ratio Arefi-Oskoui et al. [2017]	2017
Nanocomposite membranes	Various	ANN	Polymer concentration, polymer type, filler concentration, average filler size, solvent concentration (in the dope solution), solvent type, and contact angle	Solute rejection, flux recovery, and pure water flux Fetanat et al. [2021]	2021
Oscillating slotted membrane	Dilute suspension mixture of crude oil, dilute suspension mixture of tween-20	ANN	Permeate flux, shear rate, filtration time	Transmembrane pressure (TMP) Khan et al. [2022]	2022

Table 1.4: (cont'd)

Membrane Type	Water Source	ML Technique Used	Input Variables	Output	Year
Submerged membrane bioreactor	Palm oil mill effluent	RNN, nonlinear auto-regressive model	Pump voltage, airflow, transmembrane pressure OR flux	Permeate flux, transmembrane pressure (TMP) Yusof et al. [2020]	2019
Submerged membrane bioreactor (MBR) filtration system	Waste water	Feedforward neural network (FFNN), radial basis function neural network (RBFNN) and nonlinear autoregressive exogenous neural network (NARXNN)	Permeate pump voltage	Permeate flux and transmembrane pressure Mahmod et al. [2020]	2020
Reverse osmosis membrane (BW30-400)	Ground water and surface water	General regression neural network (GRNN)	Membrane operating period, time interval between consequent cleanings, water temperature, input concentration, inflow, inlet pressure of the compartments, recovery	Pressure drop (PD), salt passage (SP) Nazif et al. [2020]	2020
Reverse osmosis	Municipal wastewater	ANNs, Random forest, multiple linear regression models	Pressure, flow rate, temperature, conductivity, ORP, turbidity, dissolved organic carbon (COD), TDS	Salt passage, permeate flow rate Odabaşı et al. [2022]	2022

Table 1.4: (cont'd)

Membrane Type	Water Source	ML Technique Used	Input Variables	Output	Year
Nanofiltration system	Surface water with natural organic matter	Long short-term memory (LSTM) model	Operation time, pressure, initial permeate flux, dissolved organic carbon (DOC), modified FRI, optical coherence tomography (OCT) images	Permeate flux (PF), fouling layer thickness (FLT) Shim et al. [2021]	2021
Organic solvent nanofiltration (OSN)	Various water types	Support vector machine (SVM), boosted tree (BT), and artificial neural network (ANN)	Substrate type, nanoparticle type, nanoparticle size, nanoparticle loading, amine monomer type, amine concentration, chloride monomer type, chloride concentration, water contact angle, surface roughness, organic solvent type, solvent properties (molecular weight, viscosity, density and molar volume), solute type, solute concentration, solute charge and solute molecular weight	Relative permeability (RP) and relative selectivity (RS) Wang et al. [2023]	2023

Table 1.4: (cont'd)

Membrane Type	Water Source	ML Technique Used	Input Variables	Output	Year
Nanofiltration, reverse osmosis membranes	Pharmaceutical wastewater	ANN, SVM	Anti-inflammatory drug properties (logD, dipole moment, the effective diameter of the organic compound in water "dc", molecular length, and molecular equivalent width "eqwidth"); membrane characteristics (molecular weight cutoff "MWCO", sodium chloride salt rejection "SR (NaCl)", zeta potential, and contact angle); and filtration conditions (pH, pressure, temperature, and recovery)	Rejection percentage of the target compound Ammi et al. [2021a]	2021
Polyamide-based thin film composite (TFC) FO membrane	Effluent from primary treatment plant	ANN, SVM	Organic matters, sodium ion, and calcium ion concentrations	Water flux, membrane fouling, and removal efficiencies Im et al. [2022]	2022

Table 1.4: (cont'd)

Membrane Type	Water Source	ML Used	Technique	Input Variables	Output	Year
Polyamide nanofiltration (NF) and reverse osmosis (RO) membrane	Various water types	ANN		Molecular weight (MW), log Kow, dipole moment, molar volume, molecular length, molecular width, molecular depth, equivalent width; membrane characteristics: molecular weight cut-off (MWCO), pure water permeability, magnesium sulphate salt rejection (SR), surface membrane charge (as zeta potential), and hydrophobicity (as contact angle); operating conditions: operating pressure and permeate flux	rejection of neutral organic compounds Yangali-Quintanil et al. [2009]	2009

Table 1.4: (cont'd)

Membrane Type	Water Source	ML Technique Used	Input Variables	Output	Year
Nanofiltration (NF) and reverse osmosis (RO) membrane	Domestic wastewater	Quantitative structure-activity relationship (single neural networks "QSAR-SNN" and bootstrap aggregated neural networks "QSAR-BANN")	Pharmaceutical active compound properties (hydrophobicity "logD", dipole moment, the effective diameter of organic compound in water "dc", molecular length, and molecular equivalent width "eqwidth"); membrane characteristics (molecular weight cut-off "MWCO", sodium chloride salt rejection "SR (NaCl)", zeta potential, and contact angle); and filtration conditions (pH, pressure, temperature, and recovery)	Removal efficiency Ammi et al. [2021b]	2021
Nanofiltration and reverse osmosis membranes	Various water types	Bootstrap aggregated neural networks (BANN)	Molecular weight, ratio of the equilibrium concentration (logD), dipole moment, length, eqwidth, SR (NaCl), zeta potential, contact angle, pH, pressure, recovery, temperature	Uncharged organic compounds rejection Khaouane et al. [2017]	2017

Table 1.4: (cont'd)

Membrane Type	Water Source	ML Used	Technique	Input Variables	Output	Year
Nanofiltration and reverse osmosis membranes	Various water types	ANN		Molecular weight, ratio of the equilibrium concentration (logD), dipole moment, length, eqwidth, membrane molecular weight cutoff (MWCO)/pore size MWCO, SR (NaCl), zeta potential, contact angle, pH, pressure, recovery, temperature	Uncharged organic compounds rejection Ammi et al. [2015]	2015
Nanofiltration and reverse osmosis	Various water types	Single neural networks (SNN) and bootstrap aggregated neural networks (BANN)		Molecular weight, molecular effective diameter "dc", ratio of the equilibrium concentration (logD), dipole moment, length, eqwidth, membrane molecular weight cutoff (MWCO)/pore size MWCO, SR (NaCl), SR (MgSO ₄), zeta potential, contact angle, pH, pressure, recovery, temperature	Removal efficiency Ammi et al. [2018]	2018

Table 1.4: (cont'd)

Membrane Type	Water Source	ML Technique Used	Input Variables	Output	Year
Nanofiltration and reverse osmosis	Various water types	Random forest, neural network models	Molecular class, molecular weight, The octanol/water partition coefficient (log Kow), partition coefficient (logD), dipole moment, length, eqwidth, depth, equivalent length, membrane type, molecular weight cutoff (MWCO)/pore size MWCO, zeta potential, contact angle, pH, pressure, recovery, pH, operating pressure, recovery, salt rejection SR (MgSO₄)	Membrane Rejection Le e and Kim [2020]	2020

CHAPTER 2

Tap water fingerprinting using a convolutional neural network built from images of the coffee-ring effect

2.1 Abstract

A low-cost tap water fingerprinting technique was evaluated using the coffee-ring effect, a phenomenon by which tap water droplets leave distinguishable “fingerprint” residue patterns after water evaporates. Tap waters from communities across southern Michigan dried on aluminum and photographed with a cell phone camera and 30x loupe produced unique and reproducible images. A convolutional neural network (CNN) model was trained using the images from the Michigan tap waters, and despite the small size of the image dataset, the model assigned images into groups with similar water chemistry with 80% accuracy. Synthetic solutions containing only the majority species measured in Detroit, Lansing, and Michigan State University tap waters did not display the same residue patterns as collected waters; thus, the lower concentration species also influence the tap water “fingerprint”. Residue pattern images from salt mixtures with an array of sodium, calcium, magnesium, chloride, bicarbonate, and sulfate concentrations were analyzed by measuring features observed in the photographs as well as using principal component analysis (PCA) on the image files and particles measurements. These analyses together highlighted differences in the residue patterns associated with the water chemistry in the sample. The results of these experiments suggest that the unique and reproducible residue patterns of tap water samples that can be imaged with a cell phone camera and a loupe contain a wealth of information about the overall composition of the tap water, and thus, the phenomenon should be further explored for potential use in low-cost tap water

fingerprinting.

2.2 Introduction

Need for innovation in drinking water monitoring

With tap water crisis events that continue to occur in both developed and developing nations, the desire for low-cost tap water testing that is practical for application by citizens is high. When a teacher, student, household, or community member would like to test their tap water, they are faced with single use paper test strips, probes, standard analytical methods for measuring water quality, or water testing fees for hundreds or even thousands of different water quality parameters. Challenges exist in choosing which water constituents to test and which methods to apply, both of which can be difficult since there is little to no tap water education in typical K-12 and university systems. In this work, experiments were conducted to determine if the coffee-ring effect, precipitation reactions, and convolutional neural networks (CNN) could be harnessed for low-cost “fingerprinting” of tap water samples as a whole, rather than measuring one contaminant at a time.

How does the coffee-ring effect work

The coffee-ring effect offers low-cost separation of particles in aqueous samples due to the physics of water droplet drying on hydrophobic substrates. This phenomenon occurs when water evaporates evenly from a water droplet surface with a pinned diameter, such that the droplet shrinks in height while the diameter remains constant Wong et al. [2011], Deegan et al. [1997]. The shrinking height of the droplet correlates to a decrease in contact angle at the pinned surface through droplet drying, squishing particles into concentric circles by size Wong et al. [2011]. The phenomenon was termed nanochromatography after separation resolutions on the order 100 nm were demonstrated for mixtures of fluorescently labeled antibodies, B-lymphoma cells, and E. coli at particle

volume fractions on the order of $<0.04\%$ Wong et al. [2011]. Force balance analysis suggests nanoscale separation is possible for low particle volume fractions due to the difference in the magnitude of adhesion versus surface tension forces for large (1 μm) and small (40 nm) particles at the drop edge, where surface tension forces move particles towards the center of the drop and substrate-particle adhesion forces hold particles in place.

Most existing studies on the coffee-ring effect have been conducted on particles or biological molecules, sometimes in buffer solutions or biofluids where particle-like species deposit on the outer edge forming concentric rings of particles separated by size and soluble salts deposit throughout the center of the drop (Figure. 2.1). Particles within a drop are known to deposit on the outer edge when the fluid flow that delivers particles to the drop edge is faster than the surface capture effect, the latter which occurs if the concentration of particles at the surface of the droplet is high or if water evaporation is accelerated Li et al. [2016c]. Tap water solutions, however, are composed largely of dissolved ions rather than particles. Within dissolved salt solutions, the majority of the particles observed in the residue patterns must form as water evaporates and increases ion concentrations above solubility limits of their respective salts; however, very little work has been conducted to document the coffee-ring patterns for complex mixtures of salts Shahidzadeh et al. [2015]. It is expected that in mixed salt solutions both the coffee-ring effect and the fundamental characteristics of the salts that form will control the location, sizes, and shapes of each salt in the resulting residue pattern, with the least soluble salts that form particles quickly separated by size at the drop edge. Thus, features such as the sizes, shapes, colors, quantity, and location of particles within the coffee-ring residue of a water sample are expected to correlate to water chemistry. The coffee-ring

effect has previously been partnered with Raman spectroscopy to quantify cyanotoxins in environmental water, signs of ocular damage in human tear fluid, and osteoarthritis determinants in knee fluid; however, the patterns produced due to the coffee-ring effect have not been harnessed without expensive chemical analysis instrumentation to record composition of the deposited residues.

Image analysis via convolutional neural network (CNN)

Machine learning methods, especially deep learning artificial neural networks (ANNs) are increasing in popularity in research and engineering to solve problems that are challenging to solve with traditional analysis techniques. Convolutional neural networks (CNNs) have been widely tested and successfully used for image analysis, especially in segmentation problems, such as differentiating between an object and the background. With the development of more advanced CNN architectures (e.g., CNN models involving more layers, new activation functions, more options for objective functions to calculate error, more sophisticated model structures) and use of graphics processing units with higher computational speeds, CNNs are being developed to analyze a growing variety of data types, including medical images, electron microscopy images, cal structures. For example, CNN models have proven the ability to identify brain tumors in magnetic resonance images (MRI) faster and more accurately than the state of the art tools and can identify the pancreas in computerized tomography (CT) images, both of which are challenging analysis problems because of anatomical variability. In chemistry, CNN models are being trained using 2D and 3D images of molecular structure for quantitative structure-activity relationship (QSAR) modeling to predict toxicity Matsuzaka and Uesawa [2019] and to predict therapeutic use classes of drugs Meyer et al. [2019]. CNN models have also been trained to assign surface-enhanced Raman spectroscopy (SERS) spectra to classes of

metabolites and to assign bundles of SERS spectra (8 x 8 pixel hyperspectral images) to the concentration of rhodamine 800 dye at femtomolar concentrations for single molecule detection Lussier et al. [2019], Thrift and Ragan [2019]. Additional applications include identifying the types and positions of defect structures in silicon doped graphene from unprocessed scanning transmission electron microscopy images, predicting chemical reactivity, and diagnosing faults in the chemical process industry. Limitations of CNNs include the computational cost of model training, the sensitivity of classification to unbalanced datasets (unequal numbers of samples in different classes can result in poor model performance), and the necessity of experienced users to modify model structure and tune parameters for every individual CNN application. However, the accuracy of classification results observed and the wide variety of cases in which it can be applied ensures use of CNN will continue to grow.

The goal of this research was to determine if the residue patterns of tap water samples imaged with a cell phone camera and loupe were sufficiently reproducible, sensitive, and correlated to water chemistry to be valuable for low-cost analyses. Specific objectives were to create a library of images of residue patterns for real and synthetic tap waters, determine if the residue patterns were reproducible for a given water chemistry, document the response of the fingerprint to changes in composition of majority species (sodium, calcium, magnesium, chlorine, bicarbonate, sulfate), and apply machine learning image analysis techniques to differentiate between residue patterns. These objectives were met by photographing residue patterns for a variety of collected tap water solutions and increasingly complex synthetic water solutions with a cell phone camera through a jeweler's loupe, measuring features observed in residue patterns, and correlating residue features to water chemistry, and creating a CNN to classify residue pattern images to

groups with similar water chemistry.

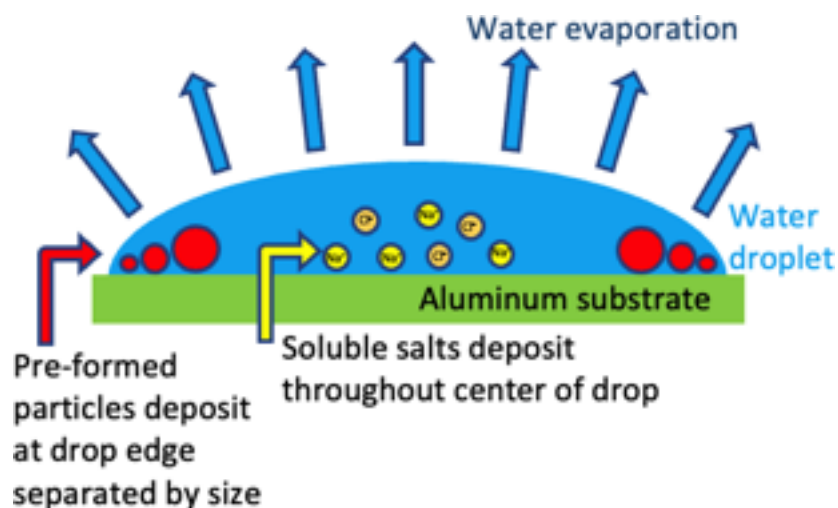


Figure 2.1: Nanoscale separation of particles within a drying droplet is provided by the phenomenon known as the coffee-ring effect.

2.3 Experimental

Water samples

Thirty tap water samples were collected from communities across southern Michigan, utilizing a variety of water treatment systems (Table. 2.1, Table. 6.1). One liter of each water sample was collected in a hydrochloric acid washed polypropylene bottle from the water supply at a public park, community center, or city building water fountain or restroom tap. Samples were stored at 4 °C until analysis using the coffee-ring effect and standard methods. Samples were not filtered before measurement. Conductivity was measured by a Hach HQ40D portable conductivity meter and intelliCAL™ CDC401 standard conductivity probe, and pH was measured with a Orion Star A211 pH meter and Orion 8135BNUWP Ross Ultra Fast pH probe (Thermo Scientific). Chlorine, sulfate, phosphate, fluoride, bromide, and nitrate concentrations were measured by ion chromatography with a Dionex series 2000i/sp instrument. Bicarbonate was measured by titration to pH of 4.5 using standard method 2320.28 Metals were measured by Varian 710-

ES Axial ICP-OES and samples were digested by nitric acid using standard method 3030 E. One replicate sample was measured for every ten samples, and values that deviated from expected (from annual municipal water quality reports or previous measurements) were repeated.

Table 2.1: Measured water quality data for tap water samples collected across Michigan and treatment information from annual municipal water quality reports and system operators. Averages and standard deviations are listed for values conducted in replica.

City	Water treatment	pH	Cond <i>uS/cm</i>	Na^+ mM	Ca^{2+} mM	Mg^{2+} mM	K^+ mM	Cl^- mM	SO_4^{2-} mM	HCO_3^- mM	PO_4^{3-} mM	Cu mM	Fe mM
MSU, academic hall	Chlorine, fluoride, phosphate, sodium hydroxide	6.96	823	1.08	2.24	1.54	0.041	0.91	0.92	6.94	0.01	6.1×10^{-3}	2.2×10^{-2}
Durand	Iron remove filters, chlorine	6.72	388	0.31	0.16	0.11	0.075	1.10	0.47	4.88	0.02	1.6×10^{-3}	2.4×10^{-3}
Kalamazoo	Chlorine, fluoride, and phosphate	8.52	976	3.17	1.06	1.29	0.06	3.11	0.39	6.23	0.01	1.2×10^{-3}	4.1×10^{-3}
Portland	Chlorine, phosphate	6.94	909	0.76	0.53	2.86	0.109	0.05	0.12	7.51	BD	1.1×10^{-3}	1.1×10^{-3}
Battle Creek Site A	Chlorine, fluoride, and phosphate	7.22	673	1.60	1.77	1.04	0.035	1.16	0.50	5.47	0.02	4.0×10^{-3}	7.9×10^{-4}
Battle Creek Site B	Chlorine, fluoride and phosphate	7.22	673	1.60	1.77	1.04	0.035	1.16	0.50	5.47	0.02	8.9×10^{-3}	2.0×10^{-2}
Charlotte	Chlorine, phosphate	7.01 ± 0.29	1215 ± 23	3.79	2.53	3.32	0.252	4.10	0.54	6.89	0.02	3.9×10^{-4}	4.4×10^{-3}

Table 2.1: (cont'd)

City	Water treatment	pH	Cond <i>uS/cm</i>	Na^+ mM	Ca^{2+} mM	Mg^{2+} mM	K^+ mM	Cl^- mM	SO_4^{2-} mM	HCO_3^- mM	PO_4^{3-} mM	Cu mM	Fe mM
Fowlerville	Chlorine, phosphate	7.14	978	4.63	1.10	0.91	0.158	3.53	0.24	6.07	0.01	7.8×10^{-4}	9.4×10^{-3}
Lansing site A	Lime softening	8.70	609	4.29	0.55	0.56	0.082	2.33	1.34	0.99	0.01	3.1×10^{-4}	2.5×10^{-3}
Lansing site B	Lime softening	7.04	535	3.79	0.63	0.49	0.079	1.91	1.16	0.83	0.01	1.4×10^{-3}	5.9×10^{-4}
East Lansing	Lime, ferric fluoride, filtration, chloramine, fluoride, phosphate	6.61	361	1.43	0.58	0.56	0.063	1.10	0.50	1.39	0.01	1.8×10^{-3}	5.3×10^{-3}
Howell	Lime softening	8.15	453	2.76	0.55	0.54	0.092	1.83	0.62	1.29	0.01	6.9×10^{-4}	6.6×10^{-3}
MSU residence hall	Iron exchange, chlorine, fluoride, phosphate, sodium, hydroxide	7.34	880	19.57	0.07	0.04	0.025	1.16	0.84	7.09	0.01	1.3×10^{-3}	2.3×10^{-2}
Williamston	Iron removal, softening, chlorine, phosphate	7.51	710	6.02	0.99	0.53	0.075	0.93	0.43	6.83	0.02	1.0×10^{-2}	6.4×10^{-4}

Table 2.1: (cont'd)

City	Water treatment	pH	Cond <i>uS/cm</i>	Na^+ mM	Ca^{2+} mM	Mg^{2+} mM	K^+ mM	Cl^- mM	SO_4^{2-} mM	HCO_3^- mM	PO_4^{3-} mM	Cu mM	Fe mM
Genoa Twp Soft	Household water softener, private well	7.04± 0.23	1920± 30	18.65± 0.47	0.20± 0.015	0.20± 0.035	0.03± 0.025	9.7± 0.3	0.61	8.55	BD	8.1× 10 ⁻⁴	BD
Genoa Twp, Untreated	Private well, untreated	7.24	1940	6.69	3.81	1.98	0.12	11.16	0.60	8.26	BD	4.5× 10 ⁻⁴	4.7× 10 ⁻²
Rest stop, Okemos	Chlorine if bacteria found	7.36	516	3.08	1.41	0.46	0.141	0.09	0.15	6.19	BD	3.4× 10 ⁻⁴	1.7× 10 ⁻³
Rest stop, Zeeland	Chlorine if bacteria found	7.05	560	3.35	1.04	0.82	0.085	0.79	0.21	5.38	BD	2.7× 10 ⁻⁴	9.3× 10 ⁻³
Rest stop, I96/M66	Chlorine if bacteria found	7.07	546	1.22	1.76	1.19	0.029	0.05	0.12	6.86	BD	2.5× 10 ⁻⁴	4.0× 10 ⁻²
Rest stop Fenton	Chlorine if bacteria found	6.96	606	2.71	1.10	1.21	0.090	1.20	0.14	5.64	BD	4.1× 10 ⁻³	1.3× 10 ⁻²
Allegan	Reverse osmosis	6.53	295	1.41	0.73	0.52	0.019	0.63	0.17	2.51	0.02	1.8× 10 ⁻⁴	6.0× 10 ⁻⁴

Table 2.1: (cont'd)

City	Water treatment	pH	Cond <i>uS/cm</i>	Na^+ mM	Ca^{2+} mM	Mg^{2+} mM	K^+ mM	Cl^- mM	SO_4^{2-} mM	HCO_3^- mM	PO_4^{3-} mM	Cu mM	Fe mM
Genoa Twp RO	Reverse osmosis of private well after household water softener	6.64	264	3.23	0.08	0.02	0.006	1.27	0.11	1.37	BD	4.8×10^{-4}	4.8×10^{-4}
Detroit	Great Lakes Water Authority (GLWA), Water Works Park plant	6.21	226	0.43	0.59	0.34	0.023	0.51	0.26	1.55	0.02	1.8×10^{-3}	5.4×10^{-3}
Flint	GLWA, Lake Huron plant	6.86	219	0.32	0.07	0.02	0.022	0.52	0.23	1.64	0.04	4.4×10^{-3}	5.6×10^{-3}
Swartz Creek	GLWA, Lake Huron plant	5.87	209	0.41	0.08	0.03	0.024	0.51	0.23	1.61	0.02	6.9×10^{-4}	4.9×10^{-3}
Grand rapids	Lake Michigan Filtration plant	7.17	304	0.44	0.89	0.26	0.030	0.63	0.33	2.2 ± 0.04	0.02	4.9×10^{-3}	1.9×10^{-2}

Table 2.1: (cont'd)

City	Water treatment	pH	Cond <i>uS/cm</i>	Na^+ mM	Ca^{2+} mM	Mg^{2+} mM	K^+ mM	Cl^- mM	SO_4^{2-} mM	HCO_3^- mM	PO_4^{3-} mM	<i>Cu</i> mM	<i>Fe</i> mM
Holland	Holland Board of Public Works Water Filtration Plant	6.76	302	0.74	0.85	0.51	0.034	0.60	0.29	2.45	BD	3.7×10^{-3}	5.7×10^{-3}
Wyoming	Donald K. Shrine Water Treatment Plant	7.16± 0.03	302±8	1.30± 0.005	0.905± 0.005	0.5 ± 0.001	0.036± 0.002	0.61± 0.01	0.34	2.17	BD	BD	BD

In order to determine the effects of specific ions on residue patterns, synthetic water samples containing various concentrations of the main components in tap water were prepared, including synthetic hard freshwater (192 mg/L NaHCO_3 , 120 mg/L MgSO_4 , 120 mg/L $\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$, and 8 mg/L KCl) and mixtures of NaCl , NaHCO_3 , CaCl_2 , MgCl_2 , CaSO_4 , MgSO_4 , and Na_2SO_4 . Salt mixtures were designed to examine ranges that may be observed in real tap waters; thus, the low and high concentrations tested of every salt do not match. Simplified synthetic tap waters were created to mimic concentrations of calcium, magnesium, sodium, chlorine, sulfate, and total carbonate species observed in tap water. Complex synthetic tap waters also contained phosphate, nitrate, fluoride, copper, and iron. Natural organic matter was not added because larger organic molecules typically deposit on the outer edge of the drop where the organics can't be identified from images alone.

Collection of coffee-ring residue patterns

Two microliter droplets of each water were gently pipetted onto aluminum substrates (6061 with mirror-like finish, McMaster-Carr 1655T1). Substrates from the manufacturer were used directly after peeling off the plastic film that protects the mirror-like finish. Samples were left uncovered for 20-30 minutes or until dry without being moved, touched, or disturbed from the moment of deposition on the slide (Figure. 6.1). Relative humidity in the lab ranged from 47-52% and temperature 23-25 °C over the course of the coffee-ring effect experiments. Samples were imaged with a Samsung S6 cell phone through a Fancii 30× triplet loupe (Amazon.com) with the LED light on (Figure. 2.2). At least five drops were imaged for each sample, and residues that were not round due to lack of pinning to the surface were repeated. Relative humidity and temperature were recorded for each experiment with a Fisher Scientific Traceable Relative Humidity/Temperature Meter (11-

661-13). Reproducibility of water residue patterns was examined by three researchers testing a subset of water samples on several substrates.



Figure 2.2: Tap water fingerprints were captured by drying droplets on aluminum and photographing with a cell phone camera through a loupe.

Image processing, principal component analysis (PCA), and cluster analysis Residue pattern photographs were cropped manually with ImageJ to dimensions of 700 by 600 pixels. Scales bars of 0.5 mm were added in ImageJ using ruler tape captured in photographs as a reference, dimensions of features in residues were measured, and processed images were saved in JPEG format. Images were converted to black and white, noise removed, and particles measured in Matlab software version R2017b (im2bw, medfilt2, and regionprops functions). Principal component analysis (PCA) was conducted on both particle measurements and on the image files themselves using Python version 3.6.4 (matplotlib, numpy, and sklearn packages; Figure. 6.2). Measured water chemistry for each tap water sample was plotted on a trilinear classification diagram using GW_Chart (Version 1.29.0.0, USGS) with samples sorted according to treatment. The cluster analysis algorithm CLARA was used to group samples into six groups using all thirteen of the measured parameters after normalization by subtracting the mean from the measured value and dividing by its standard deviation Liu and Özsu [2009]. The cluster analysis result was visualized in a two dimensional map using the two main components identified by principal

component analysis with the R factoextra package.

Convolutional neural network

A convolutional neural network (CNN) model was created to classify images. Ten residue images from each water sample were used for model training and testing, five of which were from fresh samples and five collected after storage at 4 °C. The first three replicates of each water sample for each condition (fresh and stored) were used for training the model (180 images), and the last two replicates were used for testing the model (120 images). Image pre-processing involved resizing each image from 470 by 470 pixels to 300 by 300 pixels and converting from color to gray-scale (Table. 6.3). The brightness was normalized for each image by dividing the brightness value for each pixel in an RGB channel by the overall sum of the brightness values of all pixels for that RGB channel.

A CNN model was built with two convolutional layers and three fully connected layers in Python (Figure. 6.3). In the first layer eight filters were used to extract pattern features, and sixteen filters were used in the second layer to extract deeper pattern features. After the convolutional layers, three fully connected layers were used to fit the data. The fitting method was a stochastic gradient descent (SGD) with probability calculations through the SoftMax function. The batch size was five for each optimization process. Samples were randomly selected by their weights which were set equal at the beginning but updated after each optimization process by their classification result. The learning rate was 10^{-4} in the model training process. In each iteration, five samples were randomly selected from 180 training samples by their weights with replacement, and every 36 iterations consisted of one epoch. After each epoch, training accuracy, testing accuracy, training loss, and testing loss were calculated. Two hundred epochs were processed for each model and ten independent models were trained. The test dataset accuracies of the

last one hundred epochs and the last epoch model were recorded for analysis.

2.4 Results and discussion

Coffee-ring residue patterns for each Michigan tap water are unique

Michigan State University and the surrounding communities frequently rely on groundwater sources with minimal treatment (chlorine and phosphate, sometimes with fluoride) or hardness removal by lime softening or ion exchange. Rural communities also frequently use on point-of-use or point of entry treatment such as home water softeners or reverse osmosis systems. Many communities near Great Lakes coast-lines utilize surface water sources and conventional treatment. The Great Lakes Water Authority (GLWA) treats and distributes water to a substantial fraction of Michigan's population in the east from Lake Huron or the Detroit River and many communities in the west utilize Lake Michigan. Tap water collected from the sampled Michigan communities displayed a wide range of chemical compositions (Table. 2.1).

The coffee-ring residue patterns for each type of tap water were unique, and waters with similar chemistry displayed similar residue features (Figure. 2.3). Reproducibility was evaluated initially by imaging five droplets of each sample on the same slide, and most residue patterns displayed nearly identical features across replicates (Figure. 6.4). Lime softened water showed variability across replicates, with some samples displaying a thin film of particles across the entire drop and others producing a clearing in the center. A subset of samples were analyzed by three analysts with varying levels of experience. Mirrored aluminum 6061 substrates were chosen due to low cost, availability, compatibility with the loupe and cell phone camera for imaging, and ease of use for inexperienced users; substrates were inspected before use for scratches or defects and only smooth areas without blemishes were used for the coffee-ring effect

experiments. Nanopure water and synthetic hard freshwater were applied as controls. The substrates contained residue remaining from the manufacturer that was captured in images of nanopure water controls (Table. 6.5). A trend was not observed between residue patterns for samples and the residue pattern or lack of residue pattern in the nanopure water controls (Table. 6.5). Tap water samples were tested on multiple substrates to ensure that variation observed in the patterns was not due to the substrate (Tables. 6.6). All analysts produced more consistent data across a single slide than across different slides. Despite variability between substrates, MSU water from academic buildings (hard water) displayed similar patterns on substrates tested across all researchers. Untreated groundwater from the rest stop was characteristically more variable, displaying one of two patterns with a thin film of small particles and either a white ring at the outer edge or a circular segment to one side. Residue patterns for lime softened water from East Lansing were typically consistent across a single slide, but showed two types of patterns with several concentric rings at the drop edge and either a clear center or a thin film of feathery particles across the center surface. Neither the nanopure blank nor synthetic hard freshwater were sufficient to predict which samples would produce thin films of particles for the lime softened water. A similar result was observed for softened Lansing water (Table. 6.5). Synthetic lime softened water may function as a more sensitive positive control for future experiments. Only analyst 1 observed the residue pattern for Detroit with the center scattering of particles concentrated on one side of the drop; this result was attributed to a lab bench at an angle of approximately 1° (Table. 6.5). Residue patterns that displayed variability across substrates were still sufficiently unique from samples with different chemistry to identify what type of drinking water treatment was applied. The results of these experiment suggest that a more uniform substrate and level surface may be

required to reduce variability for applications beyond identifying the tap water source from a library of residue fingerprints. It is well established that the hydrophobicity of the substrate influences the coffee-ring effect Shahidzadeh et al. [2015], Zhang et al. [2003], Ortiz et al. [2004], Zhong et al. [2017]; thus, the substrate used for training datasets must be consistent with that of unknown samples. Additional variables that must be controlled during coffee-ring effect experiments include temperature Li et al. [2016c], Takhistov and Chang [2002], humidity Li et al. [2016c], Chhasatia et al. [2010], Kaya et al. [2010], and the volume of the droplet Ortiz et al. [2006] (further evaluation of the durability of the protocol is included in the ESI and Table. 6.6).

Synthetic tap water solutions containing six main constituents do not fully explain the environmental samples

Synthetic tap water solutions were created to reflect components measured in Lansing (lime softened groundwater), MSU (minimally treated hard water), and Detroit water (surface water with conventional treatment). A synthetic mixture of simplified Lansing water containing only the six major components (calcium, magnesium, sodium, chlorine, sulfate, and total carbonate species) displayed many features observed in Lansing waters

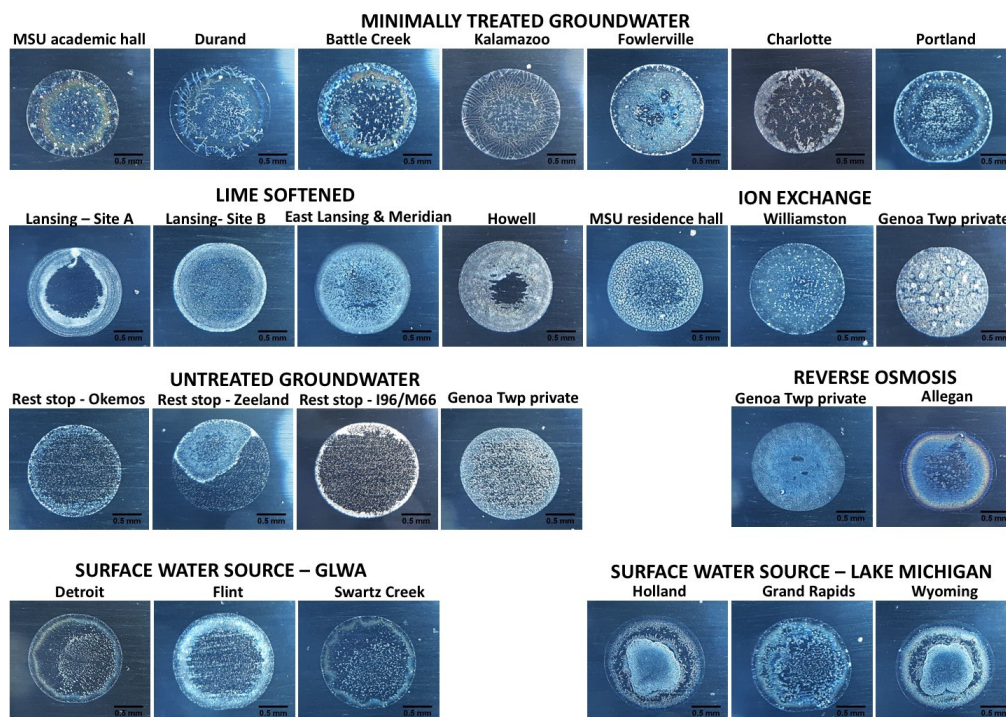
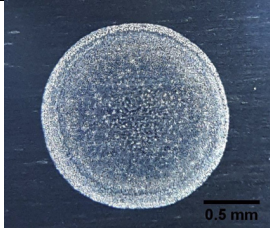
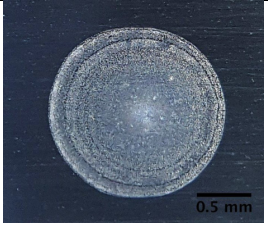
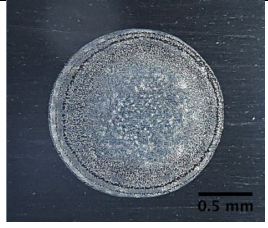
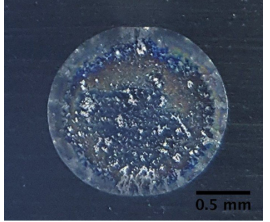
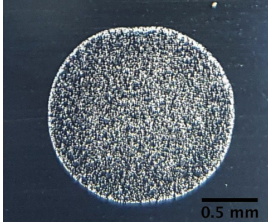
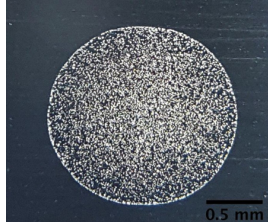
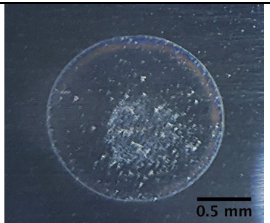
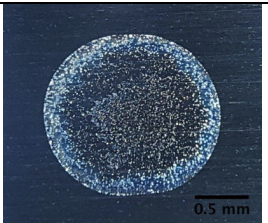
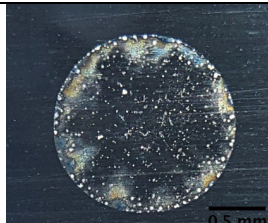


Figure 2.3: Coffee-ring residue patterns of freshly collected Michigan tap waters. The lab temperature was 24-25 °C and relative humidity 52% for this experiment. Replicates are included in Table. 6.4.

on various slides, but the simplified synthetic Detroit and MSU waters were different than the collected tap water samples (Table. 2.2). The simplified synthetic Detroit water had particles deposited at the drop edge like the environmental sample, but the rings, color, and center were different. Adding iron, copper, nitrate, fluoride, and phosphate caused the synthetic residue pattern for Detroit water to become closer to the environmental sample, but still did not capture all the features. Additional studies must be conducted to determine the influence of pH and organic matter on the residue patterns as well. The complex synthetic Detroit water sample captured the yellow and blue coloring observed in the concentric ring at the inner drop edge, possibly due to the presence of phosphate and iron forming insoluble salts. The MSU tap water still did not resemble the collected water after addition of the lower concentration components. This finding provides further evidence that lower concentration species, pH, or

particulates likely play a role in defining residue patterns.

Table 2.2: Simplified synthetic tap water compared residue patterns to real tap water, with measured pH of each solution listed below the image (24 degree C, 47% relative humidity). Replicate images are shown in Table. 6.1

	Collected tap water	Simplified synthetic	Complex synthetic
Lansing	 • 7.0-8.7	 • 8.08	 • 8.02
MSU	 • 7.34	 • 7.85	 • 8.01
Detroit	 • 6.21	 • 7.39	 • 7.35

Residue patterns document water chemistry

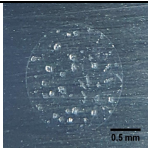
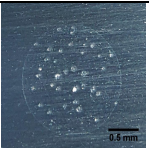
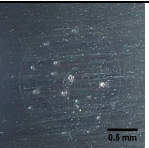
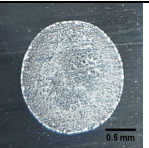
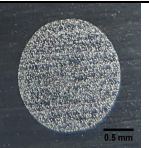
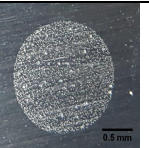
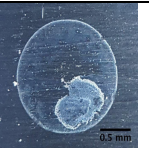
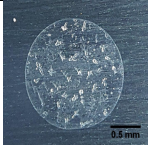
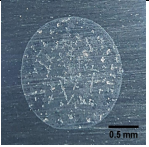
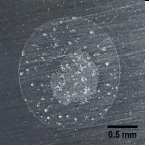
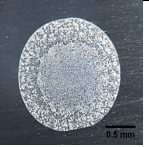
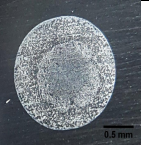
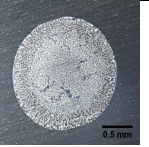
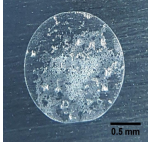
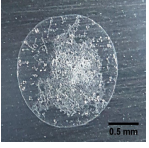
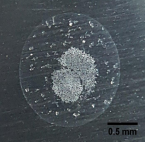
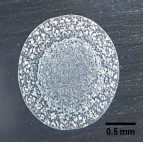
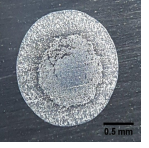
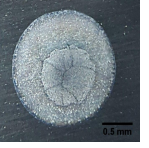
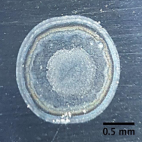
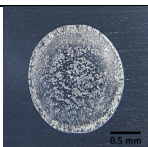
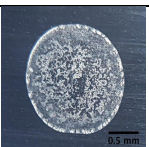
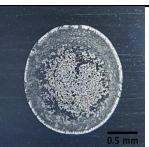
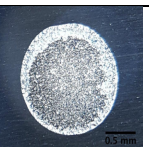
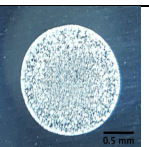
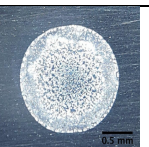
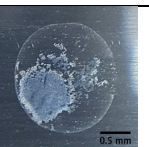
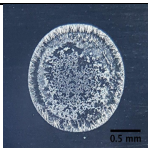
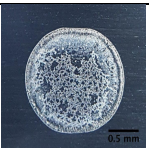
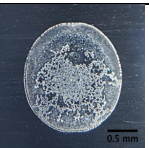
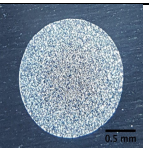
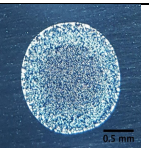
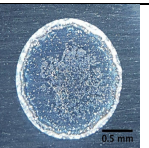
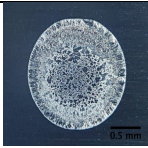
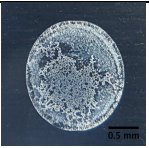
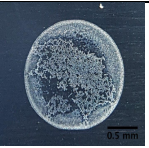
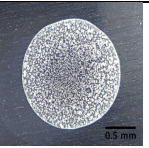
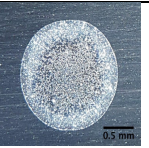
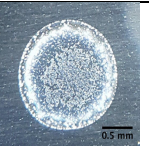
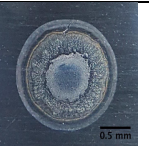
Simple synthetic mixtures demonstrate trends between water chemistry and particle, shape, size, and location of deposition. To confirm that trends in particle shapes and sizes in coffee-ring patterns are influenced by the identities and concentrations of solutes, three salt synthetic mixtures were created of $NaCl$ with $CaCl_2$ and $MgCl_2$, $NaHCO_3$ with $CaCl_2$ and $MgCl_2$, Na_2SO_4 with $CaSO_4$ and $MgSO_4$, and $NaHCO_3$ with $CaSO_4$ and

$MgSO_4$ at concentrations relevant to tap waters. In the presence of calcium and magnesium chloride, $NaCl$ caused large uniform particles to be distributed across the drop, while $NaHCO_3$ caused smaller and more densely packed flakes and feathering patterns at the higher concentrations (Table. 2.3). These features could be quantified by measuring the average area of particles and the number of particles for each set of images. For example, the average area of particles decreased with decreasing $NaCl$ concentration in the presence of 3.0 mM $CaCl_2$ and 1.5 mM $MgCl_2$, and the average number of particles decreased with decreasing $NaHCO_3$ concentration in the presence of 0.5 mM $CaCl_2$ and 0.25 mM $MgCl_2$ (Figure. 2.4). It was hypothesized that because $NaCl$ and $NaHCO_3$ are highly soluble, both produced thin films of particles that were likely deposited through surface capture or settling rather than the coffee-ring effect as ions remain dissolved through most of the droplet evaporation process. Crystal formation was sensitive to differences in slides; a similar result was found on additional slides, though the large distinct, uniformly sized $NaCl$ particles did not form at the lower concentrations of calcium and magnesium chloride (Table. 6.9). Intricate particle shapes were observed for mixtures of sodium bicarbonate with calcium and magnesium chlorides, but the shapes of the particles were not identical across all batches of slides. Additional experiments are required with higher quality substrates to determine how the shape of the bicarbonate particles correlates to the matrix water chemistry and surrounding conditions.

Simple synthetic mixtures containing sulfate salts of sodium, magnesium, and calcium had multiple concentric rings at the drop edge, likely due to differences in solubility between calcium sulfate, magnesium sulfate, and sodium sulfate. Again, the number of particles decreased with decreasing sodium sulfate concentration in the presence of 0.5 mM

CaSO₄ and 0.25 mM MgSO₄ (Figure. 2.4). Adding bicarbonate to the mixture at the same concentration of calcium and magnesium sulfate caused the concentric rings at the drop edge to be eliminated to create a thin film of densely packed very small uniform particles, except for the lowest sulfate and bicarbonate concentrations (Table. 2.3), though the number of particles still decreased with sodium bicarbonate concentration (Figure. 2.4). PCA conducted on the image files themselves (five replicates of each image) was compared to PCA on the measurements of particle sizes and numbers within the images. In both cases, three principal components were useful in clustering the images into groups with similar ions, but not sufficient to group samples by concentrations of components (Figure. 2.5). Three principal components explained around 50% of the variability of the data set for PCA conducted on the image files (Figure. 6.4). PCA is valuable for highlighting variability in a dataset, but it does not take into account subimages or sub-patterns (such as rings at the drop edge versus the center of the residue pattern) Kadappa and Negi [2016]; thus, it is not surprising that PCA on the image files was not sufficient to differentiate between images with different concentrations of ions despite clear qualitative differences in residue patterns. Specific measurements of features within the images or a convolutional neural network designed from a larger dataset may be more valuable in determining concentrations of species (Figure. 2.4).

Table 2.3: Simple synthetic mixtures analyzed at 24 °C and 48% relative humidity

	<i>NaCl</i> 10 mM	<i>NaCl</i> 5.0 mM	<i>NaCl</i> 2.5 mM	<i>NaHCO₃</i> 10 mM	<i>NaHCO₃</i> 5.0 mM	<i>NaHCO₃</i> 2.5 mM	Quality check
3 mM <i>CaCl₂</i> , 1.5 mM <i>MgCl₂</i>							
1 mM <i>CaCl₂</i> , 0.5 mM <i>MgCl₂</i>							
0.5 mM <i>CaCl₂</i> , 0.25 mM <i>MgCl₂</i>							
	<i>Na₂SO₄</i> 5.0 mM	<i>Na₂SO₄</i> 2.5 mM	<i>Na₂SO₄</i> 1.25 mM	<i>NaHCO₃</i> 10 mM	<i>NaHCO₃</i> 5.0 mM	<i>NaHCO₃</i> 2.5 mM	Quality check
3 mM <i>CaCl₂</i> , 1.5 mM <i>MgCl₂</i>							
1 mM <i>CaCl₂</i> , 0.5 mM <i>MgCl₂</i>							
0.5 mM <i>CaCl₂</i> , 0.25 mM <i>MgCl₂</i>							

Similar residue patterns were observed for collected tap water samples with similar water chemistry. Cluster analysis and trilinear classification diagrams were used to group samples with similar water chemistry, with cluster analysis taking all the collected water chemistry data into account and the trilinear diagram only using data for the species with the highest concentrations typical of fresh waters (calcium, magnesium, sodium, potassium, chlorine, sulfate, carbonate, and bicarbonate). In general, the cluster analysis and the trilinear diagrams grouped samples with those from the same treatments together (Figure. 2.6, Figure. 6.5). Cluster analysis, however, did not group ion exchange samples together, more effectively separated minimally treated groundwaters, and lumped reverse osmosis samples with surface waters. The trilinear plot showed the ion exchange samples clearly distinct from the rest, plotted the reverse osmosis samples closer to the minimally treated groundwaters, and the lime softened waters separated clearly from the surface waters. These findings highlight that the water chemistry for the ion exchanged samples are related in terms of the higher concentration components, but the overall water chemistry more closely matches samples from other groups.

Inspection of the coffee-ring residue photographs according to the groupings visualized by cluster analysis and trilinear diagrams uncovers patterns in the crystals that may associate with a given water chemistry (Figure. 2.6). For example, each ion exchange sample that clustered together on the trilinear diagram had a thin film of particles with larger crystals scattered across the drop, but each image also displayed attributes of the group assigned through cluster analysis when the lower concentration species were accounted for. Trends in the dataset can also be determined from comparing residue patterns from synthetic mixtures, samples with similar composition of the six main water components, and samples with

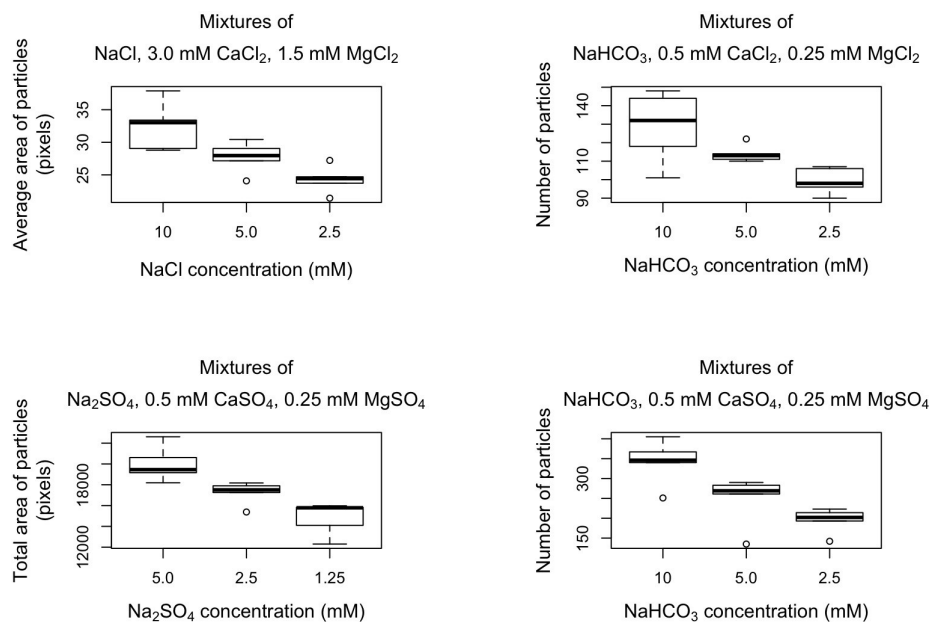


Figure 2.4: Particle areas and particle counts for simplified synthetic mixtures of three salts.

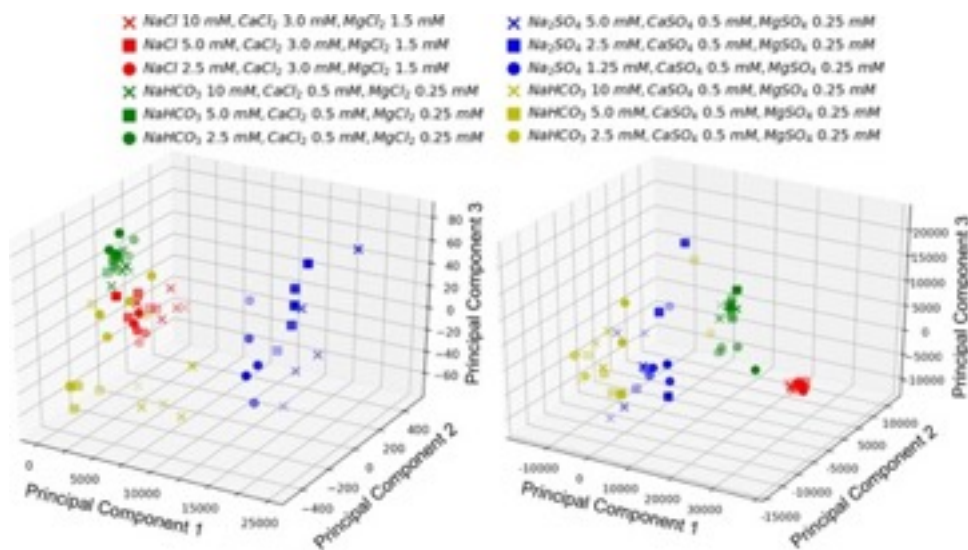


Figure 2.5: Principal component analysis (PCA) on particle measurement data (left) and PCA conducted on image files (right).

similar overall water chemistry. The residue patterns for tap waters treated by similar methods displayed characteristic features representative of that treatment, such as several concentric rings with a strong secondary ring near the outer edge for surface water, colorful concentric rings with smaller particles scattered throughout for hard groundwaters with

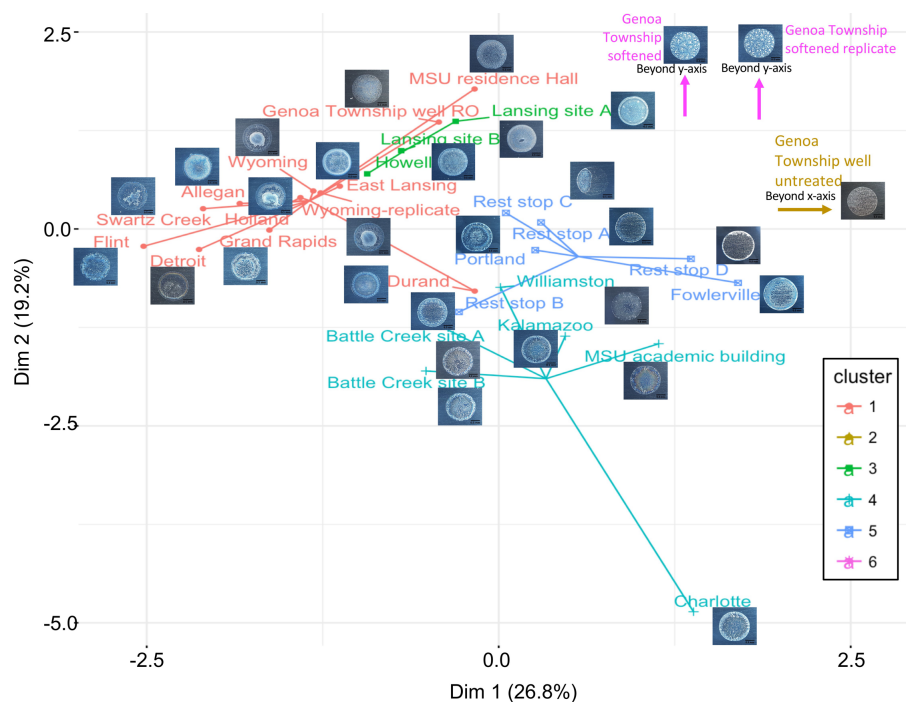


Figure 2.6: Cluster analysis of water chemistry data.

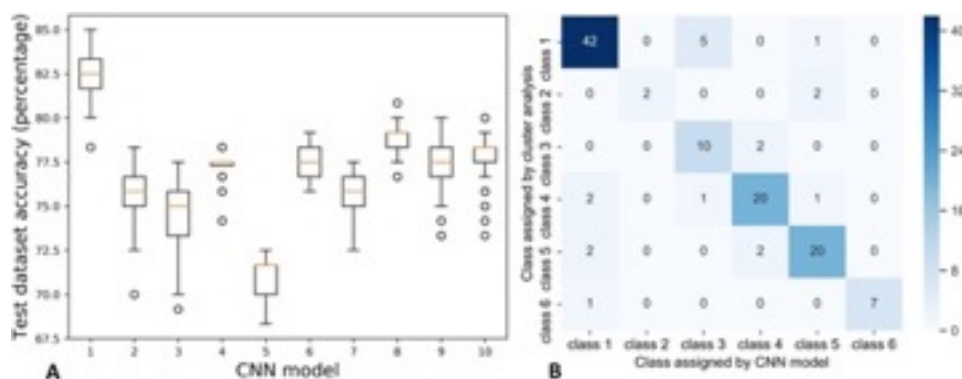


Figure 2.7: Testing dataset accuracies of ten CNN models (left) and the confusion matrix of the first trained model (right).

minimal treatment, a thin film of fine particles for reverse osmosis treated groundwater, a strong outer ring of white with small particles densely spread across the drop for untreated groundwater, large crystals scattered across the drop for ion exchange, and a white/gray thin film of small particles or dense concentric rings of small particles with feathering patterns for lime softened water (Figure. 2.3). Tap water samples contain high concentrations of dissolved ions when droplets are placed on the substrate, so particles form and grow as water evaporates from the drop as observed previously for solutions of NaCl or CaSO_4

Shahidzadeh et al. [2015]. Therefore, particles of the least soluble salts that grow quickly upon their concentrations exceeding solubility limits are expected to form particles early enough during drying to be transported by the coffee-ring effect to the drop edge, unless they grow large enough to settle first. Particles that do not form until the drop is nearly dry are expected to be deposited through the surface capture effect or settling and be found across the center of the drop. Calcium and magnesium carbonates and sulfates are less soluble than sodium and chlorine containing salts Benjamin [2014], Haynes et al. [2016]; therefore, it is logical that hard waters would display an outer ring at the drop edge and waters softened by ion exchange (containing more sodium than calcium or magnesium) would display thin films of particles. Additional mixtures must be analyzed to verify the qualitative patterns described here.

Convolutional neural network (CNN) model assigned images to groups with similar water chemistry. CNN models have previously been proven effective in object detection and image classification Krizhevsky et al. [2017], Russakovsky et al. [2015], Szegedy et al. [2015]. Herein a CNN model was developed and tested to assign residue images into classes with similar water chemistry data as determined by cluster analysis. Overall, after building the model from a library of similar training images, the CNN model was effective with 80% accuracy in assigning residue images from the test set into groups with similar water chemistry. To achieve higher accuracy, a larger dataset would be needed to train the model. Specifically, in the CNN model developed here the average and standard deviation of the accuracy for the last 100 epochs for ten independent CNN models was $76.7 \pm 3.0\%$ (Figure. 2.7). Only six of the test images were misclassified in the class one group of images that contained a total of 48 images (largely from surface waters with RO samples and a few others mixed in), but two of the test images were misclassified from class two that

contained a total of four images all from the high TDS genoa township untreated well water (Figure. 2.7). All of the misclassified images from class two were instead placed into class four that contained minimally treated groundwaters and one ion exchanged sample. Two out of twenty-four images from class four and two out of twenty-four images in class five (minimally treated and untreated groundwaters) were misclassified into class one. A few additional images were also misclassified between class four and five; in qualitative comparing residue images, images of class four and class five are more similar than images in other classes, which is logical considering these both classes largely contain minimally treated and untreated groundwaters. Confusion matrix of the ten models were provided in Figure. 2.8.

There were a few of the test images that were misclassified more often than others (Table. 6.10). Five of the test images with a misclassification percentage over 70% had a coffee-ring residue pattern that was notably different from replicates of the same sample. For example, two MSU residence hall samples had a clearing in the center of the residue pattern while the rest had a complete thin film across the entire drop; the two samples with clearings were misclassified in over 70% of the models (Table. 6.10, Table. 6.3). Two of the images with a misclassification percentage over 70% were from class two which had the lowest number of replicates. The low number of images causes the model to be less sensitive to this class despite the distinct large crystal pattern Junqué de Fortuny et al. [2013], Martens et al. [2016]. Three images were often misclassified without a clear reason (Table. 6.10).

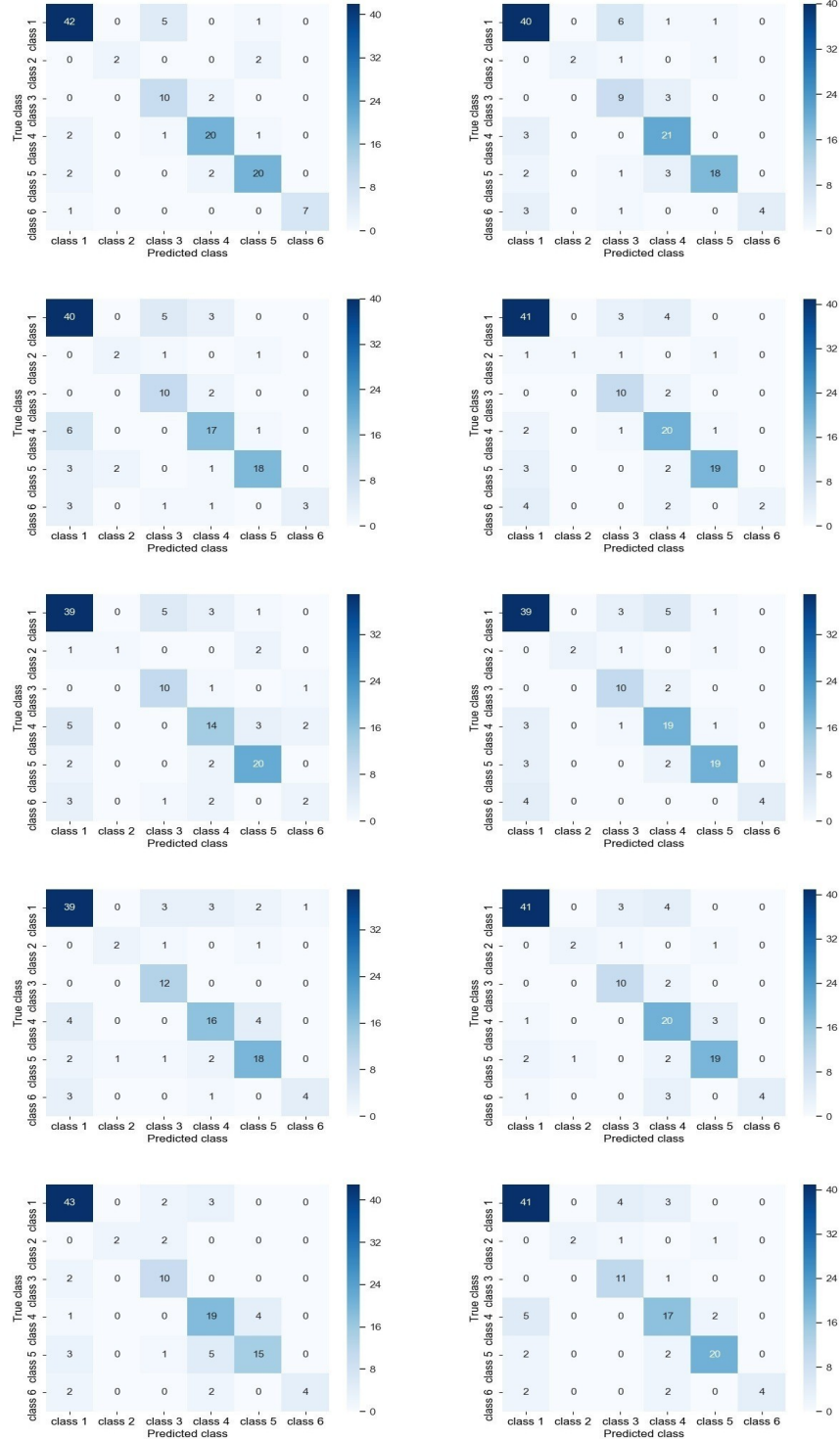


Figure 2.8: Confusion matrix of ten CNN models.

The percentage of images that were properly classified into class one was much higher than most of the other classes. Class one had the most images, so in the model training

process the model is skewed to more accurately predict the class one images Japkowicz and Stephen [2002], Krawczyk [2016]. Generally with CNN models the accuracy is improved by using a larger dataset of images during model training to allow the model to capture more information and detail Junqué de Fortuny et al. [2013], Martens et al. [2016]. Overall, class one, three, four and five had similar accuracy around 80%, but due to the low number of samples the accuracies of classes two and six were around 40-50% (Figure. 6.6). About half images in class one had less than 1% mis-classification percentage and most images in class two and six had high mis-classification percentages.

2.5 Conclusions and future outlook

Both the coffee-ring effect and convolutional neural networks (CNNs) remain underutilized techniques to be harnessed for tap water analysis. Herein we show proof of concept experiments that document the unique fingerprints provided by the coffee-ring effect for tap water solutions from various cities across Michigan and the reproducibility of the phenomenon, demonstrate that low concentration species as well as major ions influence the residue patterns, provide evidence that the patterns indeed document water chemistry within the sample, and demonstrate the ability of a CNN in assigning images to water chemistry. The low-cost substrate employed in this work caused variability between experiments, especially for batches of substrates purchased at different times; however, the variability was included in the training dataset, so the CNN was still able to classify the images with 80% accuracy. Additional work is required to identify the appropriate substrate that is widely available for a low cost test. Quality control metrics are critical for identifying variation in experiments, and lime softened water was much more sensitive to experimental variation than the hard synthetic water used as a control for this study. Traditional PCA on image files is insufficient for differentiating between images of water

samples with different concentrations of components, likely due to lack of consideration of subregions such as the outer coffee-ring; however, with a larger dataset a CNN model will be especially valuable for differentiating between water chemistries and assigning unknown images to groups from a library of images. A larger library of residue patterns and a corresponding CNN model must be trained to move this technology from qualitative tap water quality analysis to a quantitative technique and to further identify features of the residue patterns.

Despite the use of a low-cost and variable aluminum slide, using a pipette, \$18 jeweler's loupe, and cell phone camera, each type of tap water tested displayed unique characteristics, water samples with similar water chemistry produced residue patterns with similar features, waters from two locations in a city were more similar than samples from different cities, and the CNN model was able to assign samples to groups with similar water chemistry. This evidence suggests that this method should be further considered for low-cost water quality fingerprinting.

CHAPTER 3

Optimal environmental condition for contaminants separation by coffee-ring effect

3.1 Abstract

This study investigates the potential of the coffee-ring effect as a tool for tap water analysis, demonstrating its ability to produce unique fingerprints for water samples with varying compositions and environmental conditions. However, the coffee-ring effect's stability is found to be influenced by environmental conditions, presenting a challenge for its practical application. Additionally, identifying the optimal environmental conditions for separating contaminants particles is essential to enhance the technique's efficacy. Establishing the correlation between water sample coffee-ring effect patterns and element deposition compositions is also crucial for utilizing the technique to identify particle compositions. The study confirms the reproducibility of the coffee-ring effect and highlights the impact of both environmental condition and water compositions on the residue patterns produced.

Various statistical methods, such as ANOVA, MANOVA, and PERMANOVA, can differentiate coffee-ring effect residue patterns with respect to environmental conditions and water sample compositions. However, determining the most effective method for differentiating these patterns requires further research, as the results from different analyses can be inconsistent.

The study's statistical analyses indicate that environmental conditions and water chemistry significantly influence residue patterns and element distributions. Optimal environmental conditions, including 23-26°C with 45-50% relative humidity, 20-23°C with

45-50% relative humidity, and 26-29°C with 40-45% relative humidity, are identified for differentiating water samples with varying component concentrations. Nonetheless, the optimal environmental condition is a temperature range of 23-26°C and a relative humidity of 45-50%, as it yielded the highest number of optimal results in 12 separate analyses.

These findings have implications for further research on residue patterns and improving the understanding of the underlying mechanisms of the coffee-ring effect.

3.2 Introduction

Centralized drinking water supply and distribution systems in the U.S. were developed in 1854 to reduce the reliance of fast growing cities on contaminated wells and decrease incidence of cholera and typhoid diseases Burian et al. [2000]. Today, water distribution systems are currently reaching their end of life and failing faster than they can be replaced, requiring funding at a rate that strains many communities Coghill et al. [2014], Folkman [2018]. According to a 2018 report of 197,866 miles of pipes across the United States, over 16% of installed water mains are beyond their useful life, 28% of pipes of all material types are older than 50 years, and 71% of all the pipes are older than 20 years Folkman [2018]. Since 2012, the overall break rates increased 27% , primarily due to failures in asbestos cement (AC) and cast iron (CI) pipes Folkman [2018]. The most common method for prioritizing pipe replacement is based on failure data. Large, critical mains have essentially been ignored in many communities until they failed Darlene Garcia and Susan Funchion [2015]. This method ignores water quality issues related to aging pipes. Prior knowledge of pipe material or age can also be used to prioritize pipe replacement, but knowledge of where lead service lines or older pipes exist is not always available Cornwell et al. [2016]. Researchers have also developed models to prioritize pipe replacement based on pipe failure data including multiobjective genetic algorithms, failure

assessment models, rank aggregation models, etc. Giustolisi and Berardi [2009], Rogers and Grigg [2008], Tlili and Nafi [2012], Choi et al. [2017], Marzouk et al. [2015], Ho et al. [2009]. Water quality data can also be used to directly identify sections of the distribution system that negatively impact water chemistry Kirmeyer [2002]; however, collecting sufficient water quality data across a distribution system to determine which pipes are hazardous to public health is often challenging due to the time and costs required to collect and analyze enough water samples. Herein we propose to develop a fast, low-cost method that can drastically increase the number of water samples that can be collected and analyzed to aid in identification of waters across a distribution system that have been impacted by corrosion. This method will harness tap water fingerprints created by the coffee-ring effect.

Tap water fingerprints provided thru the coffee-ring effect are unique to water chemistry

When the coffee-ring effect is harnessed, tap waters leave unique residue patterns, or fingerprints that correlate to tap water chemistry (Table. 3.1), Li et al. [2020], Shahidzadeh-Bonn et al. [2008], Kaya et al. [2010], Shin et al. [2014], Shahidzadeh et al. [2015]. The residue pattern formation is a crystallization process of water contaminants and crystallization of salts or other materials in supersaturated solutions has been intensively investigated due to its practical significance in pharmaceutical purification, salt manufacturing, seawater purification, cosmetic production, deicing, and so on Li et al. [2020], Qazi et al. [2017], Wei et al. [2012], Sammalkorpi et al. [2009], Desarnaud et al. [2014], Meldrum and O'Shaughnessy [2020]. Previous researchers mainly studied the mechanisms of crystallization in electrolyte solutions without evaporation. However, Studies stressed on precipitation and crystallization from evaporating sessile droplets are far less especially when compared with the active domain of colloidal sessile droplets

Zhong et al. [2015], Feng et al. [2017], Zhong and Duan [2016], Anyfantakis et al. [2015], Zhang et al. [2016], Xu et al. [2016], Bahmani et al. [2017], Lee et al. [2017], Saxena et al. [2017], Li et al. [2016d], Chen et al. [2012], Malvadkar et al. [2010]. According to the previous study, the more complex profile of a sessile droplet characterized by the three phase contact line and the curved liquid vapor interface complexes the precipitation process as compared the easy solution configuration. The higher evaporation flux in the vicinity of the contact line can induce outward flows that result in heterogeneous distribution of ions and the associated supersaturation degree. At the mean time, microfluid formed inside the droplet sessile bring particles to the droplet substrate contact line. The curved liquid vapor interface could limit the growth and vary the motion of precipitation. The complexity caused by the multifactors in respect to evaporation, bulk flow, temperature, humidity and wettability is therefore expected to significantly vary crystallization in sessile droplets.

So far crystallization of salts from drying saline droplets has been investigated in a number of studies mainly focused on nucleation mechanisms and the dependence of precipitation profile on solid surface properties, salt concentration, and so forth Takhistov and Chang [2002], Townsend et al. [2017], Kaya et al. [2010], Shahidzadeh et al. [2015], Shin et al. [2014], Suresh [2006], Shahidzadeh-Bonn et al. [2008]. The previous study of the effects of polyelectrolyte concentration of drops and the surrounding humidity on the final salt crystallization, which exhibited profiles of concentric rings and needle-like and chainlike structures Kaya et al. [2010]. Takhistov et al. investigated the crystal formation process from microliter droplets on both hydrophilic and hydrophobic substrates. Based on their results, concentric rings of salts were formed on hydrophilic surfaces while crystalline was produced on hydrophobic surfaces Takhistov and Chang [2002]. Shahidzadeh et al. also investigated the evaporation and stain structures on

various substrates with two types of salts, sodium chloride ($NaCl$), and calcium sulfate ($CaSO_4$) with different crystalline structures and precipitation pathways. In their research, they concluded the crystalline pattern in a variety was concluded to be controlled by the interfacial properties of the emerging crystalline and the number of crystals generated Shahidzadeh et al. [2015]. The study of crystallization from saline droplet is conducted by Shin et al. They obtained three-dimensional salt structures from droplets with high aspect ratio and a rich variety of three-dimensional crystalline deposits were observed Shin et al. [2014]. The coffee-ring effect process involves the solvent evaporation on droplet surface and resulting residue ring like patterns. The formation of the coffee-ring effect pattern is complex. The contact line pinning on the substrate and the contact angle determines the pattern formation Wong et al. [2011], Larson [2014], Deegan et al. [1997], Chen and Evans [2010], Eral et al. [2013]. Wong et al. found the physics of particle separation during coffee-ring formation, which is based on a particle-size selection mechanism near the contact line of an evaporating droplet. On the basis of this mechanism, they found nanochromatography of three relevant biological entities (proteins, micro-organisms, and mammalian cells) in a liquid droplet, with a separation resolution on the order of 100 nm and a dynamic range from 10 nm to a few tens of micrometers Wong et al. [2011].

Coffee-ring effect applications

Understanding and controlling the process of solute deposition in the presence of coffee-ring effect is important in manufacturing processes involving evaporation on surfaces including printing Park and Moon [2006], Friederich et al. [2013], Kuang et al. [2014], Sun et al. [2015], Huang and Zhu [2019] and fabrication of ordered structures Han and Lin [2012], functional nanomaterials Shao et al. [2014], Zou and Kim [2014] and colloidal

crystals Park et al. [2006], Cui et al. [2009]. coffee-ring effect also improves the performance of commercial applications including fluorescent microarrays Blossey and Bosio [2002], Dugas et al. [2005], matrix assisted laser desorption ionization (MALDI) spectrometry Hu et al. [2013], Mampallil et al. [2012], Kudina et al. [2016], Lai et al. [2016], and surface enhanced Raman spectroscopy (SERS) Zhou et al. [2014a], Wang et al. [2014], Garcia-Cordero and Fan [2017]. coffee-ring effect has also implications in plasmonics Li et al. [2016a], solute separation Wong et al. [2011], diagnostics Brutin et al. [2011], Wen et al. [2013], Gulka et al. [2014] and electronics applications de Gans and Schubert [2004].

Suppression of coffee-ring effect

Coffee-ring effect can be suppressed through one of the three physical strategies (i) preventing the pinning of the contact line; (ii) disturbing the capillary flow towards the contact line and (iii) preventing the particles being transported to the droplet edge by the capillary flows. The coffee-ring effect could be suppressed by preventing contact line pinning using hydrophobic surfaces. Increasing the hydrophobicity of surfaces is often accompanied by decreasing contact angle hysteresis (CAH) Eral et al. [2013]. Lower CAH in essence means reduced contact line pinning which leads to suppression of coffee-ring effect. Lower CAH could be achieved by patterning of controllable surface wettability as reviewed previously by Tial et al. Tian et al. [2013]. These methods include chemical modification Ko et al. [2004], Tian et al. [2013] and physical modification.

On hydrophobic and partially hydrophobic surfaces, pinning can even occur when the CAH or solute concentration is high. If CAH is high, during the contact angle decreases to the receding angle, typically a few seconds depending upon the rate of evaporation, solutes can accumulate at the contact line. Such accumulation produces ring-like deposits only if the duration of pinning is above a critical value for a given substrate-

solute system Moraila-Martinez et al. [2013]. However if the pinning time is short, even with high initial solute concentration, the coffee-ring effect will just produce smaller inner rings Nguyen et al. [2013]. The nanoparticles are more prominent to form ring like patterns compared with larger particles as they can flow into the microscopic regions of the droplet edge faster. In the presence of solute particles in the droplet, electrowetting (EW) can reduce the pinned contact line on (partially)-hydrophobic surfaces Mugele and Baret [2005], Li and Mugele [2008]. A droplet is deposited on a dielectric layer covering an electrode. When a voltage is applied between the droplet and the electrode an electric force pulls the contact line outward, overcoming the pinning forces so the contact line pinning is reduced. The coffee-ring effect can also be suppressed by vibration and acoustics, marangoni flow and other factors Mampallil and Eral [2018].

Enhancement of coffee-ring effect

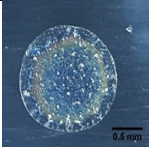
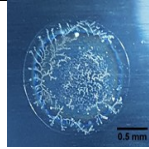
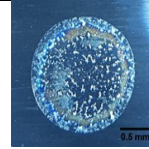
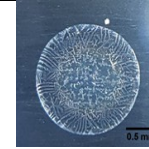
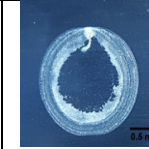
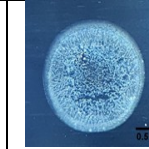
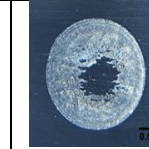
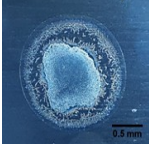
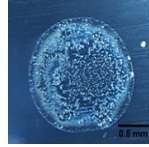
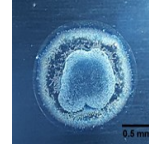
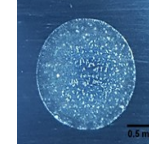
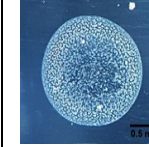
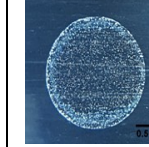
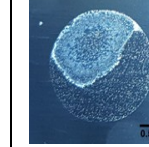
Evaporation of droplets can be utilized as a method to concentrate its solutes in it. Evaporation of the solvent can increase the analyte concentration making the reactions more probable Hernandez-Perez et al. [2016], De Angelis et al. [2011]. By the coffee-ring effect, the solutes is deposited at the contact line increasing their concentration there and separated by their size, charge and solute-substrate interactions. This deposition of solutes and particles are exploited as a pre-concentration method Figure. 1.1.

Concentrating solutes at the rim of the droplet by coffee-ring effect is called the self-ordered ring (SOR) method. It acts as a pre-concentration procedure before other analyses. To enhance the coffee-ring effect, hydrophobic surface is usually used as the substrate. Drying process on hydrophobic surfaces forms smaller rings with higher solute density as the contact line is pinned only in the later stages of the evaporation. Liu et al. demonstrated that the SOR method enhanced the fluorescence detection of orally

administrated berberine in human urine Liu et al. [2002]. Similarly, fluorescent detection of trace levels of tetracycline Huang et al. [2004a], quinidine sulfate in serum samples Yang and Huang [2006] and fluorescein Liu et al. [2006] was demonstrated based on the SOR method.

Coffee-ring effect could facilitate identifying pathogens which are associated with diseases by isolating the disease markers from body fluids Wong et al. [2011], Chen and Evans [2010]. Coffee-ring effect has also been used to enhance the deposition of gold nanoparticles(AuNPs) on cellulose nanofibers (CNFs) to enhance surface-enhanced Raman scattering (SERS) Chen et al. [2017], Wang et al. [2014], Hussain et al. [2019], Juneja and Bhattacharya [2019], Zhou et al. [2014b]. Coffee-ring effect has also been utilized for a low-resource malaria diagnostic platform Gulka et al. [2014]. Coffee-ring effect also has shown great potential to monitor tap water quality with deep neural networks Li et al. [2020].

Table 3.1: Coffee-ring residue patterns of Michigan tap waters Li et al. [2020].

MINIMALLY TREATED GROUNDWATER				LIME SOFTENED		
MSU tap water	Durand tap	Battle Creek	Kalamazoo tap	Lansing	East Lansing	Howell
						
SURFACE WATER LAKE MICHIGAN			ION EXCHANGE		UNTREATED GROUNDWATER	
Holland, MI	Grand Rapids	Wyoming	Williamston	Holmes Hall	Okemos	Zeeland
						

Tap water fingerprinting is fast, low-cost, and has potential to be automated, allowing greater numbers of samples to be analyzed across a distribution system

Compared with other methods, the coffee-ring effect method for measuring pipe corrosion indicators has benefits of being low-cost and fast, not requiring specialized technicians, and the same method can be used to see multiple analytes at once. Required equipment to complete the coffee-ring effect method includes a small aluminum substrate and one pipette which costs about 10 dollars. To collect images, a cell phone camera and a \$18, 30x jeweler's loupe can be used. Considering the wide availability of cell phone cameras already used in households, the total cost for new, reusable equipment for this method is less than forty dollars Li et al. [2020]. Common methods for contaminants elements measurement are ICP-MS (about \$25, 000 – \$40, 000 for refurbished), atomic absorption (about \$13, 000 – \$20,000), and spectroscopic methods such as phenanthroline method, neocuproine method and bathocuproine method Walter [1961] The coffee-ring effect method is not only a low-cost method, but also fast (approximately total 25 minutes including 5 minutes to drop water and 20 minutes to dry), does not use hazardous reagents, and does not require specialized technicians to conduct the experiment, and has potential to be automated for the evaluation of high numbers of samples across a distribution system.

Optimization of tap water fingerprinting for tap water contaminants

As demonstrated in previous research, tap water fingerprinting (coffee-ring effect), an innovative technique for identifying and characterizing water samples, effectively distinguishes between different tap water compositions and differentiates mixtures of salts based on their consistent and reproducible water fingerprints Li et al. [2020], Shahidzadeh-Bonn et al. [2008]. This groundbreaking approach shows promising potential for a range of applications in environmental monitoring and water quality management.

The tap water fingerprinting method produces consistent and reproducible residue patterns under constant environmental conditions 3.2, but data is not yet available to demonstrate how much the residue patterns of dried water droplets change for small changes in environmental conditions.

Table 3.2: Nine environmental conditions

Temperature, RH	20-23 (°C)	23-26 (°C)	26-29 (°C)
35%-40%	A	D	G
40%-45%	B	E	H
45%-50%	C	F	I

Under low evaporation rate conditions, particles have time to arrange by Brownian motion Mampallil and Eral [2018], Rodriguez-Navarro and Doehne [1999], Marin et al. [2011]. In contrast, when the evaporation rate is high, high-speed particles deposit into a disordered phase. Consequently, under high relative humidity and low-temperature conditions, coffee-ring fingerprints are more constant Mampallil and Eral [2018], Rodriguez-Navarro and Doehne [1999], Marin et al. [2011]. However, no research has quantified how evaporation rate (temperature and relative humidity) influences residue patterns for mixed salt solutions at concentrations relevant to tap water.

In this study, we further optimized the tap water fingerprinting methodology to enhance its capabilities for identifying contaminant particles in water samples. This optimization process involved several critical factors that significantly influence the accuracy and reliability of the fingerprinting results. Key factors considered include optimal temperature and humidity conditions, and solute properties. Experiments will be conducted to determine how much temperature and humidity control is required to minimize changes in particle positions, sizes, shapes, elemental composition, and crystal structures while also maximizing the separation of contaminant particles within the coffee-ring pattern. In this work, the question of what temperature and relative humidity ranges (within the range

of 20-29 degrees C and 35-50% relative humidity) provide reproducible fingerprints and sufficient separation of contaminant particles from other salts to facilitate detection within a photograph will be answered.

Firstly, we examined the effects of temperature and humidity on the fingerprinting process. By conducting a series of controlled experiments, we determined the optimal temperature and humidity conditions that yield the most accurate and consistent water fingerprints. These findings are crucial in ensuring that the fingerprinting method can be effectively applied under varying environmental conditions and across diverse geographical regions.

Next, we investigated the role of solute properties in the fingerprinting process. Given that the presence of various solutes can alter the characteristics of water fingerprints, understanding their effects is essential for accurately identifying contaminants in water samples. Through rigorous testing, we determined the key solute properties that influence the fingerprinting results. Furthermore, we identified the optimal conditions to concentrate similar contaminants and effectively separate different contaminants, thereby enhancing the precision and reliability of the tap water fingerprinting method.

In conclusion, our optimization of the tap water fingerprinting method has resulted in significant improvements in its ability to identify contaminant particles in water samples. By carefully considering and addressing the effects of temperature and humidity conditions and solute properties, we have established a more reliable and accurate technique for analyzing water quality and detecting potential contaminants. This optimized fingerprinting method holds great promise for enhancing water safety and protecting public health on a global scale.

3.3 Experimental Methods

3.3.1 Materials and instruments

The following substances were purchased from Fisher Scientific: sodium bicarbonate, calcium chloride, magnesium chloride, sodium sulfate, sodium phosphate monobasic, potassium fluoride, sodium hydroxide, iron nitrate nonahydrate, and copper sulfate. The surface-polished aluminum slides used were obtained from McMaster-CARR (1655T1) with a yield strength of 35,000 psi, a hardness of Brinell 95 (soft), and a fabrication of cold rolled, temper 3/8" thick T651. The slides met the specification of ASTM B209 and were polished to a #8 reflective finish without any visible grain lines. One side of these sheets and bars was polished to either a brushed finish or a mirror-like finish and protected with a peel-off film. 6061 aluminum, the most commonly used type, is used to make a wide range of products, from pipe fittings and containers to automotive and aerospace parts.

The Scanning Electron Microscopy (SEM) and Energy-Dispersive X-ray Spectroscopy (EDS) images were acquired using a high-performance JEOL 6610LV SEM system, set at an accelerating voltage of 20 kV. This advanced microscope is specifically designed for the efficient characterization and imaging of delicate structures, providing exceptional SEM imaging at magnifications ranging from 5X to 50,000X. The accelerating voltage of the JEOL 6610LV can be adjusted from 300 V to 30 kV.

X-Ray diffraction images were collected by the Oxford EDS system which was equipped on the SEM system. The JEM 6610LV Scanning Electron Microscope (SEM) is equipped with EDS. SEM/EDS provides chemical analysis of the field of view or spot analyses of minute particles. The EDS Analysis System for SEM was designed for a wide range of applications. Whether simply collecting a spectrum or performing complex phase analysis, the system is easy to get the quick results you want. EDS analysis is best

suited for: Metals and metal alloys, Ceramics, Minerals and Certain types of polymeric materials. The operation software is Scandium image processing software by Olympus Soft imaging Solutions. Coffee-ring effect patterns were also collected by SamSung S6 cell phone or a 5 MP Digital Microscope Pro-20x-200x magnification (Celestron) camera. Data analysis and statistical analysis were performed on MATLAB R2021a, R 4.1.1 and python 3.7.

3.3.2 Four-axis-autosampler

The Four-axis-autosampler is a complex device that is designed to automate the process of collecting and injecting samples. The device is composed of several components. The 3D printer stage, a CNC 3018-PRO Router Kit, is responsible for providing the the foundation for the other components to be mounted on and for providing the necessary movement and precision for the device to operate accurately. The injector, a Thermo Scientific 365CL221, is responsible for injecting the samples into the system. This component is designed to be highly precise and accurate, ensuring that the samples are injected with minimal error or variation. The Raspberry Pi-4 Model B 2019 Quad Core 64 Bit WiFi Bluetooth (4GB) serves as the controller for the stepper motors, the injector, and the sample collection system. The Raspberry Pi is also responsible for running the python code that controls the device's operations. The 3 steppers, Nema 17 Bipolar 2A Stepper Motor by OSM Technology Co (17HS19-2004S1), are responsible for moving the injector. These motors are designed to provide precise and accurate movement of the injector, ensuring that samples are injected in the correct location. The one stepper motor driver (TB6600 4A 9-42V Nema 17) is responsible for operating the sample collection and injection action. This stepper motor is responsible for moving the sample collection system, which is responsible for collecting samples, and moving the injector, which is responsible for

injecting the samples into the system. The device is operated by python code under linux system, specifically Ubuntu operating system. The sample code is used to control the various components of the device, including the stepper motors, the injector, and the sample collection system. This code is responsible for ensuring that the device operates accurately and efficiently and is able to collect and inject samples with minimal error or variation.

The Four-axis-autosampler is a highly advanced device that is designed to automatically prepare water samples based on a predefined set of water samples. The device is equipped with a sample holder that can hold up to 32 water samples at a time, making it suitable for large-scale sample preparation tasks.

The device operates in several steps, each of which is specifically designed to ensure accurate and efficient sample preparation. In the first step, the autosampler resets its syringe positions to the initial setting to ensure that the syringe is in the correct position and orientation before it begins to collect and inject samples. The syringe is then washed with nanopure water to ensure that it is clean and free from any contaminants.

In the second step, the syringe collects a 2 μL water sample at a predefined water sample location to ensure that the correct sample is collected and that the sample is collected in the correct location. The stage then moves the syringe to the desired sample location above the substrate and lowers the syringe until the syringe tip is 0.5 mm above the substrate. This step is important for ensuring that the sample is delivered to the correct location on the substrate.

In the third step, the fourth motor pushes the syringe piston to slowly push the 2 μL water sample out of the syringe. This step is important for ensuring that the sample is delivered to the substrate in a controlled and precise manner. The water sample is then

dropped on the substrate surface.

In the last step, after the water sample is dropped, the syringe is rinsed with nanopure water again and reset to its original location for collecting the next water sample. This step is important for ensuring that the syringe is clean and free from any contaminants before it collects the next sample. The whole process is then repeated for each water sample in the sample holder. This allows for efficient and accurate sample preparation for a large number of samples in a short period of time. The process flow is illustrated in Figure. 6.7.

Furthermore, the system is built on open source software and hardware, it can be easily modified and expanded according to the user's needs. The device's control system is based on a Raspberry Pi, which is a powerful and versatile platform that can be easily programmed and customized. This allows for flexibility and adaptability in the device's operation, making it suitable for a wide range of applications. The Four-axis-autosampler is a powerful and efficient device that is designed to collect water coffee-ring samples at a high speed. The device is capable of collecting samples at a rate of 45 seconds per sample, which is comparable to the speed of a human sample collector, who typically takes around 30 seconds per sample. However, the autosampler has several advantages over human sample collectors. One of the main advantages of the auto-sampler is its stability and ability to work continuously for longer periods of time. Unlike human sample collectors, the device does not tire, and it can work continuously without interruption. This is an important feature for large-scale sample preparation tasks that require a high degree of accuracy and consistency. Another advantage of the auto-sampler is that it can be placed in a small chamber with controlled temperature and humidity. This is beneficial because it allows for precise control over the sample preparation environment, which is important for maintaining the integrity and quality of the samples. Operating the same

experiments manually under this condition is tedious and time-consuming. In addition to its ability to collect water coffee-ring samples, the auto-sampler can be easily modified to work for other tasks. For example, it can be used for solution preparation, blood test and so on. This makes it a versatile and useful tool for a wide range of applications.

Overall, the Four-axis-autosampler is a powerful and efficient device that can significantly improve the speed and accuracy of sample preparation tasks. Its compact size, precise control over the sample preparation environment, and ability to work continuously make it an ideal tool for large-scale sample preparation tasks.

3.3.3 Auto temperature humidity control chamber

An auto-temperature-humidity control chamber was constructed using a chamber, two Diymore XH-M452 temperature and humidity controllers, a Space SFH-181 TP heater from Ningbo Electrical Appliance Company, a Frigidaire FFRA051WAE 5000 BTU air conditioner, and an AO-101 AquaOasis humidifier. Sodium hydroxide was used as a dehumidifier. Typically an environmental control chamber would cost on the order of \$5000; herein, to reduce overall cost of implementing the tap water fingerprinting method we built, will demonstrate use of, and will publish designs for a lower cost setup on the order of \$1000. The chamber controlling system consists of two automotive temperature and relative humidity controllers and one of them is programmed to increase temperature and relative humidity and the other is programmed to decrease temperature and relative humidity. The chamber consists of a 12V, 200W heater, a ultrasonic humidifier, a 500 ml plastic bottle with dry NaOH and desiccant and a 5,000 BTU 115V mini air-conditioner.

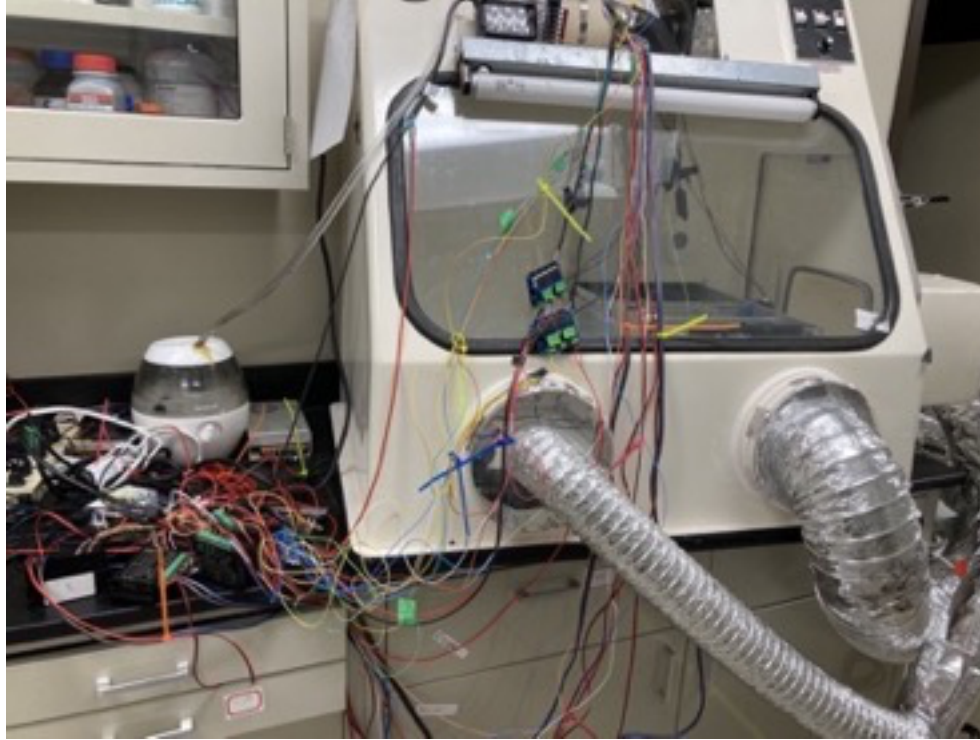


Figure 3.1: Temperature humidity control chamber.

This auto temperature humidity control chamber is capable of adjusting and maintaining the temperature and humidity automatically based on the pre-set temperature and humidity values in the two Diymore controllers. The sensitivity of temperature is 0.5 degree and relative humidity is 1%. Based on the test, the system is capable adjust temperature at a speed of 3 degrees Celsius per min and relative humidity of 2% per min. After adjusting the temperature and humidity to desired the desired range, the chamber switched to main mode. If the temperature increased and above the highest temperature limit, the air conditioning switch would be turned on to decrease the temperature to the desired range. On the other hand, if the temperature of the chamber was below the lowest limit, the switch of the heater would be turned on to increase the chamber temperature until temperature increased to the desired range. The humidifier and dehumidifier worked in the same way. When the chamber humidity was below the pre-set lowest limit, the

humidifier would be turned on until the humidity reaches the desired range and if the humidity was higher than pre-set highest limit, the dehumidifier would be turned on to lower the humidity to the desired range.

3.3.4 Water samples

In order to determine the effects of temperature and humidity on residue patterns of various water compositions, synthetic tap water samples containing various concentrations of the main components in tap water were prepared based on the range of composition concentrations of the Detroit water quality report in 2017, 2018, 2019. Detroit water is served by Great Lakes Water Authority to about 3.5 million people, 40% Michigan residents (Detroit Water and Sewerage Department 2015). Sources of Detroit tap water include the Detroit River and Lake Huron and thus, the composition of Detroit tap water varies over time. The water recipe is determined by the average of Detroit Water Quality Report from 2016 to 2018 and three recipes are designed to mimic the variability of the water chemistry Table. 3.3. Water recipes Table. Water recipes will be spiked into water samples prepared by preparing water sample with 0.7 ppm fluoride, 0.4 ppm nitrate, 0.062 ppm aluminum, 1.1 ppm potassium, 25 ppm sulfate, 0.36 ppm phosphorus in nanopure water (Table. 3.4).

Table 3.3: Detroit tap water components data sheet (Source: Detroit water quality reports 2017-2019)

Components	Average (ppm)	Average (mM)
Nitrate	0.790	0.013
Lead	0.000	0.000
Iron	0.277	0.005
Copper	0.015	0.000
Magnesium	10.800	0.444
Calcium	37.833	0.946
Sodium	9.817	0.427
Potassium	1.533	0.039

Table 3.3: (cont'd)

Sodium	9.817	0.427
Potassium	1.533	0.039
Manganese	0.004	0.000
Zinc	0.000	0.000
Sulfate	33.267	0.346
Phosphorus	1.040	0.034
Chloride	18.033	0.509
Fluoride	0.853	0.045

Table 3.4: Recipe for synthetic water samples

Sample ID / Components (mM)	$NaHCO_3$	$CaCl_2$	$MgCl_2$	Na_2SO_4	NaH_2PO_4	KF	$Fe(NO_3)_3$	$CuSO_4$
Sample A	0.1	1.5	0.5	0.35	0.033	0.4	0.005	0.00024
Sample B	0.2	1	0.35	0.35	0.033	0.4	0.005	0.00024
Sample C	0.1	0.5	0.2	0.35	0.033	0.4	0.005	0.00024
Sample D	0	1	1	1.35	0.033	0.4	0.005	0.00024
Sample E	0	1	0.5	2.35	0.033	0.4	0.005	0.00024

3.3.5 Coffee-ring effect pattern statistical analysis methods

After preprocessing the images, particles would be recognized by MATLAB and would be used to calculate particle shape, color, location from the drop edge, and size. These properties would be extracted from each residue image for each water recipe, and analysis of variance (ANOVA) would be conducted across the nine environmental condition groups and for constant evaporation rates (five replicate samples in each group). Residue patterns for two environmental conditions would be considered different from one another when a statistical difference is observed for any of the particle measurements (shape, color, location from the drop edge, and size). Residue patterns would be labeled as consistent across two environmental conditions when there is no statistical difference observed between any of the particle measurements. Analysis of variance (ANOVA) is a statistical technique to analyze variation in a response variable (continuous random variable) measured under conditions defined by discrete factors (classification variables, often with

nominal levels).

In order to determine whether or not residue patterns are consistent across two different environmental conditions, a statistical analysis would be conducted on various particle measurements. These measurements include the shape, color, location from the drop edge, and size of the particles.

If a statistical difference is observed for any of these measurements between the two environmental conditions, the residue patterns would be considered different. This means that there is a significant variation in one or more of the particle measurements between the two conditions, indicating that the residue patterns are not the same.

If there is no statistical difference observed between any of the particle measurements, the residue patterns would be labeled as consistent across the two environmental conditions. This means that there is no significant variation in any of the particle measurements, indicating that the residue patterns are the same. Overall this approach would be used to compare residue patterns between two environmental conditions only, and that further research and analysis may be required to compare residue patterns across multiple conditions or other factors.

One-Way ANOVA

The one-way analysis of variance (One-way ANOVA) is also known as single-factor ANOVA or simple ANOVA. As the name suggests, the one-way ANOVA is suitable for experiments with only one independent variable (factor) with two or more levels.

Full Factorial ANOVA (Two-Way ANOVA)

Full Factorial ANOVA, also known as two-way ANOVA, is a statistical method used to determine the effect of two or more independent variables on a dependent variable. It involves using every possible combination of levels of the independent variables in an

experiment, and analyzing the data to see if there is a significant difference in the dependent variable due to the different levels of the independent variables. Two-way ANOVA can also be used to determine if there is an interaction between the independent variables, which means that the effect of one variable on the dependent variable depends on the level of the other variable. This method is useful for experiments where there are multiple factors that could potentially affect the outcome, and allows researchers to gain a more comprehensive understanding of the relationship between the variables.

PERMANOVA

PERMANOVA is an acronym for “permutational multivariate analysis of variance”. It is best described as a geometric partitioning of multivariate variation in the space of a chosen dissimilarity measure according to a given ANOVA design, with p-values obtained using appropriate distribution-free permutation techniques (see Permutation Based Inference; Linear Models: Permutation Methods). The method is semiparametric, motivated by the desire to perform a classical partitioning, as in ANOVA (hence allowing tests and estimation of sizes of main effects, interaction terms, hierarchical structures, random components in mixed models, etc.), while simultaneously retaining important robust statistical properties of rank-based nonparametric multivariate methods, such as the analysis of similarities (ANOSIM2), namely, (1) the flexibility to base the analysis on a dissimilarity measure of choice (such as Bray-Curtis, Jaccard, etc.) and (2) distribution-free inferences achieved by permutations, with no assumption of multivariate normality. Thus, PERMANOVA opens the door for formal partitioning of multivariate data in response to complex experimental designs in a wide variety of contexts: there may be more response variables than sampling units, data may be severely non-normal, zero-inflated, ordinal or qualitative (e.g., responses to questionnaires, DNA/RNA sequences,

allele frequencies, amino acids, or protein data). Although originally motivated by ecological studies, where variables usually consist of counts of abundances (or percentage cover, frequencies, or biomass) for a large number of species, PERMANOVA is now used across many fields, including chemistry, social sciences, agriculture, medicine, genetics, psychology, economics, and more Anderson [2014]. The required assumptions are exchangeability and the linear model and homogeneity of multivariate dispersions.

MANOVA

The Multivariate analysis of variance (MANOVA) procedure provides regression analysis and analysis of variance for multiple dependent variables by one or more factor variables or covariates. The factor variables divide the population into groups. Using this general linear model procedure, the null hypotheses could be tested about the effects of factor variables on the means of various groupings of a joint distribution of dependent variables. The MANOVA could be used to investigate interactions between factors as well as the effects of individual factors. In addition, the effects of covariates and covariate interactions with factors can be included. For regression analysis, the independent (predictor) variables are specified as covariates. Both balanced and unbalanced models can be tested. A design is balanced if each cell in the model contains the same number of cases. In a multivariate model, the sums of squares due to the effects in the model and error sums of squares are in matrix form rather than the scalar form found in univariate analysis. These matrices are called SSCP (sums-of-squares and cross-products) matrices. If more than one dependent variable is specified, the multivariate analysis of variance using Pillai's trace, Wilks' lambda, Hotelling's trace, and Roy's largest root criterion with approximate F statistic are provided as well as the univariate analysis of variance for each dependent variable. In addition to testing hypotheses, Multivariate analysis of variance

(MANOVA) produces estimates of parameters O'Brien and Kaiser [1985].

ANOSIM

Classical one-way ANOSIM operates on an appropriate resemblance matrix calculated among samples, with a factor describing their a priori group structure (e.g. of different sites, times, treatments, etc.) underlying the null hypothesis to be tested, namely H_0 : 'no differences among groups of samples'. If the null hypothesis is true, then the average rank resemblance among samples within groups is expected to be the same as the average rank resemblance among samples from different groups. The ANOSIM statistic R is defined as the scaled difference between the average between-group ($\bar{r}_B$) and within-group ($\bar{r}_W$) ranks:

$$R = \frac{(\bar{r}_B - \bar{r}_W)}{M/2} \quad (3.1)$$

where $M = n(n - 1)/2$ and n is the total number of samples being considered. Clearly, under the null hypothesis, R would be expected to take values (positive or negative) 'close' to zero, and increasing departure from H_0 would result in increasingly larger positive values for R . The scaling in equation 3.1 ensures that R falls within the range -1 to 1, and takes the value $R = 1$ only under maximal separation of the groups, that is if all samples within groups (replicates) are less dissimilar to each other than any pair of samples from different groups. Values of R substantially less than 0 are not usually to be expected as this implies that samples within groups are generally less similar to each other than samples in different groups, a possibility only for a mislabeled or seriously inappropriate design. Note that the usual mathematical terminology for ranks assigns to the highest observation a rank value of 1 (the lowest number). If H_0 is true, then all samples effectively belong to a single group. The spread of possible values of R under the null hypothesis can be determined by randomly permuting the sample labels

and recalculating R for each random reallocation, or for a random subset if there is a large number of possible permutations Hope [1968]. The significance level of the observed value of R is then determined by comparing it to the range of values obtained under permutation, with rejection of the null hypothesis when the observed R is sufficiently large (positive) to have rarely or never occurred under permutation.

Jensen-Shannon divergence

The Jensen-Shannon divergence is a measure of similarity between two probability distributions. It is a symmetric and finite variant of the Kullback-Leibler divergence, also known as information radius Nielsen [2021], Manning and Schutze [1999] or total divergence to the average Dagan et al. [1997]. The square root of the Jensen-Shannon divergence is also known as the Jensen-Shannon distance Endres and Schindelin [2003], Osterreicher and Vajda [2003], Fuglede and Topsøe [2004], and it is a metric that can be used to compare two probability distributions. It is commonly used in information theory, machine learning, and natural language processing, among other fields.

Multidimensional scaling (MDS)

Multidimensional scaling is a visual representation of distances or dissimilarities between sets of objects. “Objects” can be colors, faces, map coordinates, political persuasion, or any kind of real or conceptual stimuli Kruskal and Wish [1978]. Objects that are more similar (or have shorter distances) are closer together on the graph than objects that are less similar (or have longer distances). As well as interpreting dissimilarities as distances on a graph, MDS can also serve as a dimension reduction technique for high-dimensional data Buja et al. [2008].

Noise removal with singular value decomposition (SVD)

Singular value decomposition (SVD) is a mathematical technique by which a matrix is

decomposed into a product of three matrices, which can also be written as a sum of rank-one matrices. SVD could be regarded as a generalization of eigen decomposition, a technique employed to decompose a positive semidefinite normal matrix. This relationship makes SVD connected to principal component analysis (PCA), a technique commonly utilized for data analysis and representation. One example of SVD application can be found in image processing. A digital image can be represented by a matrix, where the value of a matrix element encodes information about a specific pixel. By breaking down this matrix using SVD, the image can be simplified, and useful information can be extracted. Another application of SVD is observed in signal processing, where it is employed to remove noise from biomedical signals and construct signal and noise subspaces for analysis and approximation.

3.3.6 Experiment procedure

This research comprises three stages. In the first stage, data collection, SEM, EDS, and camera photographs related to the coffee-ring effect were gathered and the images were preprocessed. The second stage focused on method optimization, during which the required extent of temperature and humidity control to maintain consistent residue patterns was examined through the coffee-ring effect. The final stage involved identifying the optimal environmental conditions for separating contaminant particles from one another (such as calcium, sodium, magnesium, etc.) using the statistical analysis introduced earlier.

Stage 1: Collection of coffee-ring effect residue pattern

Stage 1 was divided into two subtasks. Task 1a) involved collecting the coffee-ring effect SEM, EDS, and camera photographs. Task 1b) focused on preprocessing the images gathered in Task 1a by implementing noise removal, color normalization, and other techniques.

Data collection: To investigate the effect of environmental conditions on coffee-ring effect patterns, nine temperature and relative humidity (RH) combinations were maintained by the auto temperature humidity control chamber, and a four-axis autosampler was placed inside the chamber (Table. 3.1). During the droplet dropping process, each water sample was stored in a 2 μL micro-centrifuge tube and placed in the sample holder. In each experiment, sixteen samples were positioned at once in one sample holder. The autosampler was programmed to collect 2 μL water samples and inject them onto the aluminum substrate (6061 with mirror-like finish, McMaster-Carr 1655T1) as described in previous research Li et al. [2020]. After each water sample injection, the injector tip was rinsed through a programmed procedure in nanopure water.

To prevent the influence from other droplets' drying processes, droplets were placed 1 cm apart, and ten droplets were dried at once on one aluminum substrate (1 inch wide and 3 inches long). To avoid vibrations from autosampler motors, the aluminum slides were positioned on an independent sample stage detached from the autosampler. The auto temperature humidity control chamber not only maintained the desired temperature and humidity but also prevented air flow in the environment. Two microliter droplets of each of the five water samples would be deposited on a mirrored aluminum slide and allowed to dry, separating particles that form through the coffee-ring effect Li et al. [2020]. Five water droplet replicates were collected under each environmental condition.

A low-cost camera photograph were used for all replicates, using 100X magnification and the Celestron camera, including a color bar in all images to normalize brightness, contrast, and color. The total number of collected photographs is 225 (9 environmental conditions, 5 water recipes with 5 replicates). Residues were saved for further analysis.

Image preprocessing: In image preprocessing, images were color-normalized based on

RGB distribution. Images were loaded using the `imread` function and converted to binary with the `im2bw` function (threshold set to 0.2). Noise was removed from each binary image using the `medfilt2` function with an [8, 8] square parameter. Particle edges were captured using the `edge` function with the canny method applied to the smoothed binary image. Particle properties were extracted from the smoothed binary image using the `regionprops` function with 'Area', 'Perimeter', 'Eccentricity', 'Orientation', and 'Centroid' methods. In each SEM-EDS map, a 2-D coordinate was established with origin on the center of the droplet pattern in Matlab. The deposition position of each particle for each element were recorded as a x-y value in the coordinate. Because particles were deposited as a circle around the residue center, particle locations were calculated by the distance between the particle's location and the coordinate center. The adjusted centroid was recalculated by taking the square root of the sum of squares of the differences between the centroid x-coordinate and the image center x-coordinate, and the centroid y-coordinate and the image center y-coordinate.

Stage 2: Optimization of tap water fingerprints

Stage 2 was divided into three subtasks. **Task 1a)** Determine the ranges of temperature, relative humidity that have consistent coffee-ring fingerprints. **Task 1b)** Find the optimal ranges of temperature, relative humidity to separate contaminants particles from each other. **Task 1c)** Investigate the elements deposition separation effect under each environmental condition.

$$\text{Adjusted coordinate} = \sqrt{(X_{\text{centroid}} - X_{\text{center}})^2 + (Y_{\text{centroid}} - Y_{\text{center}})^2} \quad (3.2)$$

Task 1a: Determine the ranges of temperature, relative humidity over which coffee-ring fingerprints are constant.

In order to implement this method broadly for analyzing samples across a distribution

system, it is essential to accommodate analysis in various laboratories and field settings. This task assessed the extent of temperature and humidity control needed to produce consistent tap water fingerprints. The proposed nine temperature and humidity conditions Table. 3.2 were evaluated using PERMANOVA on coffee-ring effect residue pattern features.

Task 1b: Find the optimal ranges of temperature, relative humidity that different water samples exhibit different coffee-ring effect residue patterns.

This task aimed to investigate the optimal temperature and relative humidity conditions under which different water samples exhibit different coffee-ring effect residue pattern. In the previous task, the optimal temperature and relative humidity conditions were determined to exhibit consistent coffee-ring effect residue pattern. However, only have similar residue pattern is not enough to distinguish different water samples. This task utilized PERMANOVA, MANOVA, ANOVA techniques to investigate the coffee-ring effect residue pattern feature statistics under different environmental conditions. Jensen-Shannon divergence was used to measure the similarity between different water samples and classical multidimensional scaling (NMDS) was used to visualize the differences in the coffee-ring effect residue pattern features between different water samples.

Task 1c: Investigate the optimal ranges of temperature, relative humidity to separate contaminants particles from each other.

This task is to investigate whether specific elements are associated with residue particles, EDS mapping images were used to identify particle compositions in coffee-ring effect residue patterns. The locations of elements were determined by calculating the square root of the x-axis and y-axis relative to the center of each image. Analysis of variance (ANOVA) was conducted on the element locations to examine whether there were

any significant differences in the spatial distribution of elements within the residue patterns.

Stage 3: Identify the correlation between water sample coffee-ring effect patterns and element deposition compositions

The EDS images were preprocessed using Singular Value Decomposition (SVD) and noise was filtered using the `medfilt2` function with filter size [3, 3]. After preprocessing, the element compositions were extracted from the EDS mappings.

To determine the composition ratio of each element in the corresponding particle, the particles extracted from the water samples coffee-ring effect patterns were compared with the pixel signals extracted from the EDS data. The composition ratio of each element in each particle was then calculated. To investigate whether there is a significant difference in the element composition ratios between particles, the correlations of these ratios were calculated, and ANOVA was conducted on these ratios.

3.4 Results and Discussion

3.4.1 Under what environmental conditions are coffee-ring effect fingerprints are consistent

PERMANOVA on coffee-ring effect residue pattern features (particle shape, color, location from the drop edge, and size) results has shown in Table. 3.5. In all the nine temperature and relative humidity combination conditions, the p-values are all smaller than 0.001. Based on the p-values which has the same degree of freedom of 4, all the coffee-ring effect residue patterns are consistent between sample replicates and different between different samples. However the R^2 of all the nine conditions are ranging from 0.716 to 0.957. PERMANOVA on coffee-ring effect residue pattern features visualization results has shown in Table. 6.12. According to the visualization results (manhattan distance

applied), under the condition A (20-23 °C, 35%-40%), most samples have been separated except samples A and B. However, sample A and sample B have similar recipes according to Table. 3.4. At the same time, sample D and sample E have similar water components, and their positions in the PERMANOVA visualization result are near to each other Figure. 3.2. Based on the visualization result, the samples coffee-ring effect residue pattern features are mostly differentiable under the condition C (20-23 °C, 45%-50%) (Figure. 3.3) and H (26-29 °C, 40%-45%) (Figure. 3.4). Across the nine conditions, the sample C (0.1 mM $NaHCO_3$, 0.5 mM $CaCl_2$, 0.2 mM $MgCl_2$, 0.35 mM Na_2SO_4 , 0.033 mM NaH_2PO_4 , 0.4 mM KF , 0.005 mM $Fe(NO_3)_3$, 0.00024 mM $CuSO_4$) is the most stable one that all replicates clustered in a smaller range and not overlapping with other samples. Sample E (0 mM $NaHCO_3$, 1 mM $CaCl_2$, 0.5 mM $MgCl_2$, 2.35 mM Na_2SO_4 , 0.033 mM NaH_2PO_4 , 0.4 mM KF , 0.05 mM $Fe(NO_3)_3$, 0.00024 mM $CuSO_4$) is the most unstable one that spreads most among the five water samples. This could be explained by with higher humidity, there is more vapor-liquid exchange of water molecules. So in the particles formation process, there is more time for the particles to crystalize and in the droplet-air interface, the water density gradient decreases smaller than low humidity conditions. This smaller water density induces the particles formed in a slower and gradual manner which results in crystals formed in different phases of droplet drying processes and formed unique residue patterns. At high temperature conditions, the particles residue pattern features are not only spread out but also replicates of the same sample clustered closer. The reason is at higher temperature conditions, at the moment of crystallization, crystals formed at a relative speed so the patterns features are more consistent between replicates, for example in condition H (26-29 °C, 40%-45%) and I (26-29 °C, 45%-50%). According to the overall analysis, the suitable conditions to produce

consistent residue patterns are the conditions with high temperature and relative humidity such as condition C, condition F, condition H and condition I. All condition results is shown in Table. 6.11.

Table 3.5: PERMANOVA analysis for particle features

Condition	Df	Sum of Sqs	Mean Sqs	F.Model	R2	Pr(>F)	sig.
A	4	1.09×10^{11}	2.71×10^{10}	113.28	0.95773	0.001	***
B	4	2.87×10^8	7.18×10^9	64.175	0.92772	0.001	***
C	4	1.91×10^{11}	4.78×10^{10}	15.567	0.75689	0.001	***
D	4	1.67×10^{11}	4.18×10^{10}	71.904	0.93498	0.001	***
E	4	1.49×10^{11}	3.73×10^{10}	12.651	0.71673	0.001	***
F	4	4.72×10^{11}	1.18×10^{11}	15.542	0.7566	0.001	***
G	4	5.98×10^{11}	1.49×10^{11}	8.1009	0.61835	0.001	***
H	4	2.08×10^{11}	5.20×10^{10}	27.386	0.84561	0.001	***
I	4	7.43×10^{11}	1.86×10^{11}	24.709	0.8317	0.001	***

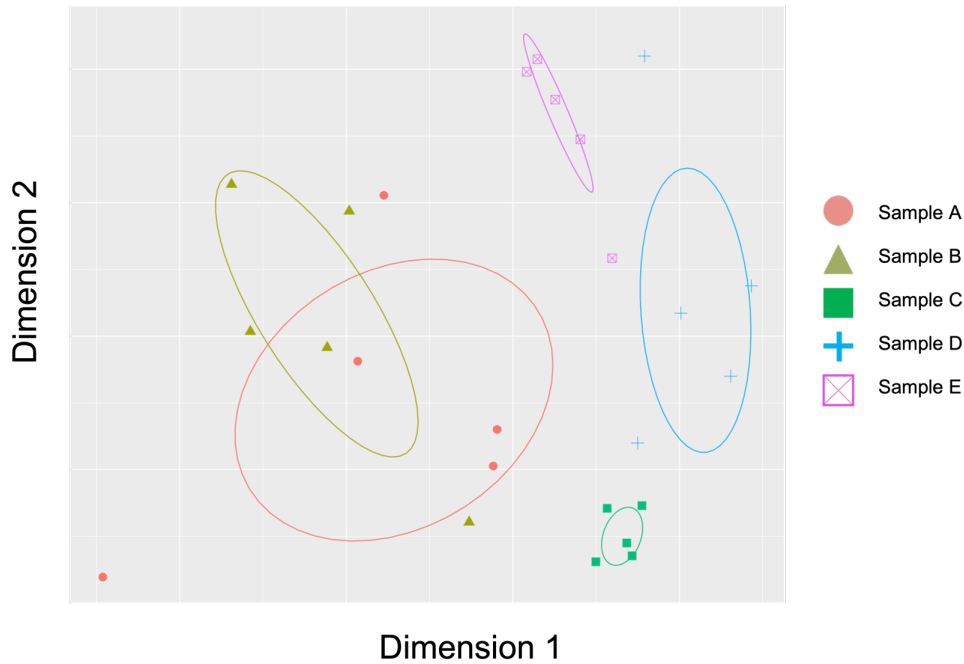


Figure 3.2: PERMANOVA of condition A

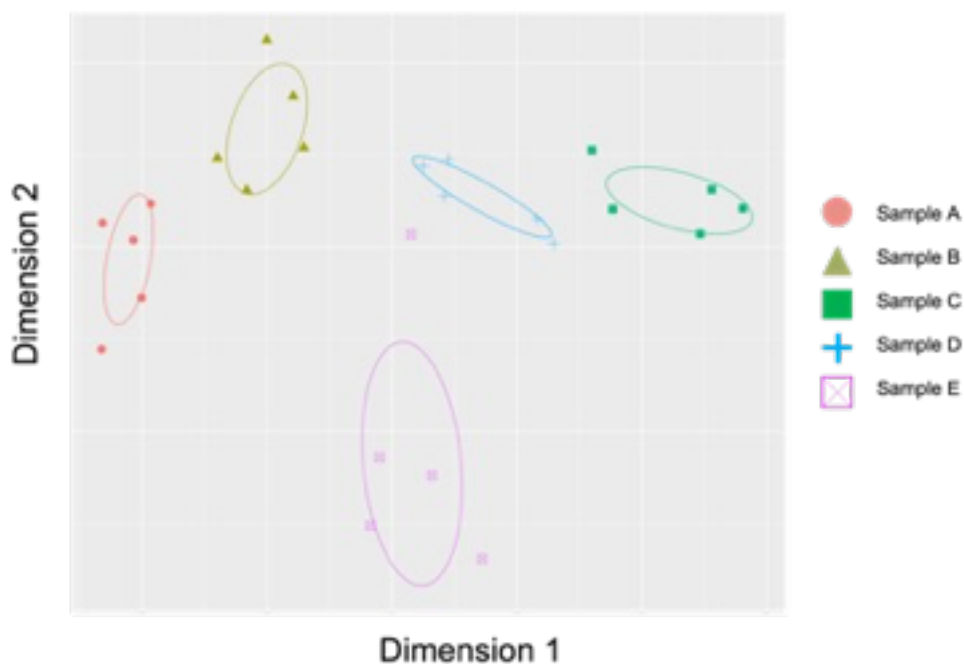


Figure 3.3: PERMANOVA of condition C

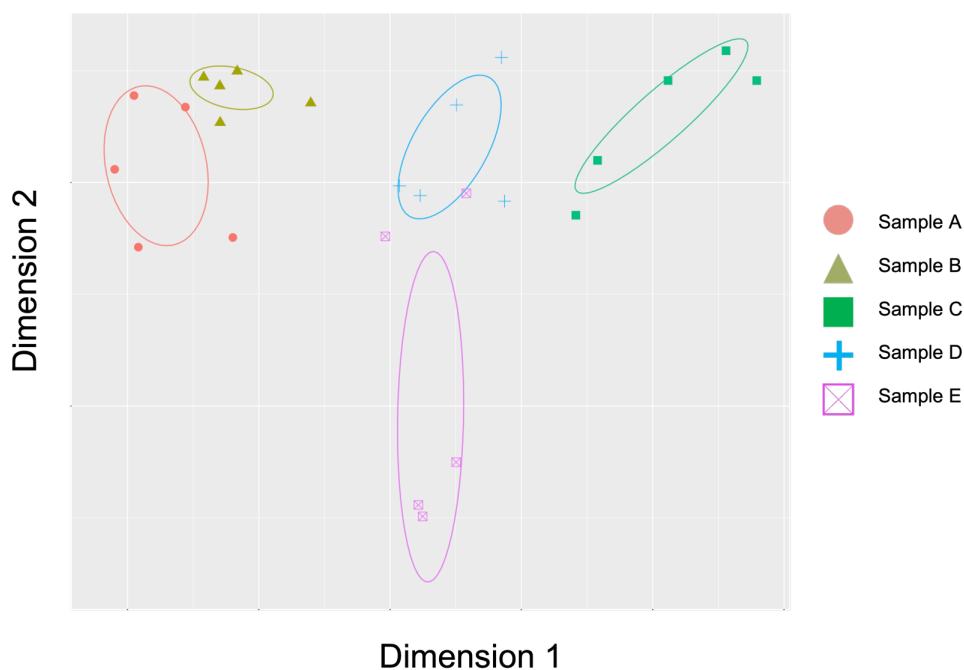


Figure 3.4: PERMANOVA of condition H

3.4.2 What are the optimal environmental conditions that different water samples exhibit mostly different coffee-ring effect residue patterns

To investigate the optimal environmental conditions for separating particles in the coffee-

ring effect residue pattern, water samples coffee-ring effect residue particles feature statistics were analyzed under varying conditions. The most statistically significant particle features were identified through multivariate analysis of variance (MANOVA) on water samples and environmental conditions.

The study found that factors such as mean area, mean perimeter, mean eccentricity, standard deviation of area, standard deviation of centroid, and standard deviation of orientation influenced the coffee-ring effect residue pattern features. These results, as shown in Table 3.6, provide insight into the conditions that promote a more visible and distinct coffee-ring effect residue pattern. According to the findings, particle features such as area, perimeter, eccentricity, and centroid are sensitive to environmental conditions, with 'class' representing water samples and 'condition' representing environmental conditions in Table 3.6.

Table 3.6: MANOVA analysis for image properties

Responses						
Response area mean						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
class	4	38382	9595.6	44.1302	$< 2.2 \times 10^{-16}$	***
condition	8	7450	931.3	4.2831	8.562×10^{-05}	***
Residuals	212	46097	217.4			
Response area std						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
class	4	1930667	482667	46.1313	$< 2.2 \times 10^{-16}$	***
condition	8	274996	34375	3.2854	0.001482	***
Residuals	212	2218133	10463			
Response eccentricity mean						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
class	4	0.121804	0.0304510	42.7562	$< 2.2 \times 10^{-16}$	***
condition	8	0.019479	0.0024348	3.4187	0.001016	**
Residuals	212	0.150987	0.0007122			
Response eccentricity std						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
class	4	0.0109677	0.00274192	21.4681	6.974×10^{-15}	***

Table 3.6: (cont'd)

condition	8	0.0017548	0.00021935	1.7174	0.09579	.
Residuals	212	0.0270767	0.00012772			
Response orientation mean						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
class	4	555.90	138.975	11.9927	8.368×10^{-09}	***
condition	8	189.02	23.628	2.0389	0.04333	*
Residuals	212	2456.72	11.588			
Response orientation std						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
class	4	56.00	14.000	2.3224	0.05782	.
condition	8	279.66	34.957	5.7991	1.074×10^{-06}	***
Residuals	212	1277.94	6.028			
Response perimeter mean						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
class	4	73599	18399.8	53.9503	$< 2.2 \times 10^{-16}$	***
condition	8	10876	1359.5	3.9863	0.0002011	***
Residuals	212	72303	341.1			
Response perimeter std						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
class	4	4692837	1173209	42.6128	$< 2 \times 10^{-16}$	***
condition	8	500897	62612	2.2742	0.0236	*
Residuals	212	5836749	27532			
Response centroid mean						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
class	4	401278	100319	12.530	3.614×10^{-09}	***
condition	8	878527	109816	13.717	5.513×10^{-16}	***
Residuals	212	1697283	8006			
Response centroid std						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
class	4	89628	22406.9	17.906	1.108×10^{-12}	***
condition	8	200351	25043.8	20.013	$< 2.2 \times 10^{-16}$	***
Residuals	212	265294	1251.4			

Table 3.6 demonstrates the coffee-ring effect residue pattern variabilities with the interaction between environmental conditions and water samples. However, the coffee-ring effect pattern variabilities of water samples without environmental conditions are not clear. In the ANOVA analysis of coffee-ring effect residue pattern features (Table 3.7),

area mean, area standard deviation, perimeter mean, perimeter standard deviation, centroid mean, centroid standard deviation, and eccentricity mean are statistically significant across the nine experiment conditions. The area mean is mostly significant at the 10^{-6} level (conditions A, C, E, F, G, H, I) and lower, with only two conditions (B, D) having larger statistical significance at 10^{-2} – 10^{-3} . This result suggests that the area mean significantly differs between water samples under most test environmental conditions. It aligns with the results in Table 6.12, where particle positions in the PERMANOVA visualization image are mixed under conditions B and D. This confirms that particles formed by different water samples exhibit distinct coffee-ring effect residue patterns.

Table 3.7: P-value of ANOVA of coffee-ring effect residue pattern features under each experiment condition

Condition	Area mean	Perimeter mean	Centroid mean	Eccentricity mean	Orientation mean	Area std	Perimeter std	Centroid std	Eccentricity std	Orientation std
A	5.62×10^{-6}	3.11×10^{-7}	1.57×10^{-3}	2.7×10^{-6}	5.07×10^{-2}	3.58×10^{-6}	1.20×10^{-6}	7.5×10^{-6}	2.64×10^{-3}	8.07×10^{-2}
B	9.51×10^{-2}	1.71×10^{-2}	4.04×10^{-8}	5.71×10^{-3}	7.16×10^{-3}	8.74×10^{-7}	2.55×10^{-6}	1.07×10^{-10}	3.83×10^{-2}	1.67×10^{-1}
C	2.46×10^{-10}	1.26×10^{-9}	2.66×10^{-2}	5.13×10^{-9}	4.32×10^{-3}	1.26×10^{-7}	3.78×10^{-7}	1.10×10^1	4.60×10^{-2}	3.12×10^{-1}
D	7.70×10^{-3}	4.56×10^{-5}	3.23×10^{-6}	1.36×10^{-5}	1.75×10^{-1}	2.73×10^{-4}	2.39×10^{-5}	8.48×10^{-4}	3.12×10^{-3}	1.22×10^{-1}
E	9.50×10^{-12}	4.19×10^{-12}	4.86×10^{-6}	1.09×10^{-4}	4.62×10^{-3}	3.23×10^{-15}	1.09×10^{-13}	6.57×10^{-5}	1.31×10^{-3}	1.12×10^{-2}
F	8.41×10^{-7}	6.43×10^{-9}	2.99×10^{-2}	5.61×10^{-5}	1.55×10^{-1}	2.50×10^{-6}	1.20×10^{-6}	7.38×10^{-5}	4.50×10^{-4}	6.48×10^{-1}
G	1.23×10^{-6}	1.62×10^{-10}	1.45×10^{-6}	4.66×10^{-7}	4.52×10^{-2}	9.30×10^{-6}	4.92×10^{-8}	8.60×10^{-5}	3.21×10^{-7}	5.14×10^{-5}
H	1.43×10^{-9}	3.65×10^{-11}	1.47×10^{-2}	9.72×10^{-6}	1.68×10^{-1}	1.24×10^{-6}	7.94×10^{-7}	2.96×10^{-2}	2.46×10^{-6}	2.80×10^{-1}
I	6.74×10^{-6}	2.41×10^{-6}	3.02×10^{-1}	5.00×10^{-3}	3.73×10^{-1}	3.16×10^{-5}	2.51×10^{-5}	9.13×10^{-7}	3.69×10^{-2}	4.76×10^{-3}

For the perimeter mean variable, the nine conditions show similar statistical results, with the perimeter mean mostly significant at the 10^{-5} level (conditions A, C, D, E, F, G, H, I) and lower, except for one condition (B) with a larger statistical significance of 1.71×10^{-2} . The larger significance value also contributes to point mixing in the PERMANOVA under condition B (Table 6.11).

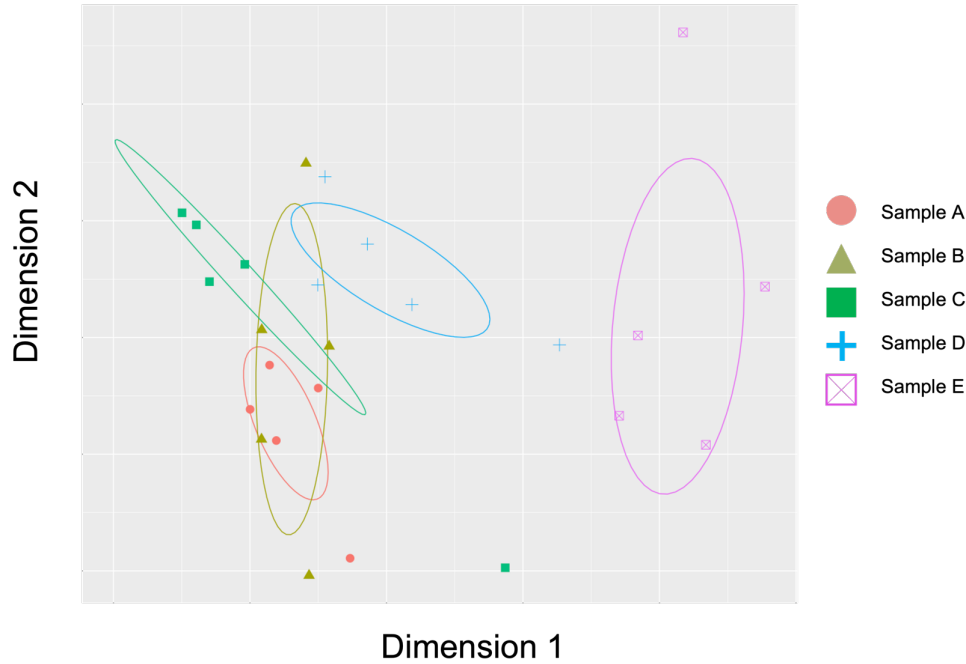


Figure 3.5: PERMANOVA of condition B

Although centroid mean is statistically significant in water sample coffee-ring effect residue pattern features, the significance levels are lower than those of area mean and perimeter mean features, with five significance values greater than 10^{-3} among nine conditions. This occurs because the shapes of formed particles are similar, leading to similar centroid calculations among particles. Interestingly, despite condition B having larger significance values for area mean and perimeter mean, the significance value of centroid under condition B is smaller than other conditions.

Eccentricity values are similar to centroid, with larger significance values than area

mean and perimeter mean but smaller values than centroid. In contrast, orientation mean shows much larger significance values due to particles forming in the droplet drying process, from the droplet edge to the droplet center, resulting in the same orientations.

The standard deviations of coffee-ring effect residue particle area, perimeter, centroid, eccentricity, and orientation display similar results to the feature means: area and perimeter standard deviations have the highest levels of statistical significance, centroid and eccentricity standard deviations have lower levels of statistical significance, and orientation standard deviation has the lowest significance levels. However, unlike the residue particle feature mean values, the feature standard deviation values do not correlate with the PERMANOVA of residue particle features.

The ANOVA on coffee-ring effect residue pattern features of each water sample follows the same trend: particle area and perimeter features have the highest statistical significance levels, centroid and eccentricity have lower statistical significance, and orientation has the lowest significance levels (Table 3.8).

Table 3.8: P-value of ANOVA of coffee-ring effect residue pattern features of water samples

Samples	Area mean	Perimeter mean	Centroid mean	Eccentricity mean	Orientation mean	Area std	Perimeter std	Centroid std	Eccentricity std	Orientation std
Sample 1	5.16×10^{-4}	4.2×10^{-3}	4.97×10^{-9}	3.92×10^{-5}	4.1×10^{-1}	2.02×10^{-1}	6.26×10^{-1}	1.87×10^{-4}	3.24×10^{-3}	4.66×10^{-1}
Sample 2	8.89×10^{-3}	7.48×10^{-3}	2.63×10^{-9}	1.37×10^{-5}	4.92×10^{-3}	9.81×10^{-2}	3.99×10^{-2}	1.05×10^{-8}	6.9×10^{-2}	7.38×10^{-4}
Sample 3	4.72×10^{-11}	1.2×10^{-12}	1.18×10^{-3}	5.23×10^{-5}	6.77×10^{-1}	3.2×10^{-8}	2.87×10^{-8}	5.64×10^{-9}	5.75×10^{-6}	1.28×10^{-5}
Sample 4	2.05×10^{-6}	1.79×10^{-6}	1.23×10^{-4}	1.14×10^{-3}	8.78×10^{-2}	4.51×10^{-6}	5.55×10^{-7}	6.68×10^{-10}	4.08×10^{-2}	2.22×10^{-3}
Sample 5	2.60×10^{-3}	1.65×10^{-4}	2.90×10^{-12}	3.95×10^{-5}	4.21×10^{-1}	1.35×10^{-3}	1.67×10^{-4}	7.45×10^{-12}	6.61×10^{-2}	1.24×10^{-3}

In a previous analysis of variance (ANOVA), the significance of each coffee-ring effect residue pattern feature was evaluated independently, considering water sample class or environmental condition separately. To confirm the statistical significance of these pattern features, a multivariate analysis of variance (MANOVA) was conducted for each environmental condition individually, as shown in Table 3.9.

Based on the MANOVA results, four conditions (A, B, D, and F) exhibited a statistically significant residue area feature. Five conditions (A, B, C, E, and H) showed a statistically significant residue eccentricity feature. Seven conditions (A, B, D, E, F, H, and I) demonstrated a statistically significant residue centroid feature. Unlike the ANOVA results, the orientation feature was not found to be statistically significant under any condition. However, only one condition (B) had a statistically significant residue perimeter feature. This discrepancy is likely due to the MANOVA algorithm accounting for the correlations between the residue features.

Table 3.9: MANOVA of coffee ring effect residue pattern features of water samples under each condition

Condition	Area mean	Perimeter mean	Centroid mean	Eccentricity mean	Orientation mean	Area std	Perimeter std	Centroid std	Eccentricity std	Orientation std
A	0.6364	0.456	0.0001	4.8×10^{-5}	0.016	0.0003	0.003	1.15×10^{-6}	0.012	0.032
B	0.408	0.059	1.12×10^{-8}	0.035	0.80	0.0001	8.4×10^{-5}	1.99×10^{-8}	0.0025	0.09
C	0.18	0.38	0.29	4.7×10^{-5}	0.062	0.01	0.2	0.5	0.254	0.22
D	0.23	0.51	0.001	0.038	0.366	0.0005	0.1	0.001	0.268	0.79
E	0.179	0.117	3.9×10^{-7}	0.006	0.05	0.01	0.014	0.007	0.004	0.19
F	0.40	0.44	0.38	0.03	0.06	0.0016	0.037	0.0001	0.8	0.31
G	0.95	0.91	0.059	0.5	0.81	0.037	0.21	0.185	0.72	0.15
H	0.52	0.54	0.006	0.0006	0.099	0.076	0.182	0.0012	0.63	0.39
I	0.83	0.756	0.249	0.6	0.59	0.042	0.173	1.75×10^{-6}	0.88	0.001

ANOSIM for coffee-ring effect residue pattern features

To account for the correlation between the experiment environmental conditions and water sample classes, ANOSIM (with Canberra dissimilarity index) was conducted on coffee-ring effect residue pattern features. The results are shown in Table 6.12.

According to the ANOSIM results, conditions C, E, G, and H are the conditions where coffee-ring effect residues of the same water recipe are more distinguishable than those of water samples with different components. Conditions A and F display comparable differences between water samples with the same components and those with different components. However, conditions B and I show the least similarity in residue patterns for the same water components and different residue patterns for different water components. The ranking of water sample residue pattern similarity for the same water components compared to water with different components is C, H, E, G, F, A, D, B, I. Statistically, the null hypothesis is that there is no difference between the means of two or more groups of (ranked) dissimilarities. The ANOSIM statistic R (Table 3.10) and significance values can be compared to test this hypothesis.

Table 3.10: R-value ANOSIM result of water samples coffee-ring effect residue patterns

Temperature °C Relative Humidity (R-value)	20-23 (°C)	23-26 (°C)	26-29 (°C)
35%-40%	0.6344	0.5459	0.7600
40%-45%	0.5366	0.7706	0.7922
45%-50%	0.8643	0.7202	0.5366

ANOSIM was conducted on each particle feature of the coffee-ring effect residue pattern features to investigate the variability of particle area, perimeter, eccentricity, and centroid in relation to water samples and environmental conditions. Under each environmental condition, Jensen-Shannon divergence was calculated based on particle area, perimeter, and eccentricity. Multidimensional scaling and classical multidimensional

scaling coordinates were then derived from the Jensen-Shannon distance matrix.

ANOSIM for coffee-ring effect residue pattern area feature

The ANOSIM result for the coffee-ring effect residue pattern area feature is shown in Table 6.13. In this result, the upper right and lower left triangles are the same due to the interchangeability of distances between two replicate residues. Also, in condition A results, for images from 11 to 15, the distances between the replicates are smaller than the distances between these replicates and other samples, demonstrating the consistency of coffee-ring effect residue patterns. Conditions C, F, and H all display relatively smaller distances within water samples than distances between samples. Under the high relative humidity conditions (conditions C, F, and I), sample E (1 mM CaCl_2 , 0.5 mM MgCl_2 , 2.35 mM Na_2SO_4 , 0.033 mM NaH_2PO_4 , 0.4 mM KF , 0.005 mM $\text{Fe}(\text{NO}_3)_3$, and 0.00024 mM CuSO_4) exhibits relatively greater distinct residue patterns compared to other water samples.

The CMDS coordinates of the ANOSIM results are shown in Table 6.14. In this table, it is clear that the coffee-ring effect residue patterns of replicates for each water sample are clustered near each other under conditions C, F, and H. However, the projected points under conditions A, B, and D are mixed together. Therefore, based on the residue pattern area feature, conditions C, F, and H are suitable for separating water contaminant particles from each other.

ANOSIM for coffee-ring effect residue pattern perimeter feature

The ANOSIM result for the coffee-ring effect residue pattern perimeter is shown in Table 6.15. Based on the results, under conditions D, G, and H, the similarities between water samples C (0.1 mM NaHCO_3 , 0.5 mM CaCl_2 , 0.2 mM MgCl_2 , 0.35 mM Na_2SO_4 , 0.033 mM NaH_2PO_4 , 0.4 mM KF , 0.005 mM $\text{Fe}(\text{NO}_3)_3$, and 0.00024 mM CuSO_4)

differ from those of water samples D (1 mM CaCl_2 , 1 mM MgCl_2 , 1.35 mM Na_2SO_4 , 0.033 mM NaH_2PO_4 , 0.4 mM KF, 0.005 mM $\text{Fe}(\text{NO}_3)_3$, and 0.00024 mM CuSO_4) and E (1 mM CaCl_2 , 0.5 mM MgCl_2 , 2.35 mM Na_2SO_4 , 0.033 mM NaH_2PO_4 , 0.4 mM KF, 0.005 mM $\text{Fe}(\text{NO}_3)_3$, and 0.00024 mM CuSO_4). The reason is that samples D and E do not contain NaHCO_3 .

Additionally, only under condition A do the replicates of water samples C, D, and E produce consistent residue patterns; under other temperature and relative humidity conditions, water samples A, B, and C produce more consistent residue patterns. Furthermore, under conditions B, C, D, F, and I, sample E produces different residue patterns than samples A, B, C, and D. In the nanochromatography (Table 6.24), sample E is prone to forming an olive-shaped residue with a strong edge. Especially under conditions D and G, sample E has difficulty maintaining a convex shape residue, which results from the shrinkage of the residue during the droplet drying process.

The CMDS coordinates of the ANOSIM results are shown in Table 6.16. The sample separation and replicate clustering results are not as strong as those for the residue pattern area feature. This is because non-convex shaped residues can produce the same sized residue pattern but with a much larger perimeter. Only under conditions C and F are the water samples with different components separated, and replicates with the same recipe are clustered together.

ANOSIM for coffee-ring effect residue pattern centroid feature

The ANOSIM result for the coffee-ring effect residue pattern centroid is shown in Table 6.17. Based on the results, only under condition C do the replicates of water samples produce similar residue pattern centroid features, and water samples with different components produce different residue patterns. Under conditions A and B, water samples

C and D produce similar residue pattern centroid features.

The reason that the centroid feature is not a suitable metric to distinguish water samples with different water components is that the formed particles in the residue have a similar centroid, which originates from the formation of the particles. During the droplet drying process, particles are formed from the droplet edge to the droplet center, and they are formed in the same direction, resulting in particles with similar centroids (see Table 6.27). The CMDS coordinates of the ANOSIM results are shown in Table 6.18. As shown in the centroid ANOSIM results, only under condition C do water samples with different components' residue patterns produce different centroid features and have different coordinates in the CMDS plot. Replicates of water samples with the same components produce similar centroid residue patterns and have similar coordinates in the CMDS plot. However, under conditions A, D, and G, the water sample C points are separable from the other water samples (see Table 6.18). This is consistent with the results in the ANOSIM results, where under condition G, water sample C (replicates 11 to 15) residue patterns have more similar centroid features than the other water replicates. This phenomenon occurs under conditions with lower relative humidity, where the concentration of 0.1 mM $NaHCO_3$, 0.5 mM $CaCl_2$, and 0.2 mM $MgCl_2$ is lower. These low component concentrations result in slower particle formation, so only when the droplet shrinks to a smaller size will particles form, and the formed particles are larger than particles formed under other conditions (see Table 6.24, Table 6.25, and Table 6.26).

ANOSIM for coffee-ring effect residue pattern eccentricity feature

The ANOSIM results for coffee-ring effect residue pattern eccentricity are shown in Table 6.19. Based on the results, under condition A, the replicates of water samples A and B have similar eccentricity features, and water samples C, D, and E have similar

eccentricity features. However, water samples A, B, and C form one group, and water samples D and E form another group. Under conditions C, D, G, and H, water sample A exhibits its own eccentricity feature. Under condition H, all five water samples exhibit distinct eccentricity features.

The CMDS coordinates of ANOSIM results are shown in Table 6.20. Under conditions B, C, and D, all replicate points are mixed together in a small region and cannot be separated effectively. Under condition G, replicate points are separated by their components; however, these points are too close together, making it difficult to find a clear rule for separating them and using them for further prediction. The water samples are separated maximally under condition H; however, there are two drawbacks in this condition. First, the replicates of water sample A are not clustered in a small region, indicating that the replicates' consistency is not optimal, as shown in Table 6.19. Second, samples B and C are too close to each other in the CMDS plot.

3.4.3 Under each environmental condition, are the elements deposition locations significantly different from each other

Previous analyses have shown that both environmental conditions and water chemistry have statistically significant effects on coffee-ring effect patterns. However, these analyses did not provide information on whether the elements were separated in each residue pattern. To investigate this, EDS mapping images were used to label particle compositions in coffee-ring effect residue patterns. The locations of elements were calculated as the square root of the x-axis and y-axis relative to the center of each image. The p-value of the analysis of variance (ANOVA) was found to be smaller than 2×10^{-16} , indicating that environmental conditions and water sample class have significant statistical effects on element distributions. This suggests that different elements are separated by the coffee-ring

effect.

The carbon, chlorine and sulfur elements Two-way ANOVA results is shown in Table. 6.21. All the tests on these nine conditions have degree of freedom of 4 for class variable, degree of freedom of 2 for elements variable and degree of freedom of 8 for class:elements (class stands for water samples, elements stands for elements, carbon, chlorine and sulfur in this case). Based on these tests, all these nine conditions have shown statistical significance that the p-value is smaller than 2×10^{-16} . Comparing the F values respect to the elements of under these nine conditions, condition A and C have the value around 5600 and condition D have the value around 8600 which is the highest value in these nine conditions. This results concludes the carbon, chlorine and sulfur are mostly separated under condition D than condition A and C and other conditions. Comparing the F values respect to the class variable, condition C, D and G all have shown largest F values (F values in the range of 400-470) which means the carbon, chlorine and sulfur elements are mostly separated in the coffee-ring effect residue pattern under these environmental conditions with respect to the water components recipe. Comparing the class to elements correlation, the carbon, chlorine and sulfur are mostly separated under the C, D, F and G (F values in the range 400-600) conditions which is consistent with the ANOSIM of residue pattern features result.

The Two-way ANOVA results for calcium, magnesium, and sodium elements are presented in Table 6.22. All nine tests have a degree of freedom of 4 for the class variable, 2 for the elements variable, and 8 for class:elements (where 'class' represents water samples, 'elements' represents the elements calcium, magnesium, and sodium in this case). Based on these tests, all nine conditions showed statistical significance, with p-values smaller than 2×10^{-16} .

When comparing the F-values with respect to the elements under these nine conditions, conditions A, B, C, and I had values around 3000, while condition D had the highest value at around 3800. This indicates that calcium, magnesium, and sodium are more effectively separated under condition D compared to A, B, C, and the other conditions.

When comparing the F-values with respect to the class variable, conditions B, C, D, and E showed the largest F-values (in the range of 50 to 150), suggesting that calcium, magnesium, and sodium elements are more effectively separated in the coffee-ring effect residue pattern under these environmental conditions with respect to the water components recipe.

Furthermore, when comparing the class to elements correlation, calcium, magnesium, and sodium were more effectively separated under conditions B, C, D, and E (with F-values in the range of 100 to 180), which is consistent with the ANOSIM analysis of residue pattern features for carbon, chlorine, and sulfur.

Overall, these results suggest that conditions B, C, D, and E are the most effective for separating calcium, magnesium, and sodium elements in the coffee-ring effect residue pattern. Previous analyses have demonstrated that environmental conditions and water chemistry have statistically significant effects on the coffee-ring effect pattern and the distribution of element components in water samples. However, it remains unclear whether there is a correlation between the coffee-ring effect patterns and the element compositions of water samples, which is crucial for building models to recognize and quantify contaminants. In previous analyses, we identified several optimal conditions that produced consistent replicates of water sample residue patterns and distinct residue patterns for different water components. The following analysis aims to investigate under which environmental conditions the coffee-ring effect patterns of water samples are correlated

with element compositions. This analysis will provide insight into the relationship between the residue patterns and the underlying elemental components, which can be used to develop more accurate models for detecting and quantifying contaminants.

3.4.4 Do the water sample coffee-ring effect patterns have significant statistical correlation with element composition

The heat-map correlations between the coffee-ring effect residue particles' area, eccentricity, and the percentage of elements such as sulfur, chlorine, carbon, sodium, magnesium, and calcium are shown in Table 6.23. The strongest correlations between residue particle features and element percentage are observed under conditions A, G, and H.

Under condition G, the correlation between calcium and magnesium is -0.0093, indicating that these two elements in the residue are well separated in the residue pattern. Conversely, under condition B, the correlation between calcium and magnesium is 0.0087, suggesting that these two elements present in similar positions in the residues are not well separated.

Another important phenomenon observed under conditions A, G, and H is that the correlation between particle area feature and elements is higher than other conditions. For instance, the correlation between particle area and sulfur percentage is 0.01, which is higher than condition B (0.0045) and condition D (0.0057). Additionally, the correlation between area and chlorine is 0.027, which is the highest correlation among these nine conditions.

Overall, these results suggest that conditions A, G, and H are more effective in separating the elemental components in the coffee-ring effect residue pattern and producing a higher correlation between the particle features and element compositions.

Table 3.11: Optimal condition analysis for consistent replicates residue pattern and distinct water samples particle features.

Conditions analysis vs	A	B	C	D	E	F	G	H	I
Temperature (°C)	20-23	20-23	20-23	23-26	23-26	23-26	26-29	26-29	26-29
Relative humidity (%)	35-40	40-45	45-50	35-40	40-45	45-50	35-40	40-45	45-50
PERMANOVA on CRE pattern features			✓			✓		✓	✓
MANOVA on CRE pattern area	✓	✓		✓		✓			
MANOVA on CRE pattern perimeter		✓							
MANOVA on CRE pattern eccentricity	✓	✓	✓		✓			✓	
MANOVA on CRE pattern centroid	✓	✓		✓	✓	✓		✓	✓
CRE pattern features ANOSIM			✓		✓	✓	✓	✓	
CRE area ANOSIM			✓			✓		✓	
CRE perimeter ANOSIM		✓	✓	✓		✓			
CRE centroid ANOSIM	✓			✓			✓		
CRE eccentricity ANOSIM			✓	✓		✓	✓		
EDS elements ANOVA	✓	✓	✓	✓	✓	✓	✓	✓	
Particles EDS ANOVA	✓						✓	✓	
Summary-number of optimal results out of 12 analyses	6	6	7	6	4	8	5	7	2

Based on the previous analysis presented in Table 3.11, the optimal environmental conditions for separating the elemental components in the coffee-ring effect residue pattern are 23-26°C with 45-50% relative humidity, 20-23°C with 45-50% relative humidity, and 26-29°C with 40-45% relative humidity. Nonetheless, the optimal environmental condition is a temperature range of 23-26°C and a relative humidity of 45-50%, as it yielded the

highest number of optimal results in 12 separate analyses. These conditions produced the highest correlation between particle features and element compositions, indicating that the particles and elements were well separated in the residue pattern. These optimal environmental conditions can be useful for developing models to detect and quantify contaminants in water samples using the coffee-ring effect residue pattern analysis.

3.5 Conclusion

The study demonstrates the potential of the coffee-ring effect as a tool for tap water analysis. It shows that the coffee-ring effect can produce unique fingerprints for water samples with different recipes and environmental conditions. The results also confirm the reproducibility of the coffee-ring effect, which is essential for establishing it as a reliable analytical technique. Additionally, the study highlights that both environmental conditions and water compositions impact the residue patterns produced by the coffee-ring effect, and that these patterns reflect the water chemistry within the sample. This study demonstrated the effectiveness of the auto temperature humidity control chamber in maintaining temperature and relative humidity, as well as the four-axis autosampler for conducting nanochromatography experiments.

The study concludes that statistical methods such as ANOVA, MANOVA, and PERMANOVA can differentiate coffee-ring effect residue patterns with respect to environmental conditions and water sample compositions. However, the results from different analysis methods are inconsistent, so further research is needed to determine the best method for differentiating these patterns. The research presents the findings of various statistical analyses conducted to investigate the coffee-ring effect residue patterns. These analyses included ANOVA and MANOVA tests on residue pattern features, such as area, perimeter, centroid, eccentricity, and orientation, ANOSIM tests on residue pattern

features and element distributions, and two-way ANOVA tests on element distributions.

The results of these analyses indicate that both environmental conditions and water chemistry significantly influence residue patterns and element distributions. In particular, certain conditions, such as 23-26°C with 45-50% relative humidity, 20-23°C with 45-50% relative humidity, and 26-29°C with 40-45% relative humidity, are well-suited for differentiating between water samples with varying concentrations of different components. Nonetheless, the optimal environmental condition is a temperature range of 23-26°C and a relative humidity of 45-50%, as it yielded the highest number of optimal results in 12 separate analyses. It is important to note that these findings have implications for the study of residue patterns and the understanding of the coffee-ring effect. Specifically, they suggest that further research is needed to better understand how environmental factors and water chemistry work together to impact residue patterns.

CHAPTER 4

CNN-Vision-transformer model for elements concentration

estimation by coffee-ring effect residue patterns

4.1 Abstract

This study investigates the effectiveness of the machine learning technique in detecting multiple contaminants in a tap water's dried residue's coffee-ring effect "fingerprint" with high accuracy. The use of the coffee-ring effect on water droplets dried on low-cost aluminum substrates allows low-cost separation of solutes within water samples, forming unique "fingerprints" for each tap water that can be photographed and analyzed using machine learning. Three models were evaluated in this research: the One-stage point estimation model (OnePeM), the Two-stage vision-transformer point estimation model (TwoVtPeM), and the Two-stage vision-transformer multiple output estimation model (TwoVtMoM). The TwoVtPeM technique achieved the best performance of the models tested (OnePeM, TwoVtPeM and TwoVtMoM), with OnePeM also performing well and TwoVtMoM falling short. The TwoVtPeM relative percentage errors were $\pm 17.1\%$ for oxygen, $\pm 4.5\%$ for sulfur, $\pm 19.9\%$ for sodium, $\pm 5.7\%$ for chlorine, $\pm 19.8\%$ for calcium, $\pm 25.8\%$ for magnesium, and $\pm 20.1\%$ for carbon. The R^2 was 0.95 which is higher than OnePeM with 0.90 R^2 and TwoVtMoM which was 0.54. The TwoVtPeM had a higher error mean than OnePeM, but it exhibited lower relative standard deviations of estimation; the TwoVtPeM relative standard deviations values were: 3.9% for oxygen, 3.0% for sulfur, 5.3% for sodium, 3.9% for magnesium, 5.3% for chlorine, 10.0% for calcium, and 5.9% for carbon. Moreover, 79.2% of water samples were correctly classified for hardness based on the estimated element concentrations by TwoVtPeM. The OnePeM

model correctly classified 67.2% of water samples, however the TwoVtMoM model achieved only 60.2% accuracy rate in classifying water samples for hardness.

The study's findings reveal the advantages of the deep learning technique (TwoVtPeM) potential for water analysis over other screening methods such as test strip test kits, due to its ability to estimate multiple contaminants simultaneously, speed and low cost. Further improvements can be made, including addressing certain limitations such as the quality of the substrate and the size and complexity of the dataset and models. Advances in camera technology and deep learning techniques have the potential to improve the method's ability to detect low concentrations of elements. In conclusion, this study highlights the potential of machine learning to transform water quality monitoring, leading to better health outcomes for individuals and communities.

4.2 Introduction

Ensuring sustainable and clean access to water is crucial for water and wastewater treatment plants as well as other natural and industrial systems that depend on this vital resource. These plants not only have to meet the needs of consumers and upgrade infrastructure to improve their quality of life, but they also face increasingly stringent regulatory measures to meet rising quality standards Faherty [2021]. Unfortunately, heavily polluted waterways are becoming more common in many countries, posing a threat to human, aquatic, and terrestrial life Ebenstein [2012]. To address these challenges, researchers worldwide are exploring methods to optimize, remediate, and enhance water usage Lages Barbosa et al. [2015], Yang et al. [2020], Vu and Wu [2022], Podder et al. [2021]. Many are focusing on creating and simulating optimized, cost-effective, and intelligent models to tackle these issues. Artificial intelligence (AI) has become an important tool in this effort, enabling the analysis and interpretation of vast amounts of

data to facilitate better decision-making and more effective management of water resources.

The water industry is increasingly turning to emerging AI and ML technologies, as well as smart systems, to address challenges that have traditionally been underserved by conventional methods and approaches. These technologies are anticipated to offer cost savings and process optimization through their resilience, generalization, and ease of design, helping to model and overcome complex water-related issues Alam et al. [2022], Taoufik et al. [2022], Gordanshekan et al. [2023], Xie et al. [2022]. Applications that have already benefited from ML include water and wastewater treatment, natural-systems monitoring, and precision agriculture. The most commonly used ML techniques in these studies include artificial neural networks (ANNs), recurrent neural networks (RNNs), random forest (RF), support vector machine (SVM), and adaptive-neuro fuzzy inference systems (ANFISs), with occasional use of AI techniques such as fuzzy inference systems (FISs). Some studies have also explored hybrid approaches, such as ANN-RF and SVM-RF, with positive outcomes in water-related modeling processes.

4.2.1 Coffee-ring effect residue provides particles structure information

The coffee-ring effect creates unique residue patterns or fingerprints correlating to tap water chemistry when harnessed Li et al. [2020], Shahidzadeh-Bonn et al. [2008], Kaya et al. [2010], Shin et al. [2014], Shahidzadeh et al. [2015]. These patterns result from the crystallization process of water contaminants and are influenced by various factors, such as evaporation, bulk flow, temperature, humidity, and wettability Li et al. [2020], Qazi et al. [2017], Wei et al. [2012], Sammalkorpi et al. [2009], Desarnaud et al. [2014], Meldrum and O'Shaughnessy [2020]. Crystallization of salts from drying saline droplets has been investigated in some studies, which analyzed nucleation mechanisms and the dependence

of precipitation profile on factors like surface properties and salt concentration. The complexity of the coffee-ring effect pattern formation is influenced by contact line pinning on the substrate and the contact angle. Previous research has found particle separation during coffee-ring formation to be based on a particle-size selection mechanism near the contact line of an evaporating droplet, leading to nanochromatography of various biological entities with high separation resolution and dynamic range Wong et al. [2011], Larson [2014], Deegan et al. [1997], Chen and Evans [2010], Eral et al. [2013]. This mechanism has the potential to be used to estimate crystal structures and even particle concentrations.

4.2.2 Applications of AI and ML methods in Water Treatment

ML techniques for modeling membrane-filtration processes aim to output several variables, such as transmembrane pressure, permeate flux, and solute rejection. Inputs in published studies include pH, temperature, contact/filtration time, transmembrane pressure, and flux rate, among others. ANN, RNN, and SVM models consistently performed well, achieving R^2 values greater than 0.9 and often greater than 0.99. AI and ML methods have also been used to control chlorination, estimate disinfection by-product (DBP) concentration, model significant parameters for adsorption and membrane-filtration processes. Statistical measures used to evaluate results include the coefficient of correlation (R), coefficient of determination (R^2), mean average error (MAE), mean square error (MSE), root mean square error (RMSE), and relative error (RE).

Chlorination and Disinfection By-Product Estimation

Disinfecting water is crucial for killing or inactivating microorganisms and viruses. Chlorine-based disinfectants Li et al. [2017], Xu et al. [2015, 2013] are often used, but they pose health hazards and can create DBPs Sedlak and von Gunten [2011], Bull et al. [1995]. AI methods can be used to control chlorination, while ML technologies can predict and

mitigate DBP formation. Studies have tested models on surface waters treated with chlorine and noted success in modeling DBP concentrations in treated water distribution networks and at consumer taps Librantz et al. [2018], Godo-Pla et al. [2021], Singh and Gupta [2012], Mahato and Gupta [2022], Park et al. [2018], Lin et al. [2020], Xu et al. [2022], Peleato [2022], Okoji et al. [2022], Cordero et al. [2021].

Adsorption Processes

Adsorption processes remove various contaminants in the water and wastewater treatment industries. Predictive models using ML can optimize the adsorption process and extend the media's life, increasing the plant's effectiveness and confidence in meeting applicable regulations. Studies have modeled adsorption processes with water streams contaminated with metals, industrial dyes, and organic compounds using various adsorbent media including carbonaceous materials and metal-based nanocomposites Bhagat et al. [2021], Mazloom et al. [2020], Mesellem et al. [2021a], Al-Yaari et al. [2022], Mazaheri et al. [2017], Ahmad et al. [2020], Fawzy et al. [2016], Ullah et al. [2020], Mahmoud et al. [2019], Mesellem et al. [2021b].

Membrane-Filtration Processes

Membrane processes separate contaminants in water and wastewater treatment by passing the water through a barrier or filter using high-pressure differentials. These processes are typically used for contaminants that are difficult or costly to remove by chemical or physical means Hube et al. [2020], Pronk et al. [2019]. AI and ML models have been used to treat various water sources contaminated with pollutants and natural compounds Zoubeik et al. [2019], Fetanat et al. [2021], Khan et al. [2022], Yusof et al. [2020], Nazif et al. [2020], Shim et al. [2021], Ammi et al. [2021a]. ANN is the most commonly used model, although ANFIS, SVM, and specific forms of ANNs have also been used for membrane-

filtration-process modeling. ANN, RNN, and SVM models consistently performed well, achieving R^2 values greater than 0.9 and often greater than 0.99 Zoubeik et al. [2019], Khan et al. [2022], Yangali-Quintanilla et al. [2009].

Vision Transformer in computer vision

Deep neural networks (DNNs) are the backbone of AI systems today. Different types of networks are suited for different tasks. For instance, the multi-layer perceptron (MLP) or fully connected (FC) network, made up of multiple linear layers and nonlinear activations, is a classical type of neural network Rosenblatt [1957]. Convolutional neural networks (CNNs) use convolutional and pooling layers to process shift-invariant data like images LeCun et al. [1998], Krizhevsky et al. [2017]. Recurrent neural networks (RNNs) apply recurrent cells to handle sequential or time series data Hochreiter and Schmidhuber [1997]. Transformer is a novel neural network that uses self-attention mechanisms Bahdanau et al. [2014], Parikh et al. [2016] to extract intrinsic features Vaswani et al. [2017] and demonstrates great potential for broad AI applications. It was first used in NLP tasks, where it showed significant improvement Vaswani et al. [2017], Devlin et al. [2018], Brown et al. [2020]. For instance, Vaswani et al. Vaswani et al. [2017] first proposed the transformer, which is based on the attention mechanism, for machine translation and English constituency parsing tasks. Devlin et al. Vaswani et al. [2017] introduced BERT (Bidirectional Encoder Representations from Transformers), a language representation model that pre-trains the transformer on unlabeled text, considering the context of each word in a bidirectional manner. BERT obtained state-of-the-art results on 11 NLP tasks upon publication. Brown et al. Brown et al. [2020] pre-trained a massive transformer-based model, GPT-3 (Generative Pre-trained Transformer 3), using 45 TB of compressed plaintext data and 175 billion parameters. It performed well on various downstream NLP

tasks without the need for fine-tuning. These transformer-based models, with their robust representation capacity, have brought about significant advances in NLP.

Recently, the success of transformer architectures in NLP has inspired researchers to apply it to computer vision tasks. Although CNNs have been traditionally considered the foundation of vision applications He et al. [2016], Ren et al. [2015], the transformer is proving to be a potential alternative. Chen et al. [2020] trained a sequence transformer to predict pixels through auto-regression, achieving comparable results to CNNs in image classification tasks. The vision transformer model, ViT, was proposed by Dosovitskiy et al. [2020], which directly applies a pure transformer to sequences of image patches for full image classification and has achieved state-of-the-art results on multiple image recognition benchmarks. Transformer has also been used to solve various other vision problems, such as object detection Carion et al. [2020], Zhu et al. [2020], semantic segmentation Zheng et al. [2021], image processing Chen et al. [2021], and video understanding Zhou et al. [2018]. Its exceptional performance is attracting more researchers to propose transformer-based models for a wide range of visual tasks.

However, there has not yet been research conducted using the coffee-ring effect in conjunction with machine learning and deep learning models, particularly the vision transformer model, to estimate the concentration of elements in water samples. The vision transformer model has the potential to not only utilize the particle morphology and location information from one element to make estimations but also incorporate the physical chemistry interactions between elements to correct noise and increase accuracy. This approach could offer a novel method for screening water quality and even understanding the underlying interactions between various elements within them. Another contribution of the study is the use of SEM-EDS images as training data to build the

model. This approach allows for the extraction of much more detailed information regarding crystal structure. Additionally, EDS images serve as guidance for the model to estimate the locations of deposited elements, which helps reduce estimation errors and increase the coefficient of determination from 0.90 to 0.95. This innovative method provides improved accuracy and insights into the complex relationships between elements within the samples.

4.2.3 Model for elements recognition and concentration estimation

The proposed components estimation model is a two-stage deep learning approach for determining the elements concentrations in water samples using the coffee-ring effect. The coffee-ring effect is a phenomenon in which a ring-shaped deposit of coffee particles is formed around the perimeter of a droplet of coffee on a substrate. The effect is caused by the combination of coffee particles' Brownian motion and evaporation, which causes the particles to be transported to the edge of the droplet. The coffee-ring effect is of interest in various fields such as materials science, physics, and biology, as it can be used to pattern surfaces and deposit particles in a controlled manner.

The first stage of the model utilizes a deep learning model to estimate the locations and abundances of seven elements (calcium, magnesium, sodium, sulfur, carbon, oxygen, and chlorine) in the sample, based on the crystal structure and location information present in images of the coffee-ring effect. The input to the model are the SEM images and EDS images of the coffee-ring effect, which are pre-processed to ensure they are of good quality and that the features of interest are clearly visible. The model uses a convolutional neural network (CNN) architecture to extract features from the images, as the information extracted from one element can be useful for understanding the presence and behavior of other elements, and the crystal deposition location plays a critical role in determining the

crystal composition.

The output of the first stage are seven binary images, each indicating the estimated location and abundance of a specific element. The binary output images are threshold images that have been processed to get a binary image where the pixels with signal corresponds to the location of the estimated element and the abundance of signal pixels indicates the abundance of the element in that area.

The second stage of the model utilizes a Vision-transformer deep learning model to estimate the concentrations of the elements in the sample, based on the locations and abundances estimated in the first stage. The model uses the outputs from the first stage as input, and considers the relationships between elements, such as the low solubility of calcium sulfate, to improve the accuracy of the concentration estimates. For example, the estimated concentration of sulfur can be used to refine the concentration estimation of calcium, and vice versa. This stage also uses a CNN architecture to extract features from the inputs and a regression model to estimate the concentrations.

Overall, this proposed model utilizes the latest machine learning techniques to study the coffee-ring effect and estimate the composition of elements in water samples. The two-stage approach, with the co-learning and attention technique, allows for more accurate estimation of the locations, abundances, and concentrations of the elements, and can provide new insights into the dynamics of the coffee-ring effect and aid in the development of new techniques for controlling the deposition of particles.

In this study, three models were built and evaluated on water samples that have been prepared using scanning electron microscopy (SEM) images. The end-to-end model is a single-stage model designed to estimate the concentrations of elements in the water samples. The input to the model is a three-layer SEM image and the output is the

estimated concentration of elements. The model consists of a Unet module with a ResNet50 encoder, ImageNet encoder weights, and a sigmoid activation function. This is followed by three convolutional layers, max pooling layers with ReLU activation, and a final linear layer that outputs the estimated element concentrations (refer to Figure 4.1). The Two-stage vision-transformer point estimation model is made up of two modules (stages). The first module is identical to the Unet structure in the end-to-end model, producing seven binary 2D outputs used to estimate the elements' concentrations. The second module is a vision-transformer module that extracts elements' location information to estimate their corresponding concentrations (refer to Figure 4.2). The third model, the Two-stage vision-transformer multiple output model, is similar in structure to the Two-stage vision-transformer point estimation model, but it produces a range of elements' concentrations (refer to Figure 4.3).

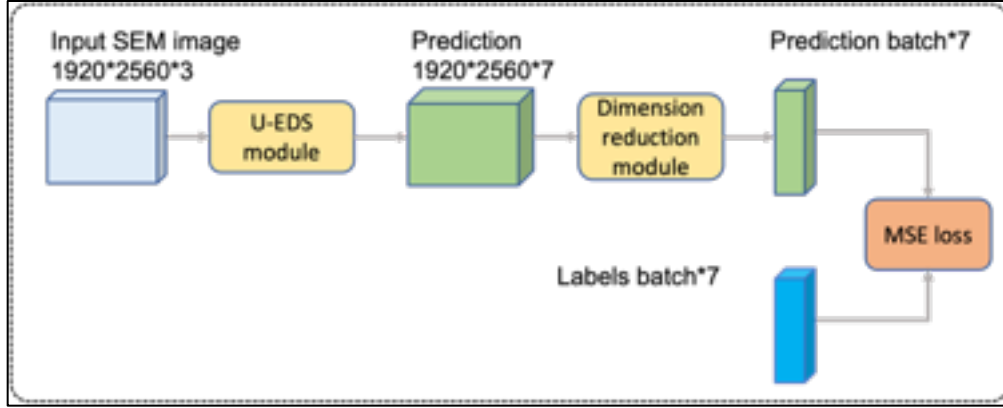


Figure 4.1: One-stage point estimation model structure.

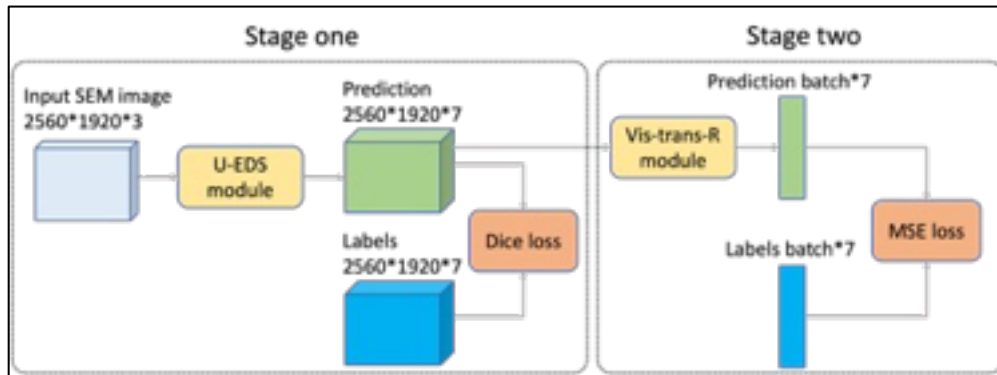


Figure 4.2: Two-stage vision-transformer point estimation model structure.

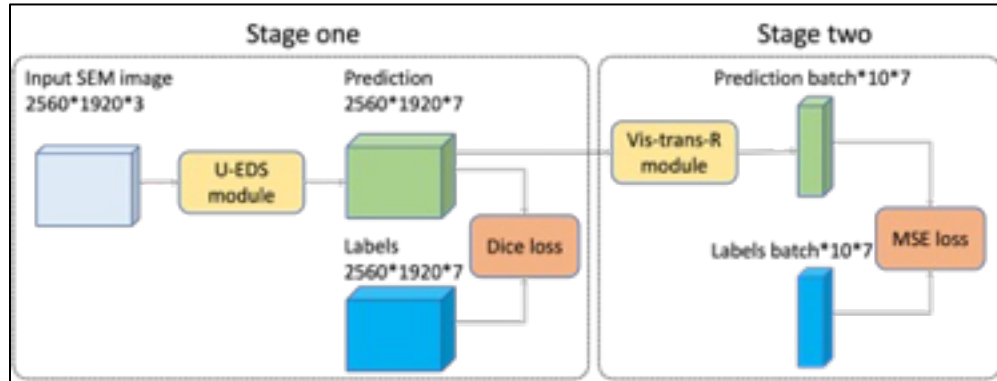


Figure 4.3: Two-stage vision-transformer point estimation model structure.

4.3 Experimental Methods

4.3.1 Develop a deep learning model to identify corrosion indicators and quantify their concentrations in tap water

A CNN model has been developed to identify corrosion indicators in tap water,

utilizing similar methods as those previously employed for assigning tap water SEM fingerprints to groups with similar water chemistry with an accuracy of $76.7 \pm 3.0\%$ Li et al. [2020]. Features of the previous model that are applicable to the new work include the convolutional layers, fully connected layers, and the Relu activation function Li et al. [2020]. Parameters of the model have been adjusted to fit this research, including the number of convolutional layers, the output layer, and the loss function (three-channel RGB images will be analyzed instead of black and white images). The output of this study consists of maps depicting expected elemental deposition and concentrations of each contaminant, in contrast to the previously published work where the output was a classification of the image into a group with similar water chemistry. Loss will be calculated for the proposed work using mean square error instead of the cross-entropy method used previously Li et al. [2020].

The experiment has been divided into three steps. In the first step, additional tap water SEM fingerprints have been collected and evaluated for synthetic Detroit water samples under the optimal environmental condition (23-26 (°C), 45-50% relative humidity) obtained from a chapter 3. In the second step, a deep learning model has been developed using tap water SEM fingerprints (SEM images) and SEM-EDS map images to assign elements to the crystals that formed. Finally, in the last step, three vision-transformer models have been constructed to utilize the predicted element depositions in order to estimate concentrations of each element.

Tap water fingerprints (SEM and photographs) collection for an array of synthetic waters

Water sample recipes were developed based on Detroit water reports from 2017 to 2019. Components were prepared in a broader range to accurately represent the variability of

water constituents. The recipe details can be found in the supplementary file. The SEM residue patterns and EDS mapping of contaminant particles in tap water samples have been collected from droplets of each sample with five replicates that were dried under optimal temperature and relative humidity conditions (23-26 (°C), 45-50% relative humidity). Photographs of each residue were captured with the Celestron camera, SEM images of whole droplets were taken, and EDS maps were obtained for sodium, calcium, magnesium, chlorine, carbon, sulfur, and oxygen using the same method as in previous research Li et al. [2020] section. 3.3. Water sample recipes were designed to mimic the range of tap water components Table. 6.28 Table. 6.29 Table. 6.30. Table. 6.31 Table. 6.32. The SEM image and EDS mapping of the same area are shown in Figure. 4.4.

Elements mapping estimation model for recognition of contaminants particles

Elements mapping estimation model has been built and trained based on the SEM and EDS mapping data collected in previous step. The model takes water SEM fingerprints as input and, however, EDS image maps of contaminants elements as output instead of classification of the image. To evaluate the model performance, the output images have been overlaid with the EDS map, the pixel positions in these two maps has been measured and accuracy has been calculated. The model was built with `segmentation_models_pytorch` package with `resnet34` encoder, seven output classes, sigmoid activation and model weights initialized with ImageNet weight initializer. Multilabel dice loss was applied in the training process. All the three models trained for 1000 epochs with 0.1 learning rate.

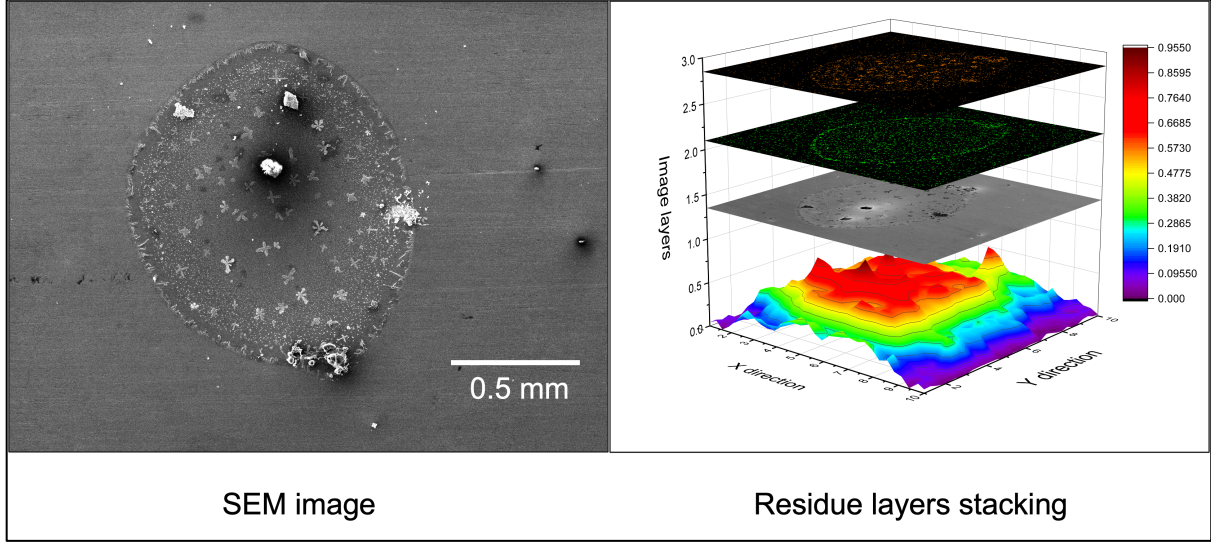


Figure 4.54: 3D stacking of residue surface scanning, SEM image, oxygen EDS, chlorine EDS bottom up.

Dice loss

In cross entropy loss, the overall loss was calculated as the average of per-pixel loss. However, the per-pixel loss was calculated discretely without considering whether its neighboring pixels are boundaries or not. As a result, cross entropy loss only takes into account the loss in a micro sense, rather than considering it globally, leading to limitations in image-level prediction. Dice loss Eqn. 4.1 originates from Sorensen-Dice coefficient, which is a statistic developed in 1940s to gauge the similarity between two samples. It was brought to computer vision community by Milletari et al. in 2016 for 3D medical image segmentation Milletari et al. [2016] which is widely used for image segmentation and boundary detection.

$$D = \frac{2 \sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \quad (4.1)$$

The equation for the Dice coefficient, shown in Eq. 4.1, calculates the similarity between the prediction and ground truth in boundary detection. The variables p_i and g_i represent corresponding pixel values, with a value of 1 indicating the presence of a

boundary and 0 indicating its absence. The denominator is the sum of total boundary pixels in both the prediction and ground truth, while the numerator is the sum of correctly predicted boundary pixels (i.e., those where p_i and g_i both have a value of 1).

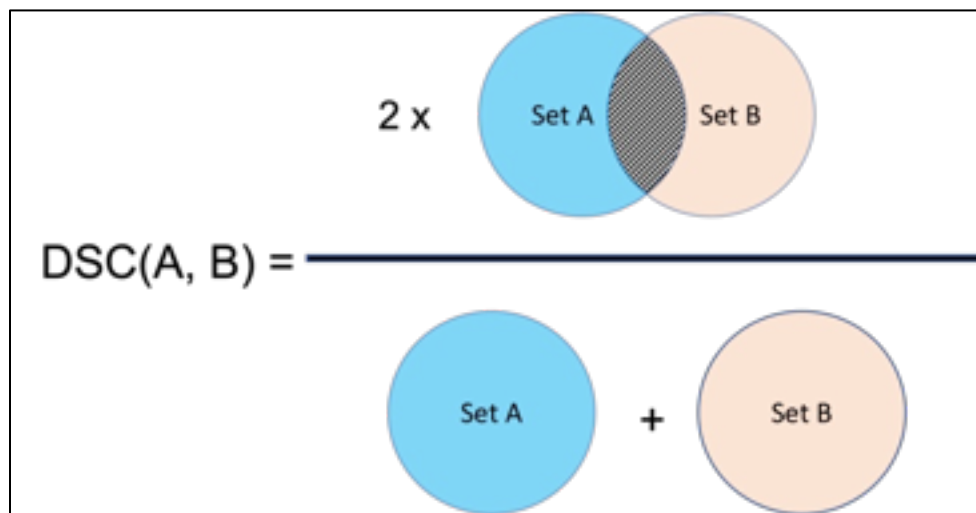


Figure 4.5: Dice coefficient (set view)

Persistent Homology of Point Clouds

In practice, the sliding window embedding of a video X is a finite set $SW_{d,\tau} X = SW_{d,\tau} X(t) : t \in T$, determined by a finite choice of $T \subset R$. As $SW_{d,\tau} X \subset R^{W \cdot H(d+1)}$, the ambient Euclidean distance equips $SW_{d,\tau} X$ with the structure of a finite metric space. Such discrete metric spaces, or point clouds, are topologically trivial, with N points having N connected components and no higher-dimensional features like holes. However, when a point cloud is sampled from or around a continuous space with non-trivial topology (e.g., a circle or torus), one would expect simplicial complexes built on the point cloud vertices to reflect the underlying continuous space's topology. Persistent homology is applied to discrete collections of points such as sliding window embeddings Zomorodian and Carlsson [2004].

For a point cloud (X, d_X) , where X is a finite set and $d_X : X * X \rightarrow [0, \infty)$ represents a distance function, the Vietoris-Rips complex (also known as Rips complex) at scale $\epsilon \geq 0$

consists of non-empty subsets of X with a diameter less than or equal to ϵ :

$$R_\epsilon(X) := \sigma \subset X : d_X(x_1, x_2) \leq \epsilon, \forall x_i, x_j \in \sigma \quad (4.2)$$

The $R_\epsilon(X)$ is a simplicial complex with its vertex set equivalent to X . It is formed by adding an edge between any pair of vertices with a distance of at most ϵ , incorporating all 2-dimensional triangular faces (i.e., 2-simplices) with existing bounding edges, and, more generally, including all k -simplices with included $(k-1)$ -dimensional bounding facets. Figure 4.6 illustrates the evolution of the Rips complex for a set of points sampled around the unit circle.

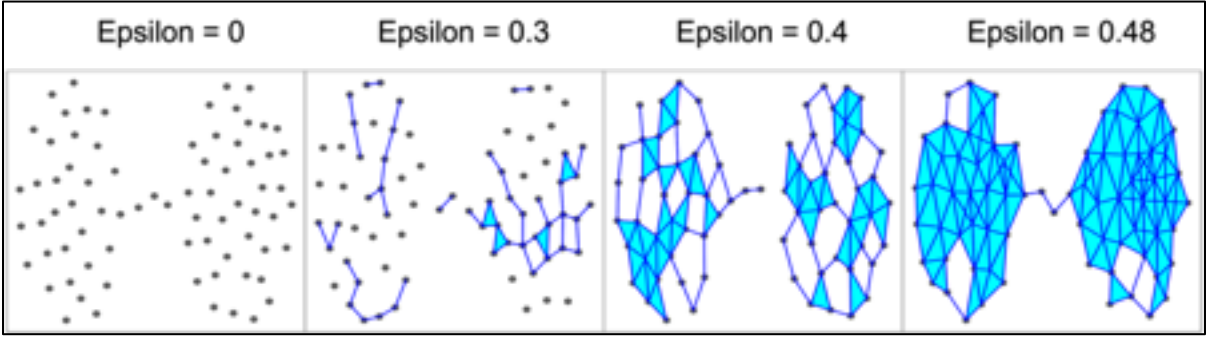


Figure 4.6: The Rips complex, at three different scales ($\epsilon = 0, 0.30, 0.40, 0.48$), on a point cloud with 40 points sampled around $S^1 \subset R^2$.

For an open cover given by $\{B_\alpha(l_j)\}_{l_j \in L}$, where L is the landmark set and α is the radius of the balls, we have an associated partition of unity defined as

$$\phi_j(b) = \frac{|\alpha - d(b, l_j)|_+}{\sum_k |\alpha - d(b, l_k)|_+} \quad (4.3)$$

Persistent Homology of Point Cloud

In topological analysis, the nerve complex, or the nerve of a family of sets, is a concept used to represent the intersection patterns of these sets. Given a collection of sets, the nerve complex is an abstract simplicial complex where each set corresponds to a vertex, and a collection of vertices forms a simplex if and only if the intersection of the corresponding sets is nonempty. In other words, the nerve complex encodes how the sets in

a family overlap with each other. This concept is particularly useful in various applications, including topological data analysis, where it can help analyze the structure of complex data sets Dey et al. [2017], Carlsson [2020].

Let I be a set of indices and C be a family of sets $(U_i)_{i \in I}$. The nerve of C is a set of finite subsets of the index set I Geoghegan [2007]. It contains all finite subsets $J \subseteq I$ such that the intersection of the U_i whose subindices are in J is non-empty Eqn. 4.4.

$$N(C) = \{J \subseteq I: \bigcap_{i \in J} U_i \neq \emptyset, J \text{ finite set}\} \quad (4.4)$$

Build vision-transformer model to use elements locations to estimate concentrations of contaminants elements

The Elements mapping estimation model from previous step has been trained to recognize element particles to output elements mappings. By utilizing the estimated contaminants particles EDS mapping images, these particles concentrations has been quantified. In this study, the vision-transformer model has been built and trained on EDS mapping predictions to estimate contaminants concentrations. The vision-transformer composed by two Multi Head Attention module with FeedForward module, Norm module, Positional layer, Encoder module, Decoder module and Feature Extraction module.

One-stage point estimation model, two-stage vision-transformer point estimation model and two-stage vision-transformer multiple output model comparison

To measure the model performance, One-stage point estimation model (OnePeM) was built to estimate the concentrations of elements in the water samples. This model consists of two modules. The first module is identical to the Unet structure of elements mapping estimation model and the second module is to use the 2D layers to estimate the elements concentrations. Different from the two-stage vision-transformer point estimation model (TwoVtPeM), in this model, the elements EDS mapping weren't used to train the first

module of the module, however, the model was trained end-to-end to estimate the elements concentrations. To produce robust concentration estimations, a two-stage vision-transformer multiple output model (TwoVtMoM) was built to produce multiple elements estimations. These two models have the same model backbone and the same activation function, weight initializer as the two-stage vision-transformer point estimation model.

To train the model stochastic gradient descent (SGD) with 0.001 learning rate and MSE loss were used to optimize the model parameters for 500 epochs. To accelerate the training speed, the model is trained by distributed data parallel (DDP) module with eight A100 (80GB SXM4) GPUs. The training time is about 10 hours.

Model training

Three of the five replicates of each image collected in task were randomly assigned to the training dataset and the remaining two replicates were assigned to the testing dataset. All models were trained on the training set and model performance was tested on the testing set. The accuracy of the particle recognition was computed by comparing two features of the element SEM-EDS mapping image and CNN model output: 1) whether or not a pixel occurs in the same location, and 2) the size of pixel clusters. Specifically, the pixel occurrence was evaluated by first overlaying the CNN output map onto the EDS map for contaminants particles. Both the EDS images and the CNN output are maps where each pixel was assigned either a value of 0 or 1.

In the evaluation stage, the CNN model output were analyzed to determine whether or not a pixel value of 1 exists in the same position or in a circle with a radius of 3 pixels drawn around the corresponding location on the EDS map. The pixel will be labeled as correctly identified if there exists at least one pixel for indicating the contaminants particles in the EDS map or labeled as incorrectly if not. The model accuracy, percentage of the

pixels that matched the EDS output for each image were calculated. Stochastic gradient descent (SGD) with 0.001 learning rate and MSE loss were used to optimize the model parameters and training were conducted for 500 epochs. To accelerate the training speed, the models were trained by distributed data parallel (DDP) module with eight A100 (80GB SXM4) GPUs.

4.4 Results and Discussion

4.4.1 Elements correlations between coffee-ring effect subrings

To investigate the correlations between elements in each coffee-ring effect residue subring, the droplet residue were separated into fifteen subrings with the evenly 4.9. The elements correlations between coffee-ring effect residue subrings were analyzed by Pearson correlation coefficient. The Pearson correlation coefficient is a measure of the linear correlation between two variables. It is a dimensionless number between -1 and 1, where 1 is total positive linear correlation, 0 is no linear correlation, and -1 is total negative linear correlation. The Pearson correlation coefficient is calculated by Eqn. Eqn. 4.5.

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4.5)$$

The strongest correlation was observed between sodium and chlorine, particularly within the second subring of both elements. This phenomenon suggests that sodium and chloride ions tend to form crystals in the second subring area. The highest correlations among oxygen, calcium, and sulfur were found in the outermost subring, indicating the formation of calcium sulfate ($CaSO_4$) in this region Figure. 4.7. Meanwhile, the highest correlations between chlorine and calcium occurred in the middle subring areas, signifying the formation of calcium chloride ($CaCl_2$) in those regions Figure. 4.8.

4.4.2 Elements mapping estimation model analysis

The estimated calcium-carbon and oxygen sulfur EDS mappings are displayed in separate 2D histograms in Figure 4.10. As observed, oxygen and sulfur are more prominently present in the droplet residue pattern area, while calcium is distributed throughout the entire image, although it is primarily located in the residue pattern. This is likely due to the presence of calcium in the substrate during the manufacturing process. To overcome this issue, a higher-quality substrate with a lower calcium content could be utilized. From the histogram results, discerning the correlation between calcium and carbon proves to be difficult. However, the relationship between calcium and sulfur is more apparent. SEM example image is shown in Figure. 4.7.

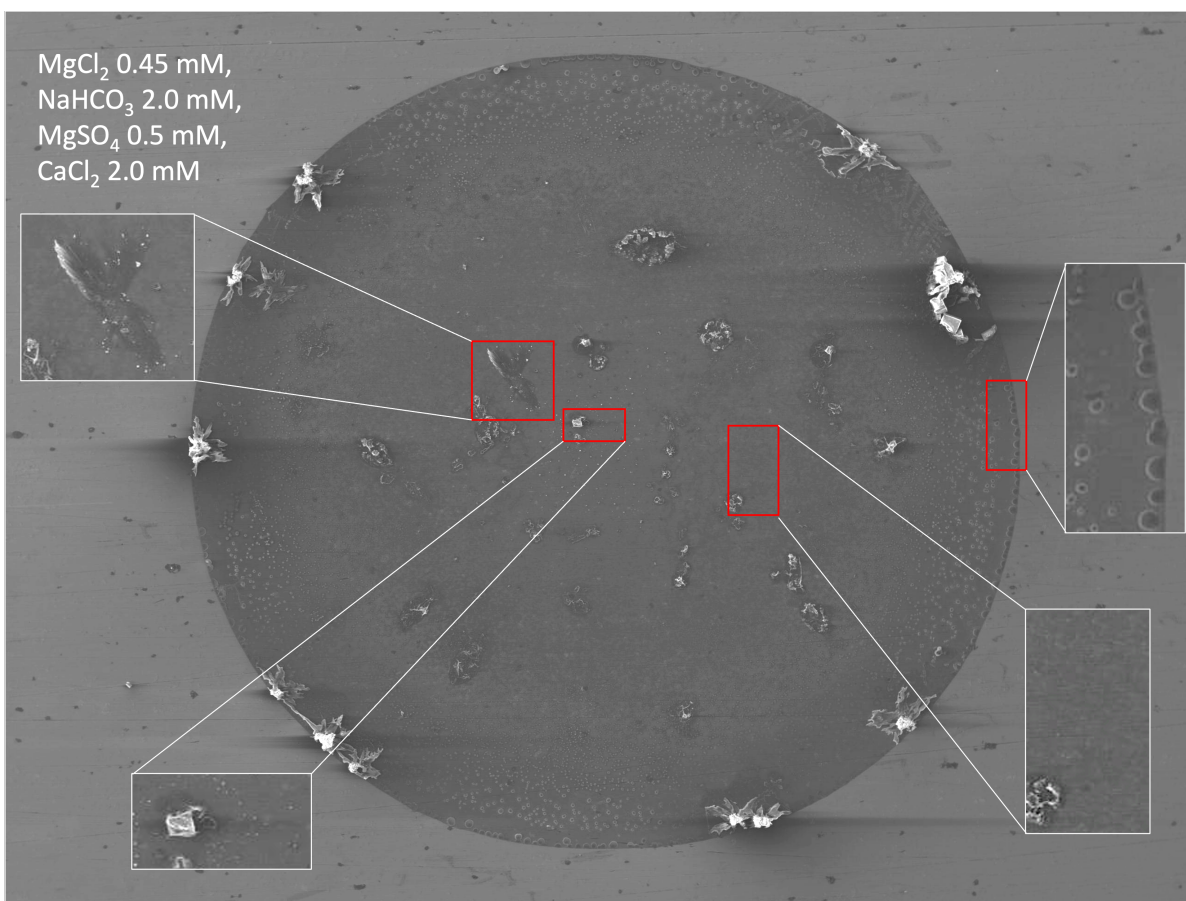


Figure 4.7: SEM image of water sample coffee-ring effect residue pattern (with detailed subregion presentation). Water sample with $MgCl_2$ 0.45 mM, $NaHCO_3$ 1.0 mM, $CaCl_2$ 1.5 mM, $MgSO_4$ 0.5 mM.

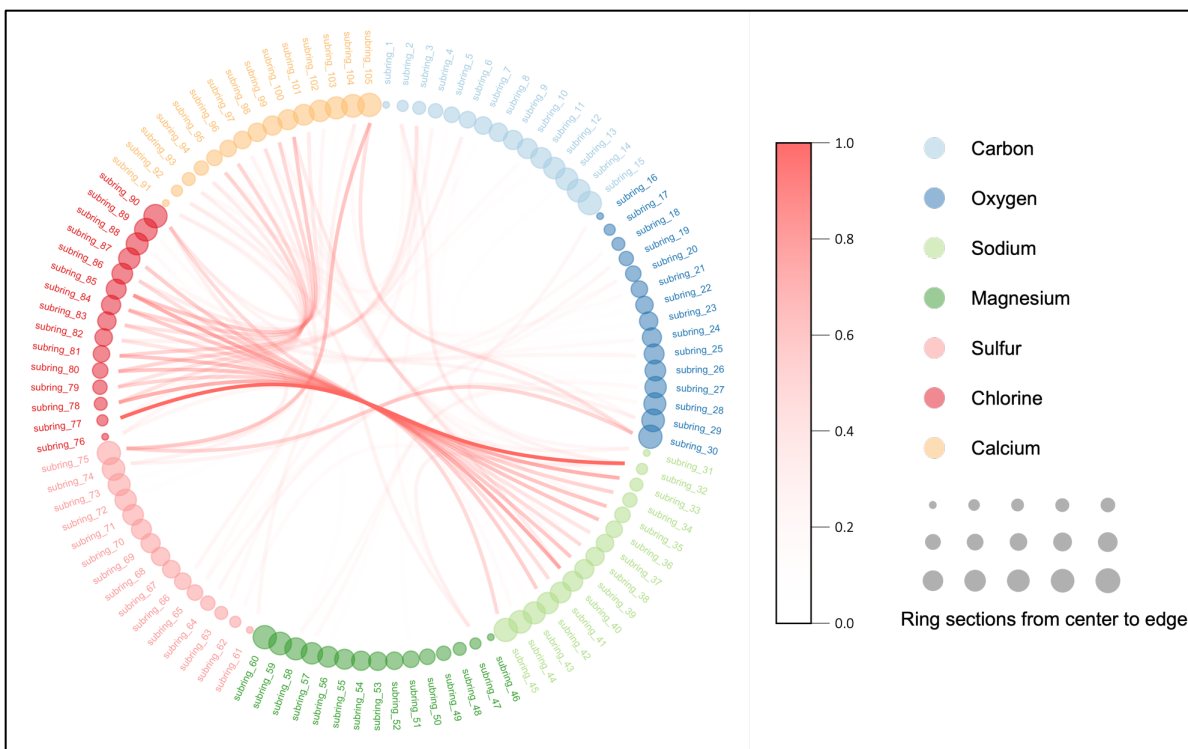
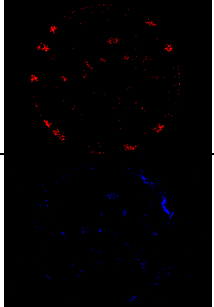
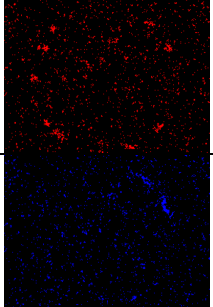
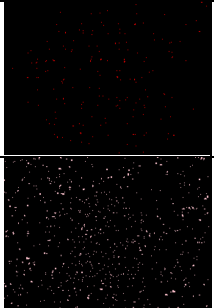
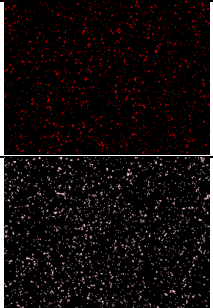
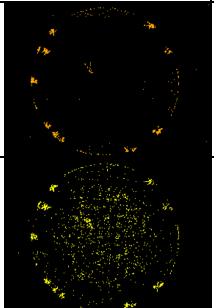
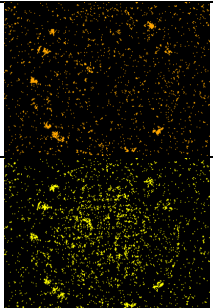
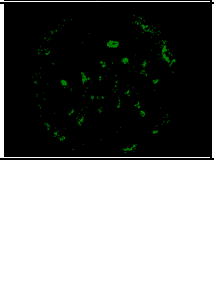
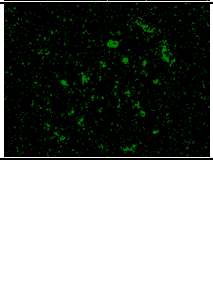


Figure 4.8: Pearson correlation of water contaminants in Coffee-ring effect residue subrings. Water sample with $MgCl_2$ 0.45 mM, $NaHCO_3$ 1.0 mM, $CaCl_2$ 1.5 mM, $MgSO_4$ 0.5 mM.

Table 4.1: Coffee ring effect elements deposition prediction by Unet model. Water sample with $MgCl_2$ 0.45 mM, $NaHCO_3$ 1.0 mM, $CaCl_2$ 1.5 mM, $MgSO_4$ 0.5 mM

Elements	Predicted EDS mapping	Target EDS mapping
Calcium		
Sodium		
Carbon		
Magnesium		
Oxygen		
Sulfur		
Chlorine		

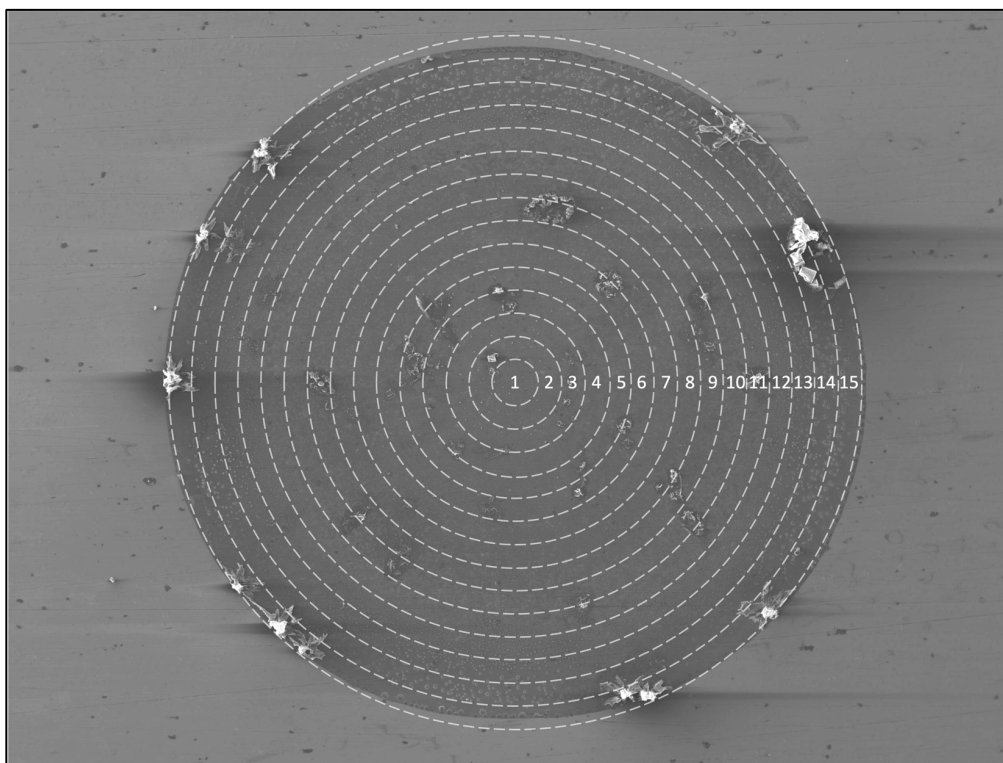


Figure 4.9: Coffee-ring effect residue pattern were separated to fifteen subrings with the evenly. Water sample with $MgCl_2$ 0.45 mM, $NaHCO_3$ 1.0 mM, $CaCl_2$ 1.5 mM, $MgSO_4$ 0.5 mM.

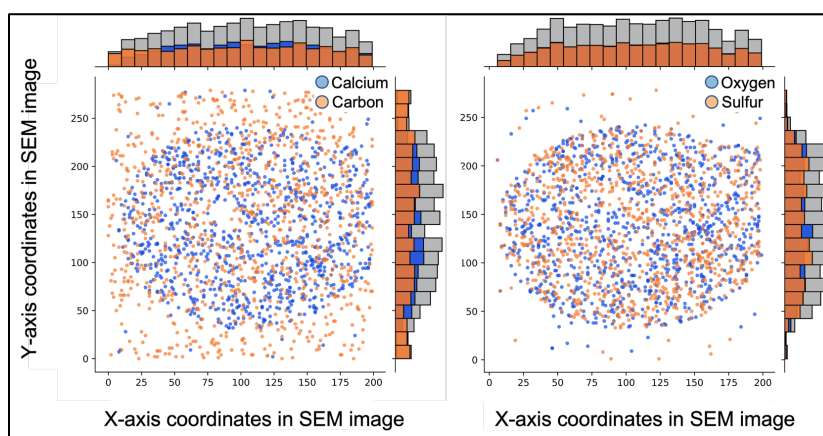


Figure 4.10: Topological nerve complex of estimated Calcium-Carbon EDS and Oxygen-Sulfur EDS. The left is Calcium-Carbon 2D histogram and the right is Oxygen-Sulfur 2D histogram. x-axis and y-axis are scaled. Water sample with $MgCl_2$ 0.45 mM, $NaHCO_3$ 1.0 mM, $CaCl_2$ 1.5 mM, $MgSO_4$ 0.5 mM.

The calculation of the nerve complex was based on the combination of calcium and carbon EDS mappings and the combination of calcium and sulfur EDS mappings. In both combinations, the two EDS mappings were compared, with one serving as the reference

and the other as the target. If a predicted element pixel was found in the reference, the location of the corresponding pixel in the target was recorded as a positive signal if it was present within a 3x3 area. To minimize noise, 1000 randomly selected points from the resulting pixels were used. This method resulted in the creation of the calcium-carbon and calcium-sulfur combination mappings with radius 0.008, which were then used to calculate the nerve complex. The nerve complex of the calcium-carbon and calcium-sulfur combination mappings are shown in Figure 4.11. The calcium-sulfur nerve complex formed at different locations of than calcium-calcium which was consistent the teh claim that different composition particles formed at different locations in the droplet residue pattern.

The results of the Unet elements deposition estimation are presented in Figure 4.12. The three tables, from left to right, represent accuracy, false positive, and false negative (sensitivity). The y-axis of each table represents 625 water samples, while the x-axis lists the elements in the order of Oxygen, Sulfur, Sodium, Magnesium, Chlorine, Calcium, and Carbon. The two-stage vision-transformer point estimation model, the one-stage point estimation model, and the two-stage vision-transformer multiple output model all include this module and were trained independently.

As shown in the accuracy results, sulfur and magnesium have the highest overall accuracy, while calcium and carbon have the lowest accuracy. This is also evident in Figure 4.17 where the predicted calcium values are mostly lower than the true values. The high accuracy of sulfur and magnesium can be attributed to the more accurate collection of sulfur and magnesium EDS mappings, compared to the high noise present in the calcium EDS mapping (as seen in Figure 4.13), as the EDS instrument is more sensitive to these two elements. Additionally, the substrate contains fewer sulfur and magnesium impurities, and these elements are more separated from other elements such as

oxygen and are prone to form crystals, such as SO_4^{2-} ions. The false positive and false negative values for magnesium are also lower than for other elements.

However, the EDS detector is not as sensitive to carbon, and the substrate contains a high concentration of calcium, leading to an inaccurate collection of EDS mappings for carbon. As a result, the model has difficulty learning the relationship between crystal structure and elements composition for carbon.

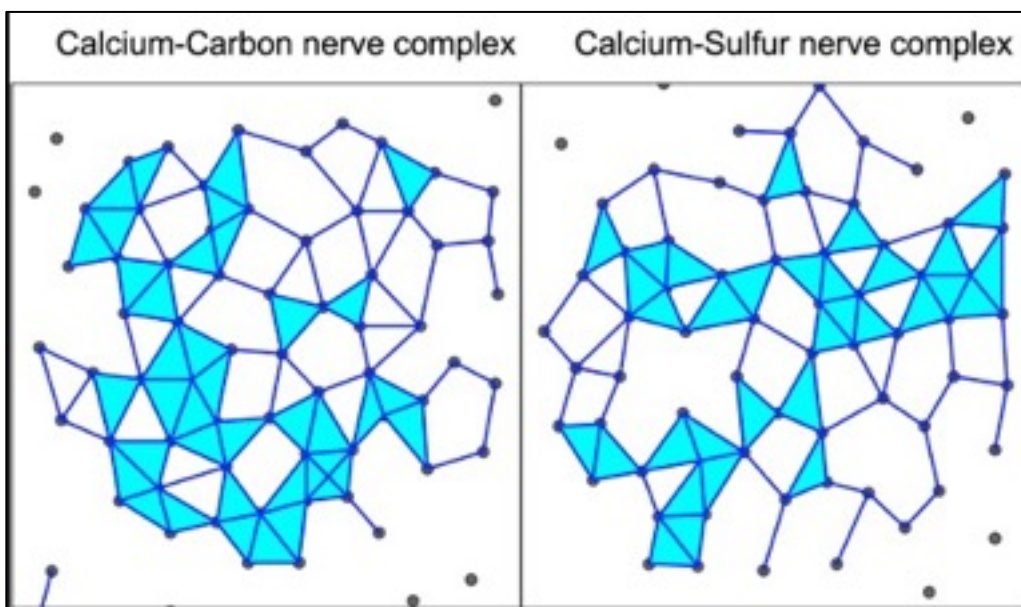


Figure 4.11: Topological nerve complex of estimated Calcium-Carbon EDS and Calcium-Sulfur EDS. The diagram on the left represents the Calcium-Carbon EDS nerve complex, while the one on the right shows the Oxygen-Sulfur nerve complex. A radius of 0.008 was used in the calculations. The coffee-ring effect residue pattern resulted in the formation of calcium carbon crystals ($CaCO_3$) and calcium sulfur crystals ($CaSO_4$) at different locations.

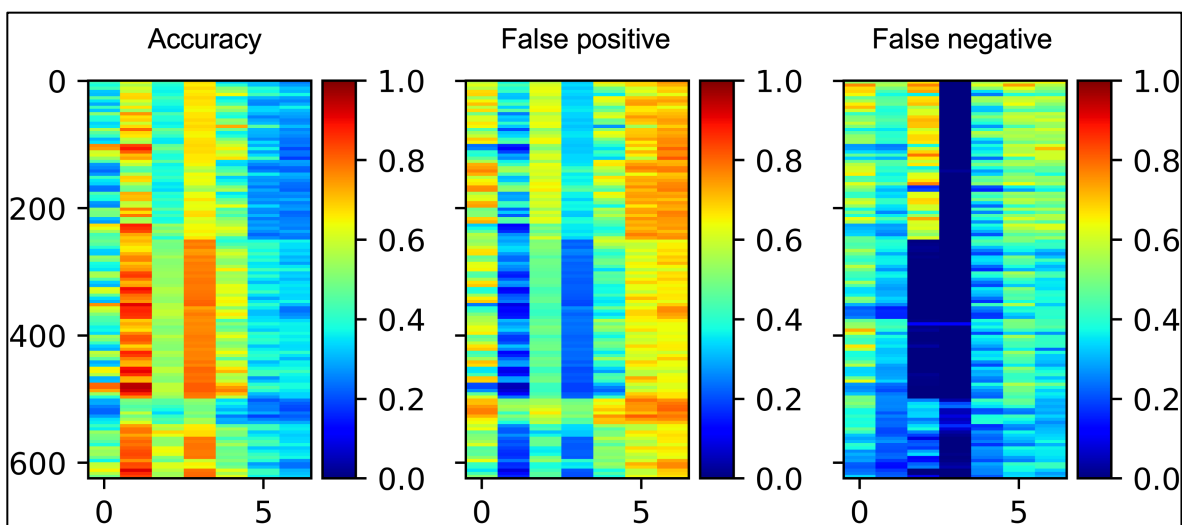


Figure 4.12: Accuracy, False Positive, False Negative (Sensitivity) tables from left to right; O, S, Na, Mg, Cl, Ca, C elements in each table from left to right. Result is averaged across five replicates.

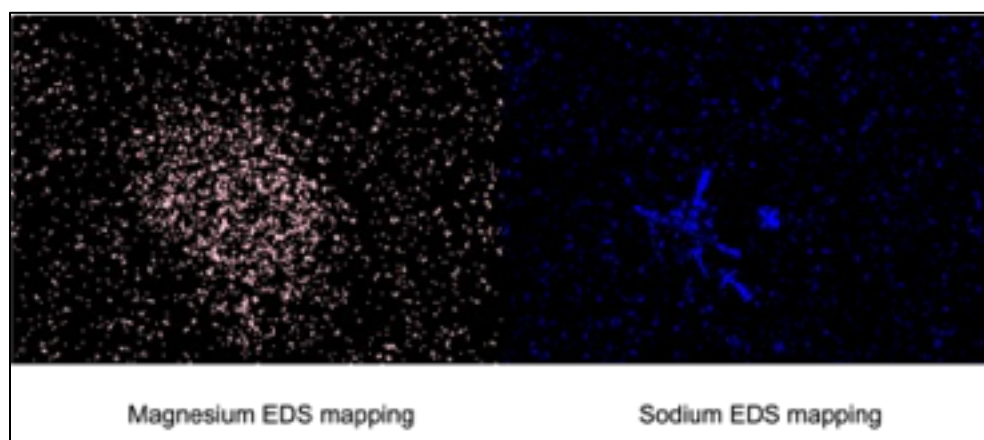


Figure 4.13: Magnesium Sodium EDS mapping comparison. water sample was prepared with the following components: 0.45 mM Magnesium Chloride ($MgCl_2$), 0.25 mM Sodium Bicarbonate ($NaHCO_3$), 2.0 mM Magnesium Sulfate ($MgSO_4$), and 0.75 mM Calcium Chloride ($CaCl_2$).

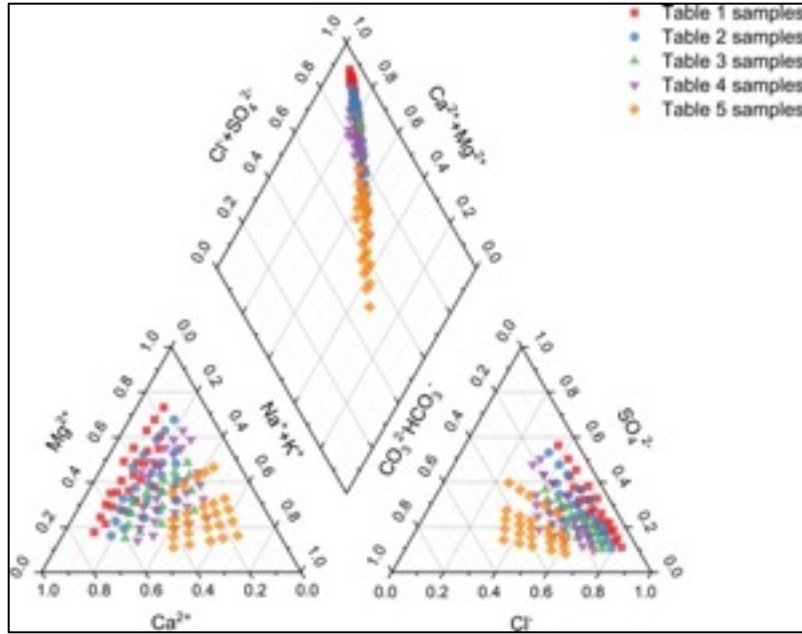


Figure 4.14: Trilinear plot of water recipes.

The trilinear plot of water sample recipes, as depicted in Figure. 4.14, effectively demonstrates the wide range of element concentrations found in various tap water samples. These samples are distributed across the plot to account for the inherent variability of tap water components that may be encountered in different geographical regions and under diverse environmental conditions. This comprehensive representation of tap water compositions enables a more thorough analysis and understanding of the various factors influencing water quality, ultimately supporting the development and evaluation of the vision-transformer model in this study.

4.4.3 Two-stage model produces better results than one-stage model

Water contaminants elements concentrations were predicted by the two-stage vision-transformer point estimation model, one-stage point estimation model and two-stage vision-transformer multiple output model. Results were plotted independently by target concentrations (x-axis) versus predicted concentrations (y-axis). Elements were labeled by independent color.

Two-stage vision-transformer point estimation model (TwoVtPeM)

Figure 4.15 displays the predicted and true (target) chlorine-sulfur mass ratios. The predicted chlorine to sulfur mass ratio is found to be higher than the true values, particularly when the true values are larger. This is consistent with the overestimation of concentration seen in the results of the TwoVtPeM (Fig. 4.17). The reason for this overestimation will be discussed in the following sections.

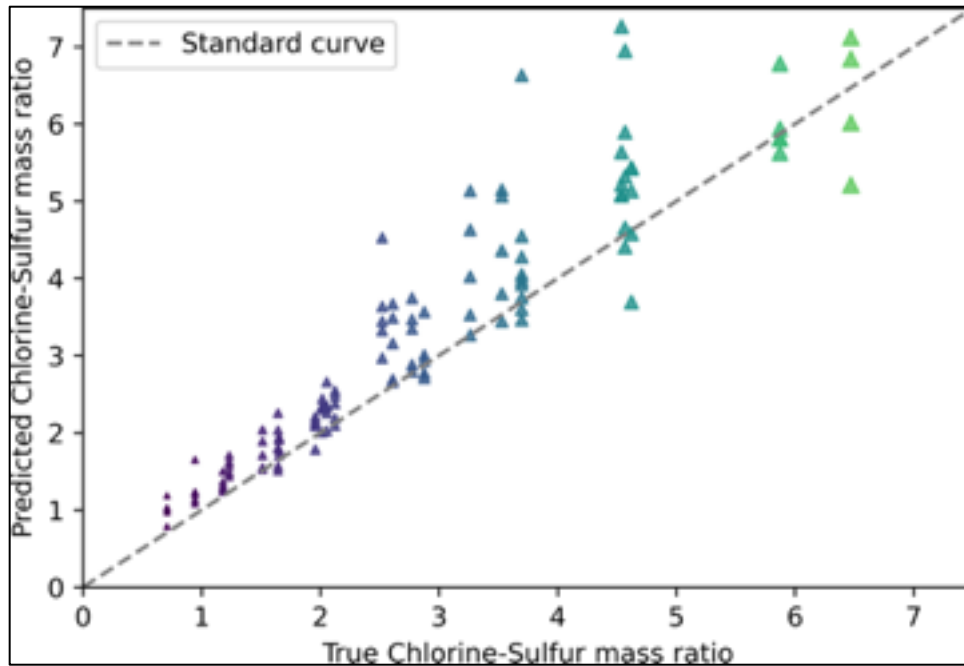


Figure 4.15: TwoVtPeM chlorine to sulfur mass ratio. Targets chlorine to sulfur mass ratio vs predictions chlorine to Sulfur mass ratio. Marker colors relates target chlorine to sulfur ratio value.

The predicted water hardness values tend to be higher than the true hardness values of the water samples, as shown in Fig. 4.16. For instance, twenty hard water samples were predicted as very hard, and five moderately hard water samples were predicted as hard. Nineteen hard water samples and eighty very hard water samples were correctly predicted. Only one sample had a predicted hardness lower than its true hardness. This is due to the overestimation of calcium concentrations, as seen in Fig. 4.17. The reason for this overestimation will be discussed in the following section.

The concentrations of contaminants estimated by the TwoVtPeM model are displayed in Figure 4.17. The target concentrations (elements concentrations in the water preparation recipe) are plotted on the x-axis, while the predicted concentrations are plotted on the y-axis. The results indicate that the predicted chlorine concentrations are generally higher than the true chlorine concentrations. This is consistent with the EDS mapping results in Figure 4.12 which show that the false negative value is lower than the false positive value. This suggests that some of the estimated chlorine crystals are not actually chlorine, leading to an overestimation of the chlorine concentration. Additionally, the estimation of chlorine has a larger standard deviation, which is likely due to the relatively high concentrations of chlorine compared to other elements in the water samples. As shown in Table 4.1, the predicted chlorine crystals are larger than true chlorine crystals.

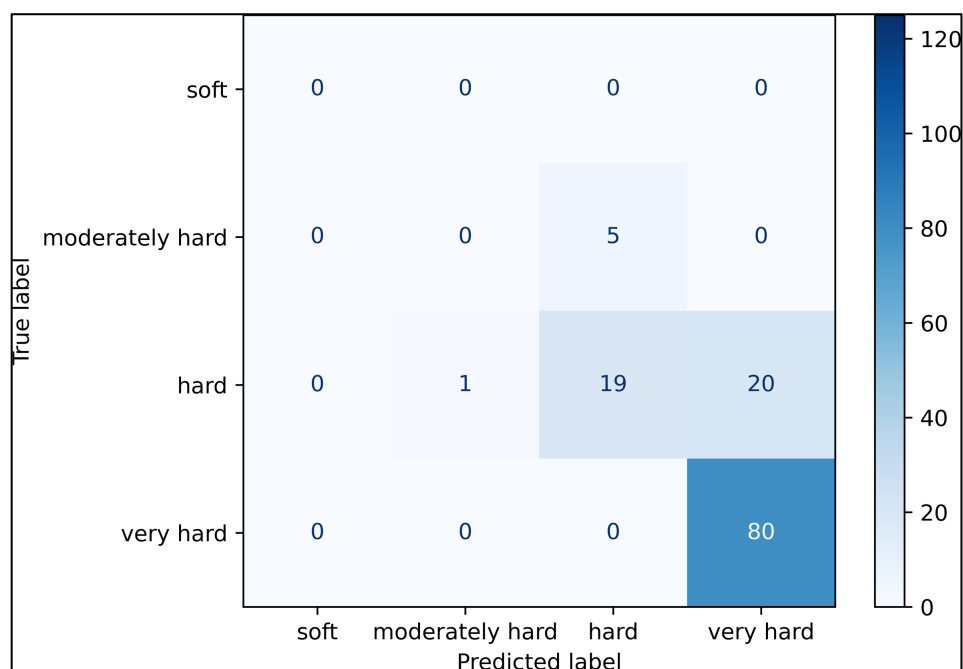


Figure 4.16: TwoVtPeM of water samples hardness category classification results.

The trilinear plot of the estimated element concentrations by the TwoVtPeM is presented in Figure. 4.18. When comparing this result with the true element concentrations

trilinear plot in Figure. 4.14, it is apparent that the water samples in the same table of water recipes are situated in similar locations. This observation indicates that the TwoVtPeM has successfully estimated element concentrations, demonstrating the effectiveness and accuracy of the model in analyzing and characterizing various tap water compositions.

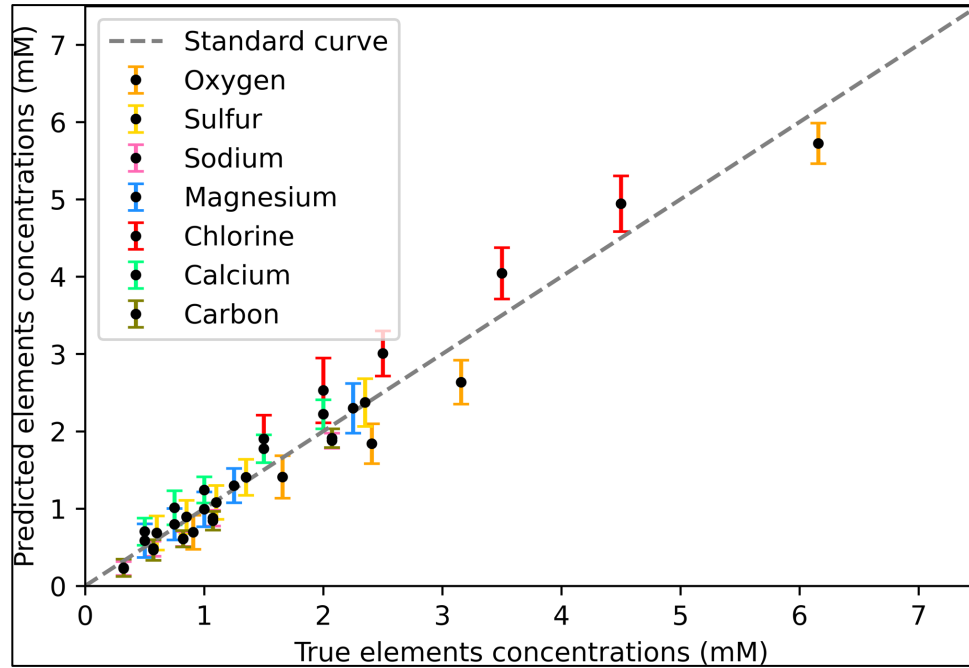


Figure 4.17: TwoVtPeM results. Targets (x-axis) vs predictions (y-axis).

One-stage point estimation model (OnePeM)

Figure 4.19 displays the predicted and true (target) chlorine to sulfur mass ratios. Different from the overestimated chlorine to sulfur mass ratio in the TwoVtPeM, the estimated chlorine to sulfur mass ratio is overestimated when the true chlorine to sulfur mass ratio is low but underestimated by the OnePeM especially when the true chlorine to sulfur mass ratio is high. This is consistent with the elements concentrations estimations Figure 4.21 that chlorine concentration is overestimated under its low concentrations condition but underestimated under its high concentration condition.

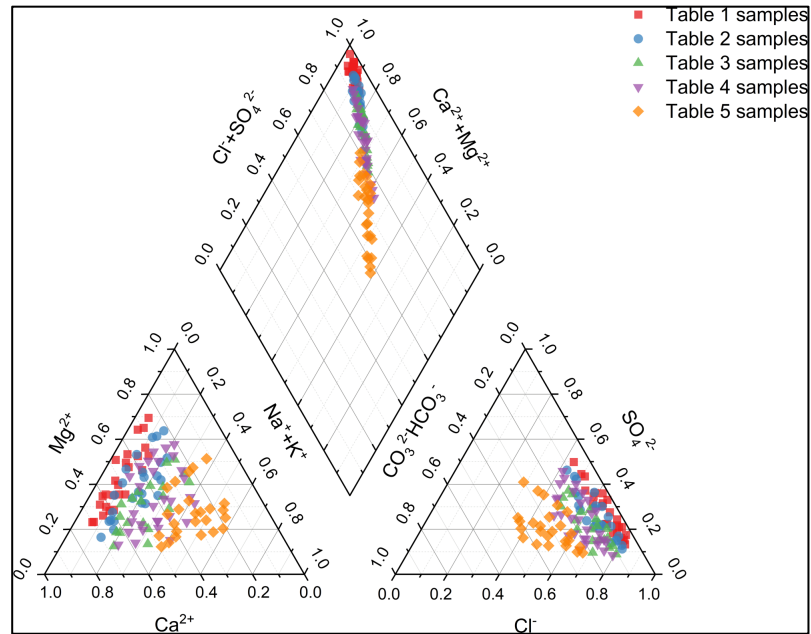


Figure 4.18: TwoVtPeM of water samples trilinear plot.

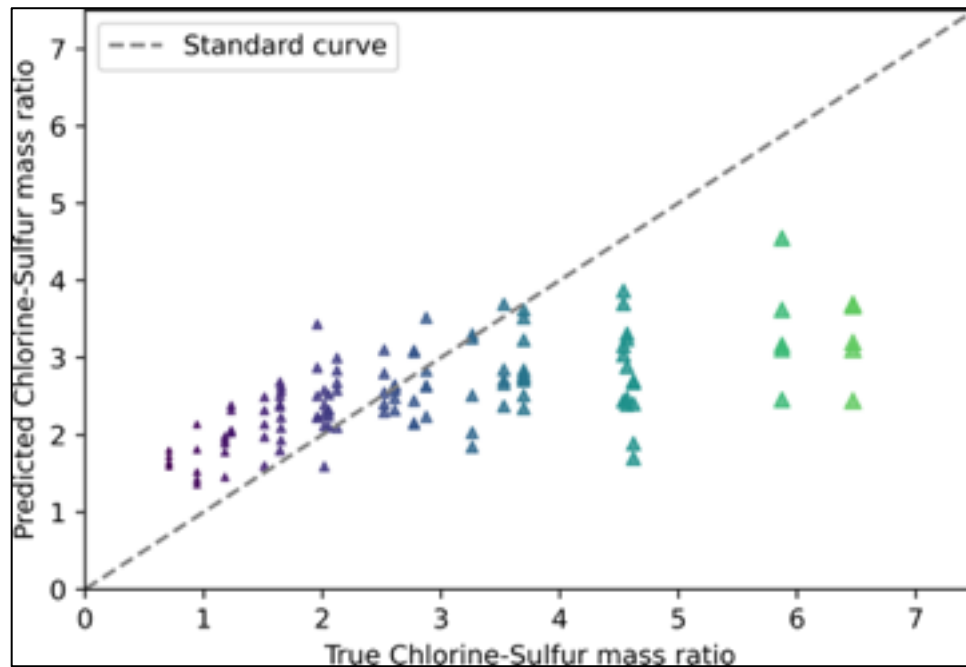


Figure 4.19: OnePeM chlorine to sulfur mass ratio. Targets chlorine to sulfur mass ratio vs predictions chlorine to sulfur mass ratio. Marker colors relates target chlorine to sulfur ratio value.

The predicted water hardness values also tend to be higher than the true hardness values of the water samples, as shown in Fig. 4.20. For instance, thirty-four hard water samples were predicted as very hard, three moderately hard water samples were predicted

as hard and two moderately hard water samples predicted as very hard. six hard water samples and eighty very hard water samples were correctly predicted. Only two samples had a predicted hardness lower than its true hardness. This is due to the overestimation of calcium and magnesium concentrations under low concentration conditions, as seen in Fig. 4.21. The reason for this overestimation will be discussed in the following section.

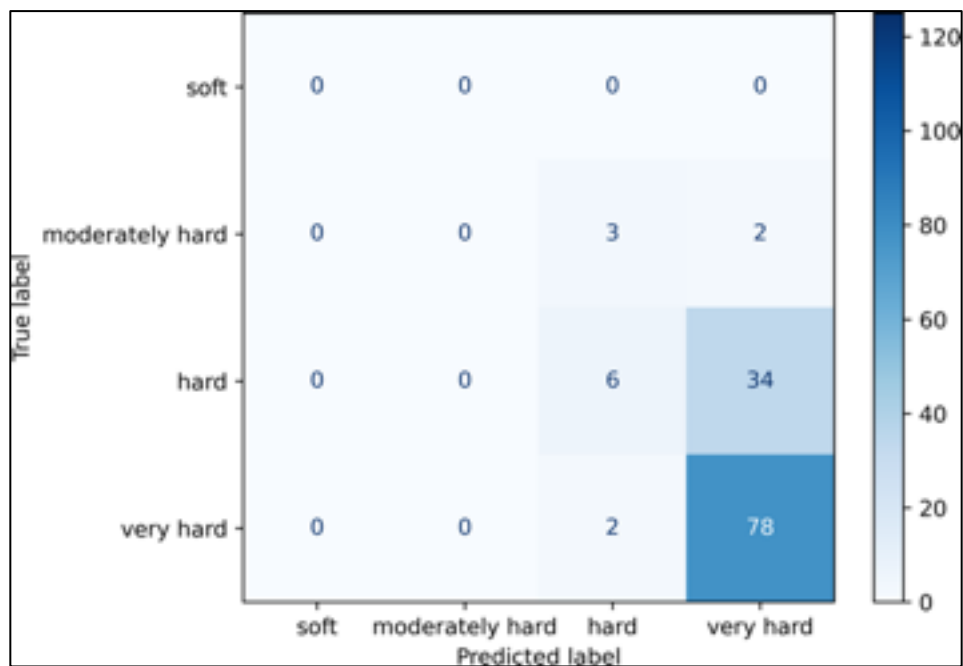


Figure 4.20: OnePeM of water samples hardness category classification results.

Figure 4.21 displays the concentrations of contaminants estimated by the one-stage point estimation model. In comparison to the TwoVtPeM, the OnePeM results in a greater standard deviation in the predicted concentrations. Additionally, the model tends to overestimate low concentrations and underestimate high concentrations of each element. For example, the predicted calcium concentration is higher than its true concentration when it is around 2 mM, but lower than its true concentration when it is around 3.5 mM. This is because the one-stage model is trained end-to-end, lacking the correction step present in the TwoVtPeM that adjusts for the EDS mapping estimation. As a result, the model requires more training epochs and fine-tuning to effectively learn the features.

The trilinear plot of the estimated element concentrations by the OnePeM is presented in Figure. 4.22. When comparing this result with the true element concentrations trilinear plot in Figure. 4.14, it is apparent that the water samples in the same table of water recipes are situated in similar locations, but not as accurately as the TwoVtPeM. This observation indicates that while the OnePeM is capable of estimating element concentrations, its performance is not as precise as the TwoVtPeM.

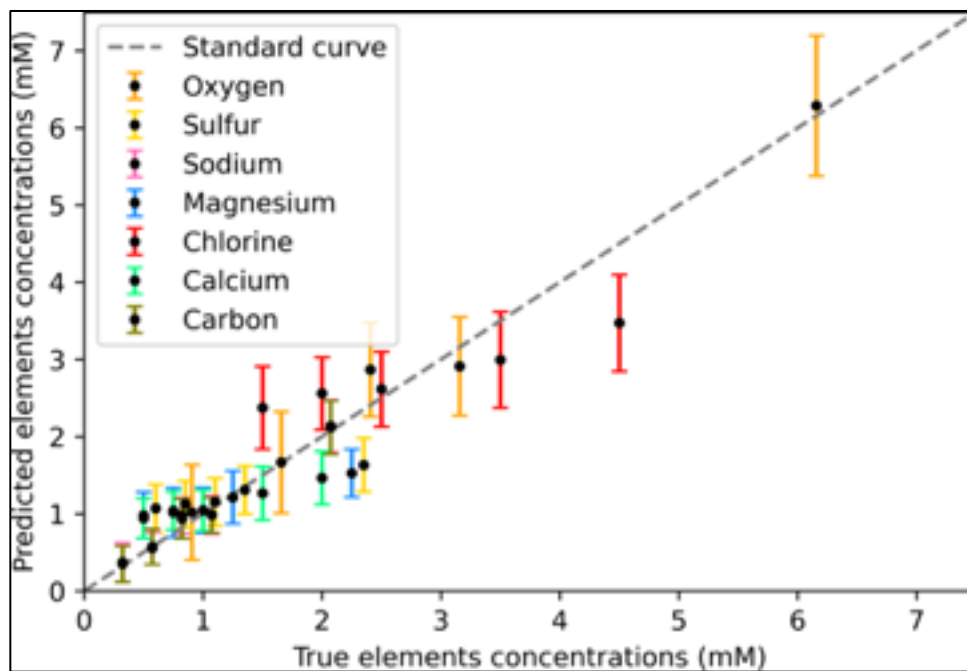


Figure 4.21: OnePeM results. Targets (x-axis) vs predictions (y-axis).

Two-stage vision-transformer multiple output estimation model (TwoVtMoM)

While the TwoVtMoM was expected to produce more accurate results than the TwoVtPeM, its element concentration estimations are actually worse. The model tends to overestimate low true element concentrations and underestimate high true element concentrations. This is due to the larger number of parameters in the TwoVtMoM model, which requires more training epochs and fine-tuning to effectively learn the features.

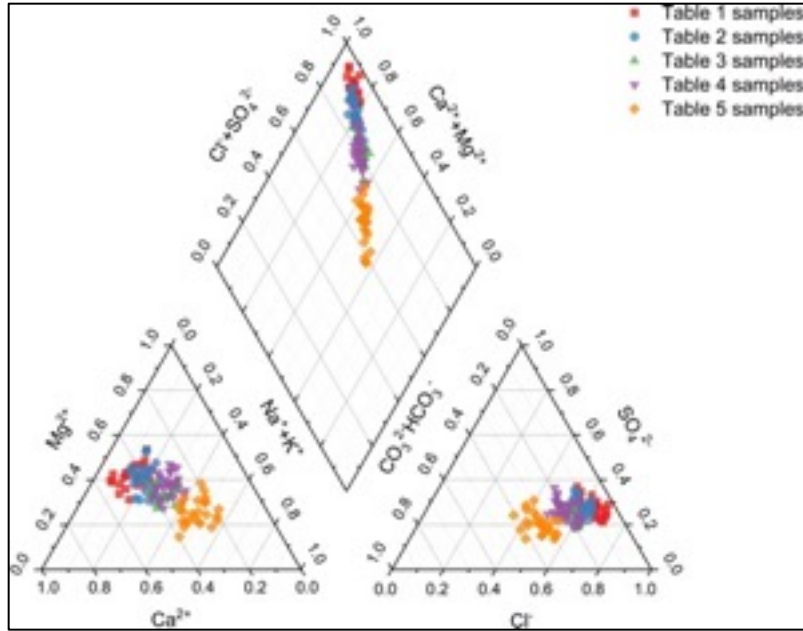


Figure 4.22: OnePeM model of water samples trilinear plot.

The results of each element estimation for the different models are summarized in Figure 4.24. As illustrated in the figure, the TwoVtPeM (row 1) exhibits lower variance compared to the OnePeM (row 2). The one-stage point estimation model tends to predict lower element concentrations than the actual values. This is due to the fact that the TwoVtPeM more accurately maps the elements' locations compared to the OnePeM. Although crystals form in a 3D structure, the EDS mapping can only represent the elements' 2D deposition. The TwoVtPeM can utilize relative location information from other elements to construct the crystal deposition structure and infer the corresponding concentrations.

The error mean (calculated as the percentage difference between the mean of the estimated element concentrations and their true concentrations) and standard deviation of concentration estimations (calculated as the standard deviation of estimated element concentrations) are presented in Table 4.2. The OnePeM has the lowest error mean for five elements (oxygen, sodium, chlorine, calcium, and carbon) out

of the seven elements, while the TwoVtPeM has the lowest error mean for the remaining two elements (sulfur and magnesium). Although the OnePeM has the lowest error mean, the TwoVtPeM has the lowest standard deviation for all seven element concentration estimations. This demonstrates that the TwoVtPeM is more stable than the OnePeM, which is due to the elements EDS mapping estimation module in its first stage, resulting in an R^2 of 0.95, which is higher than the 0.9 of the OnePeM.

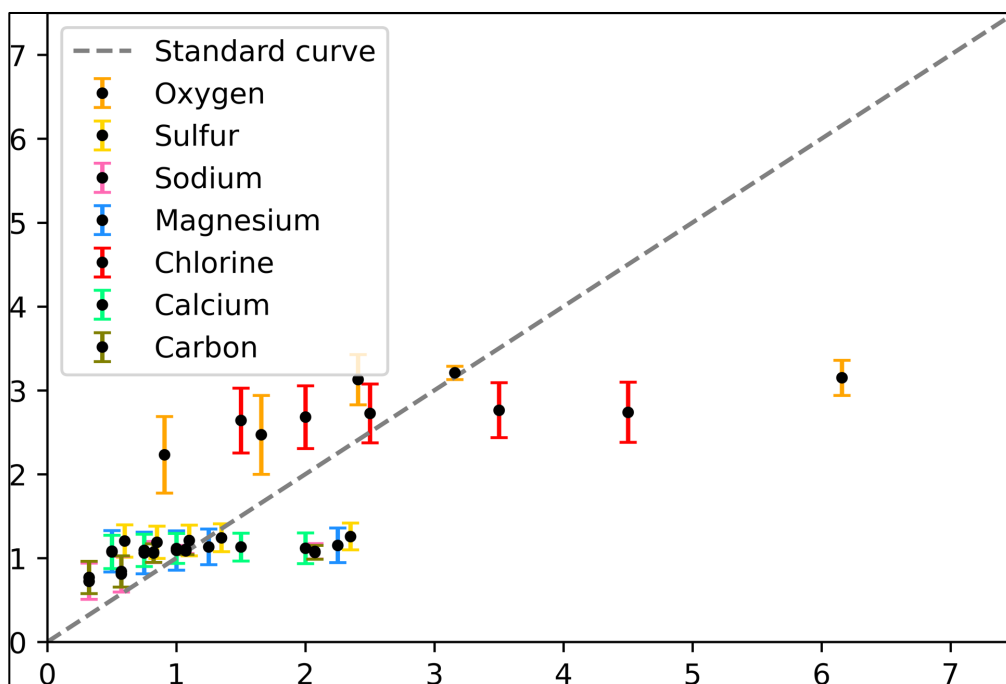


Figure 4.23: TwoVtMoM results. Targets (x-axis) vs predictions (y-axis).

Model comparison

In Section 4.4, the individual results of the three models regarding their element concentration estimations are presented. To compare the three models, the element concentration estimations and relative standard deviations are illustrated in Figure 4.24 and Table 4.2. From this figure, it is evident that TwoVtPeM outperforms the other models with lower variance and higher R^2 . The OnePeM concentration estimations are accurate for nonmetals oxygen, chlorine, and sulfur; however, its estimations are not precise for metals sodium, calcium, magnesium, and carbon. The

TwoVtPeM is more accurate for all elements. The TwoVtMoM is the least effective model, with the highest variance and lowest R^2 .

According to model performance analysis: the TwoVtPeM technique achieved the best performance of the models tested (OnePeM, TwoVtPeM and TwoVtMoM), with OnePeM also performing well and TwoVtMoM falling short. The TwoVtPeM relative percentage errors were $\pm 17.1\%$ for oxygen, $\pm 4.5\%$ for sulfur, $\pm 19.9\%$ for sodium, $\pm 5.7\%$ for chlorine, $\pm 19.8\%$ for calcium, $\pm 25.8\%$ for magnesium, and $\pm 20.1\%$ for carbon. The R^2 was 0.95 which is higher than OnePeM with 0.90 R^2 and TwoVtMoM which was 0.54. The TwoVtPeM had a higher error mean than OnePeM, but it exhibited lower relative standard deviations of estimation; the TwoVtPeM relative standard deviations values were: 3.9% for oxygen, 3.0% for sulfur, 5.3% for sodium, 3.9% for magnesium, 5.3% for chlorine, 10.0% for calcium, and 5.9% for carbon. Moreover, 79.2% of water samples were correctly classified for hardness based on the estimated element concentrations by TwoVtPeM. The OnePeM model correctly classified 67.2% of water samples, however the TwoVtMoM model achieved only 60.2% accuracy rate in classifying water samples for hardness Table 4.2.

Although the OnePeM has the relative error for oxygen, sodium, chlorine, calcium, and carbon, it exhibits larger relative standard deviations than the estimations of TwoVtPeM, indicating that the OnePeM is less stable. The TwoVtPeM has the lowest standard deviation for all seven element concentration estimations, demonstrating greater stability than the OnePeM. This is attributed to the element EDS mapping estimation module in its first stage. The TwoVtPeM can utilize relative location information from other elements to construct the crystal deposition structure and infer the corresponding concentrations.

The TwoVtMoM was expected to have the lowest relative error and highest R^2 , but this was not the case. This is due to the larger number of parameters in the TwoVtMoM model,

which necessitates more training epochs and fine-tuning to effectively learn the features. To apply this method in water quality monitoring, further research is required to investigate the reasons for the TwoVtMoM's poor performance and explore methods to enhance it.

Another necessary step is to develop a model that transfers from the element concentration estimation model based on water SEM fingerprints to one based on water photograph fingerprints. The rationale is that SEM images are more accurate than photographs, but SEM images are not available in households or in the field. The model built from water SEM fingerprints is only used for learning crystal features from water residue patterns, and this information is solely for constructing the element concentration estimation model from water photograph fingerprints. Thus, in the future, when the element concentration estimation model from water photograph fingerprints is developed, only water photograph fingerprints will be needed for element concentration estimation.

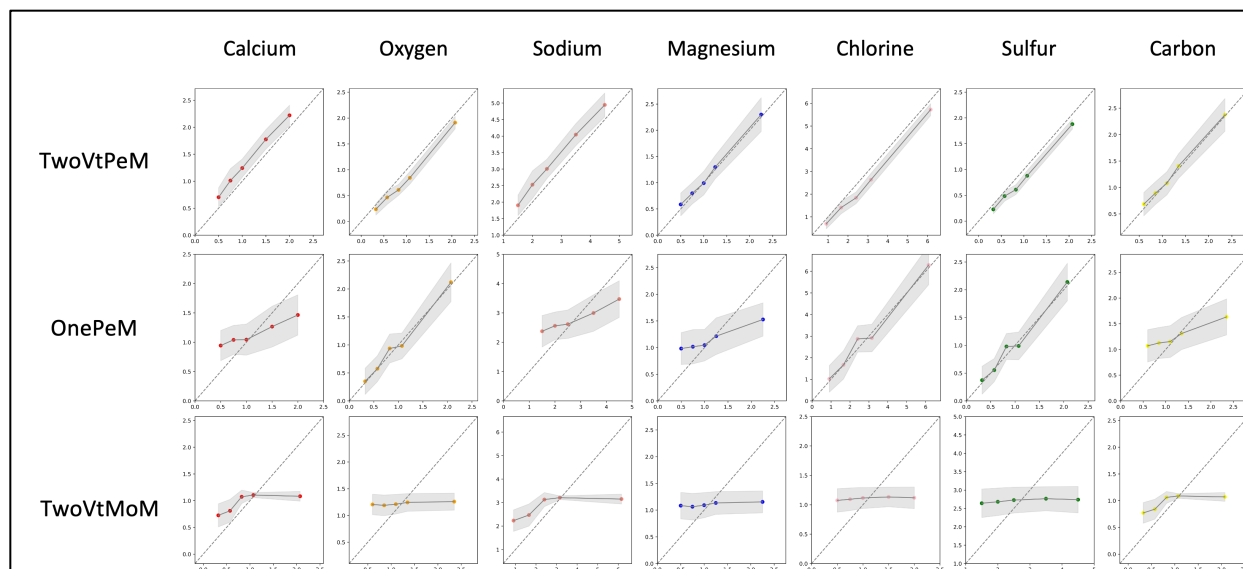


Figure 4.24: TwoVtMoM (row 1) produces lower variance than one-stage point estimation model (row 2). OnePeM predicts lower elements concentrations than their real values.

Table 4.2: Comparing Estimation Results of Model Element Concentrations.

Models	Oxygen	Sulfur	Sodium	Magnesium	Chlorine	Calcium	Carbon	R^2
Relative Error (%)								
OnePeM	±5.2%	±16.4%	±5.2%	±20.0%	±10.7%	±17.9%	±3.2%	0.90
TwoVtPeM	±17.1%	±4.5%	±19.8%	±5.7%	±19.7%	±25.8%	±20.1%	0.95
TwoVtMoM	±35.5%	±19.3%	±30.2%	±21.9%	±11.8%	±20.7%	±33.3%	0.54
Relative Standard Deviation Error (%)								
OnePeM	6.9%	19.7%	8.0%	27.9%	12.2%	24.6%	6.2%	
TwoVtPeM	3.9%	3.0%	5.3%	3.9%	5.3%	10.0%	5.9%	
TwoVtMoM	59.0%	31.0%	46.8%	42.3%	20.3%	39.9%	53.1%	
Coefficient of Variation (%)								
OnePeM	33.5%	20.7%	36.4%	34.7%	22.0%	30.2%	36.5%	
TwoVtPeM	13.0%	22.4%	14.1%	25.0%	14.1%	20.5%	17.6%	
TwoVtMoM	19.4%	18.1%	25.6%	26.4%	15.1%	20.7%	22.7%	
Mean Absolute Percentage Error (%)								
OnePeM	±18.1%	±33.2%	±20.3%	±37.3%	±38.9%	±27.1%	±17.4%	
TwoVtPeM	±17.1%	±13.3%	±20.7%	±25.9%	±15.7%	±19.8%	±20.3%	
TwoVtMoM	±55.2%	±42.2%	±49.6%	±49.2%	±47.6%	±36.8%	±52.8%	
Root Mean Square Error								
OnePeM	0.52	0.44	0.18	0.39	0.45	0.79	0.17	
TwoVtPeM	0.45	0.18	0.18	0.27	0.18	0.54	0.18	
TwoVtMoM	1.57	0.60	0.51	0.54	0.60	0.40	1.09	
Mean Square Error								
OnePeM	0.27	0.19	0.04	0.16	0.20	0.62	0.03	
TwoVtPeM	0.20	0.03	0.03	0.07	0.03	0.29	0.03	
TwoVtMoM	2.47	0.36	0.26	0.29	0.37	1.19	0.27	

4.5 Conclusion

Machine learning is transforming the way we approach water quality and public health. This study shows the potential of machine learning to revolutionize water quality monitoring. With the use of low-cost aluminum substrates, the overall cost of the experiment is significantly lower than traditional analytical methods, making this technique a cost-effective solution for water quality monitoring. The method is especially useful in rural areas and in the event of potential pollution incidents, where early detection is crucial.

The findings of this study reveal that the TwoVtPeM technique achieved the best performance of the models tested (OnePeM, TwoVtPeM and TwoVtMoM), with OnePeM also performing well and TwoVtMoM falling short. The TwoVtPeM relative percentage errors were $\pm 17.1\%$ for oxygen, $\pm 4.5\%$ for sulfur, $\pm 19.9\%$ for sodium, $\pm 5.7\%$ for chlorine, $\pm 19.8\%$ for calcium, $\pm 25.8\%$ for magnesium, and $\pm 20.1\%$ for carbon. The R^2 was 0.95 which is higher than OnePeM with 0.90 R^2 and TwoVtMoM which was 0.54. The TwoVtPeM had a higher error mean than OnePeM, but it exhibited lower relative standard deviations of estimation; the TwoVtPeM relative standard deviations values were: 3.9% for oxygen, 3.0% for sulfur, 5.3% for sodium, 3.9% for magnesium, 5.3% for chlorine, 10.0% for calcium, and 5.9% for carbon. Moreover, 79.2% of water samples were correctly classified for hardness based on the estimated element concentrations by TwoVtPeM. The OnePeM model correctly classified 67.2% of water samples, however the TwoVtMoM model achieved only 60.2% accuracy rate in classifying water samples for hardness.

Advances in camera technology and deep learning techniques hold great potential for improving the method's ability to detect low concentrations of elements. By using substrates

with varying surface properties, such as roughness, wettability, charge, and others, different crystal formations can be produced that can be designed to monitor specific contaminants. The two-stage vision-transformer multiple output model produces a smaller variance, but the concentration estimation is not always accurate, requiring more fine-tuning and training epochs.

To detect low concentrations of elements, water samples with lower concentrations need to be prepared and the coffee-ring effect residue pattern collected. Confirmation of the crystal structure can be obtained through Raman spectroscopy on the water sample residue. To analyze the one-stage point estimation model performance, the intermediate output of the seven element mappings can be compared with the predicted EDS mapping of the two-stage vision-transformer point estimation model. This will provide insights into the strengths and weaknesses of each model, allowing for further improvements to be made.

An additional avenue for improvement is the creation of a loss function that takes into account not only the pixel classes but also their structure. Contaminants often have distinct 3D lattice structures, and this information could be leveraged in the loss function. Additionally, incorporating domain knowledge from physical chemistry could also be beneficial. For instance, magnesium and calcium crystals are unlikely to form crystals at the same location, but calcium and sulfur are more likely to form calcium sulfate first due to their relatively low K_{sp} values compared to other crystals such as sodium chloride and calcium chloride.

In conclusion, this study highlights the potential of machine learning to revolutionize water quality monitoring. By improving the efficiency and effectiveness of water quality management systems, machine learning has the potential to lead to better health outcomes for individuals and communities. With continued advancements in technology

and machine learning techniques, we can expect to see even more exciting developments in this field in the future.

CHAPTER 5

Implications

Machine learning is revolutionizing both water quality and public health. In the realm of water quality, machine learning is employed to create predictive models that shed light on the relationships between various water quality parameters and the impact of different factors. This results in the creation of early warning systems that can identify potential water quality problems, enabling proactive solutions. Machine learning also enables the analysis of large amounts of data and extraction of previously hidden insights, leading to a deeper understanding of water quality and new methods for managing this critical resource. By automating certain tasks and simplifying data analysis processes, machine learning has the potential to enhance the efficiency and effectiveness of water quality management systems. In public health, machine learning algorithms are trained on medical images and patient records to diagnose diseases and predict future health outcomes. They are also utilized to analyze and forecast the spread of infectious diseases, providing crucial support to public health officials.

Machine learning is integrated into environmental monitoring systems, providing real-time data analysis for environmental facilities and resulting in more informed management decisions. Additionally, machine learning algorithms can predict the risk of specific environmental issues, such as pollution events or habitat degradation, allowing for early interventions and preventive measures. Machine learning has the potential to significantly improve the efficiency and effectiveness of environmental initiatives, leading to better environmental outcomes for ecosystems and communities. The impact of machine

learning on water quality and public health is substantial and has the potential to fundamentally change the way we approach and manage these critical resources. Through the use of advanced machine learning techniques, we can gain a deeper understanding of water quality, create new and innovative solutions for preserving this essential resource, and protect public health for future generations.

This study underscores the potential of machine learning to transform water quality monitoring. By enhancing the efficiency and effectiveness of water quality management systems, machine learning can be utilized for various image formats, including SEM, EDS, X-ray Powder Diffraction (XRD), Raman spectroscopy, images collected in rural areas, and even satellite data covering larger areas. Consequently, machine learning could potentially result in better health outcomes for individuals and communities. As technology and machine learning techniques continue to advance, we can anticipate further groundbreaking developments in this field that will contribute to ensuring cleaner water and healthier environments for all. As a screening method, this research demonstrates the effectiveness of machine learning techniques in water quality monitoring. With improvements in camera technology, material science, and model design, such as the development of multimodal techniques incorporating local weather, groundwater information, pipe information, and environmental incidents, this approach shows great promise as a fast, low-cost, and accurate water quality monitoring technique.

BIBLIOGRAPHY

Tak-Sing Wong, Ting-Hsuan Chen, Xiaoying Shen, and Chih-Ming Ho. Nanochromatography driven by the coffee ring effect. *Analytical chemistry*, 83(6):1871–1873, 2011.

Pavlo Takhistov and Hsueh-Chia Chang. Complex stain morphologies. *Industrial & engineering chemistry research*, 41(25):6256–6269, 2002.

Noushine Shahidzadeh-Bonn, Salima Rafai, Daniel Bonn, and Gerard Wegdam. Salt crystallization during evaporation: impact of interfacial properties. *Langmuir*, 24(16):8599–8605, 2008.

Xin Zhong, Junheng Ren, and Fei Duan. Wettability effect on evaporation dynamics and crystalline patterns of sessile saline droplets. *The Journal of Physical Chemistry B*, 121(33):7924–7933, 2017.

D Kaya, VA Belyi, and M Muthukumar. Pattern formation in drying droplets of polyelectrolyte and salt. *The Journal of chemical physics*, 133(11):114905, 2010.

Bongsu Shin, Myoung-Woon Moon, and Ho-Young Kim. Rings, igloos, and pebbles of salt formed by drying saline drops. *Langmuir*, 30(43):12837–12842, 2014.

Sooheyong Lee, Haeng Sub Wi, Wonhyuk Jo, Yong Chan Cho, Hyun Hwi Lee, Se-Young Jeong, Yong-Il Kim, and Geun Woo Lee. Multiple pathways of crystal nucleation in an extremely supersaturated aqueous potassium dihydrogen phosphate (kdp) solution droplet. *Proceedings of the National Academy of Sciences*, 113(48):13618–13623, 2016.

Hee-Soo Kim, Sung Soo Park, and Frank Hagelberg. Computational approach to drying a nanoparticle-suspended liquid droplet. *Journal of Nanoparticle Research*, 13:59–68, 2011.

Andrew Stannard. Dewetting-mediated pattern formation in nanoparticle assemblies. *Journal of Physics: Condensed Matter*, 23(8):083001, 2011.

Mark J Robbins, AJ Archer, and Uwe Thiele. Modelling the evaporation of thin films of colloidal suspensions using dynamical density functional theory. *Journal of Physics: Condensed Matter*, 23(41):415102, 2011.

Dinesh Gupta and Michael H Peters. A brownian dynamics simulation of aerosol deposition onto spherical collectors. *Journal of Colloid and interface Science*, 104(2):375–389, 1985.

Jim C Chen and Albert S Kim. Brownian dynamics, molecular dynamics, and monte carlo modeling of colloidal systems. *Advances in colloid and interface science*, 112(1-3):159–173, 2004.

Saeed Jafari Kang, Vahid Vandadi, James D Felske, and Hassan Masoud. Alternative mechanism for coffee-ring deposition based on active role of free surface. *Physical Review*

E, 94(6):063104, 2016.

Benjamin J Fischer. Particle convection in an evaporating colloidal droplet. *Langmuir*, 18(1):60–67, 2002.

Leonid Shmuylovich, Amy Q Shen, and Howard A Stone. Surface morphology of drying latex films: Multiple ring formation. *Langmuir*, 18(9):3441–3445, 2002.

L Pauchard and C Allain. Stable and unstable surface evolution during the drying of a polymer solution drop. *Physical Review E*, 68(5):052801, 2003.

Yuri O Popov. Evaporative deposition patterns: spatial dimensions of the deposit. *Physical Review E*, 71(3):036313, 2005.

T Heim, S Preuss, B Gerstmayer, A Bosio, and R Blossey. Deposition from a drop: morphologies of unspecifically bound dna. *Journal of Physics: Condensed Matter*, 17(9):S703, 2005.

Peter J Yunker, Tim Still, Matthew A Lohr, and AG Yodh. Suppression of the coffee-ring effect by shape-dependent capillary interactions. *nature*, 476(7360):308–311, 2011.

Ronald G Larson. Transport and deposition patterns in drying sessile droplets. *AIChE Journal*, 60(5):1538–1571, 2014.

Jake Graser, Steven K Kauwe, and Taylor D Sparks. Machine learning and energy minimization approaches for crystal structure predictions: a review and new horizons. *Chemistry of Materials*, 30(11):3601–3612, 2018.

Jungho Park and Jooho Moon. Control of colloidal particle deposit patterns within picoliter droplets ejected by ink-jet printing. *Langmuir*, 22(8):3506–3513, 2006.

Andreas Friederich, Joachim R Binder, and W Bauer. Rheological control of the coffee stain effect for inkjet printing of ceramics. *Journal of the American Ceramic Society*, 96(7): 2093–2099, 2013.

Minxuan Kuang, Libin Wang, and Yanlin Song. Controllable printing droplets for high-resolution patterns. *Advanced materials*, 26(40):6950–6958, 2014.

Jiazhen Sun, Bin Bao, Min He, Haihua Zhou, and Yanlin Song. Recent advances in controlling the depositing morphologies of inkjet droplets. *ACS applied materials & interfaces*, 7(51):28086–28099, 2015.

Qijin Huang and Yong Zhu. Printing conductive nanomaterials for flexible and stretchable electronics: A review of materials, processes, and applications. *Advanced Materials Technologies*, 4(5):1800546, 2019.

Wei Han and Zhiqun Lin. Learning from “coffee rings”: Ordered structures enabled by controlled evaporative self-assembly. *Angewandte Chemie International Edition*, 51(7):

1534–1546, 2012.

Jiao-Jing Shao, Wei Lv, and Quan-Hong Yang. Self-assembly of graphene oxide at interfaces. *Advanced Materials*, 26(32):5586–5612, 2014.

Jianli Zou and Franklin Kim. Diffusion driven layer-by-layer assembly of graphene oxide nanosheets into porous three-dimensional macrostructures. *Nature communications*, 5(1): 1–9, 2014.

Jungho Park, Jooho Moon, Hyunjung Shin, Dake Wang, and Minseo Park. Direct-write fabrication of colloidal photonic crystal microarrays by ink-jet printing. *Journal of colloid and interface science*, 298(2):713–719, 2006.

Liying Cui, Yingfeng Li, Jingxia Wang, Entao Tian, Xingye Zhang, Youzhuan Zhang, Yanlin Song, and Lei Jiang. Fabrication of large-area patterned photonic crystals by ink-jet printing. *Journal of Materials Chemistry*, 19(31):5499–5502, 2009.

Ralf Blossey and Andreas Bosio. Contact line deposits on cdna microarrays: a “twin-spot effect”. *Langmuir*, 18(7):2952–2954, 2002.

Vincent Dugas, Jérôme Broutin, and Eliane Souteyrand. Droplet evaporation study applied to dna chip manufacturing. *Langmuir*, 21(20):9130–9136, 2005.

Jie-Bi Hu, Yu-Chie Chen, and Pawel L Urban. Coffee-ring effects in laser desorption/ionization mass spectrometry. *Analytica chimica acta*, 766:77–82, 2013.

D Mampallil, HB Eral, D Van Den Ende, and F Mugele. Control of evaporating complex fluids through electrowetting. *Soft Matter*, 8(41):10614–10617, 2012.

Olena Kudina, Burak Eral, and Frieder Mugele. e-maldi: an electrowetting-enhanced drop drying method for maldi mass spectrometry. *Analytical chemistry*, 88(9):4669–4675, 2016.

Yin-Hung Lai, Yi-Hong Cai, Hsun Lee, Yu-Meng Ou, Chih-Hao Hsiao, Chien-Wei Tsao, Huan-Tsung Chang, and Yi-Sheng Wang. Reducing spatial heterogeneity of maldi samples with marangoni flows during sample preparation. *Journal of The American Society for Mass Spectrometry*, 27(8):1314–1321, 2016.

Wei-dong Zhou, Jia-nan Cai, Long Sun, and Chen Shen. Time–space difference based gps/sins ultra-tight integrated navigation method. *Measurement*, 58:87–92, 2014a.

Weidong Wang, Yongguang Yin, Zhiqiang Tan, and Jingfu Liu. Coffee-ring effect-based simultaneous sers substrate fabrication and analyte enrichment for trace analysis. *Nanoscale*, 6(16):9588–9593, 2014.

Jose L Garcia-Cordero and Z Hugh Fan. Sessile droplets for chemical and biological assays. *Lab on a Chip*, 17(13):2150–2166, 2017.

Penghui Li, Yong Li, Zhang-Kai Zhou, Siying Tang, Xue-Feng Yu, Shu Xiao, Zhongzhen Wu, Quanlan Xiao, Yuetao Zhao, Huaiyu Wang, et al. Evaporative self-assembly of gold nanorods into macroscopic 3d plasmonic superlattice arrays. *Advanced Materials*, 28(13): 2511–2517, 2016a.

David Brutin, Benjamin Sobac, Boris Loquet, and José Sampaol. Pattern formation in drying drops of blood. *Journal of fluid mechanics*, 667:85–95, 2011.

Jessica T Wen, Chih-Ming Ho, and Peter B Lillehoj. Coffee ring aptasensor for rapid protein detection. *Langmuir*, 29(26):8440–8446, 2013.

Christopher P Gulka, Joshua D Swartz, Joshua R Trantum, Keersten M Davis, Corey M Peak, Alexander J Denton, Frederick R Haselton, and David W Wright. Coffee rings as low-resource diagnostics: detection of the malaria biomarker plasmodium falciparum histidine-rich protein-ii using a surface-coupled ring of ni (ii) nta gold-plated polystyrene particles. *ACS applied materials & interfaces*, 6(9):6257–6263, 2014.

Berend-Jan de Gans and Ulrich S Schubert. Inkjet printing of well-defined polymer dots and arrays. *Langmuir*, 20(18):7789–7793, 2004.

HB Eral, DJCM t Mannetje, and Jung Min Oh. Contact angle hysteresis: a review of fundamentals and applications. *Colloid and polymer science*, 291(2):247–260, 2013.

Dongliang Tian, Yanlin Song, and Lei Jiang. Patterning of controllable surface wettability for printing techniques. *Chemical society reviews*, 42(12):5184–5209, 2013.

Hwa-Young Ko, Jungho Park, Hyunjung Shin, and Jooho Moon. Rapid self-assembly of monodisperse colloidal spheres in an ink-jet printed droplet. *Chemistry of materials*, 16(22):4212–4215, 2004.

Huaiguang Li, Darren Buesen, Rhodri Williams, Joerg Henig, Stefanie Stapf, Kallol Mukherjee, Erik Freier, Wolfgang Lubitz, Martin Winkler, Thomas Happe, et al. Preventing the coffee-ring effect and aggregate sedimentation by in situ gelation of monodisperse materials. *Chemical Science*, 9(39):7596–7605, 2018.

Carmen L Moraila-Martinez, Miguel A Cabrerizo-Vilchez, and Miguel A Rodriguez-Valverde. Controlling the morphology of ring-like deposits by varying the pinning time of driven receding contact lines. *Interfacial Phenomena and Heat Transfer*, 1(3), 2013.

Tuan AH Nguyen, Marc A Hampton, and Anh V Nguyen. Evaporation of nanoparticle droplets on smooth hydrophobic surfaces: the inner coffee ring deposits. *The Journal of Physical Chemistry C*, 117(9):4707–4716, 2013.

Frieder Mugele and Jean-Christophe Baret. Electrowetting: from basics to applications. *Journal of physics: condensed matter*, 17(28):R705, 2005.

F Li and F Mugele. How to make sticky surfaces slippery: Contact angle hysteresis in

electrowetting with alternating voltage. *Applied Physics Letters*, 92(24):244108, 2008.

Dileep Mampallil and Huseyin Burak Eral. A review on suppression and utilization of the coffee-ring effect. *Advances in colloid and interface science*, 252:38–54, 2018.

Ruth Hernandez-Perez, Z Hugh Fan, and Jose L Garcia-Cordero. Evaporation-driven bioassays in suspended droplets. *Analytical chemistry*, 88(14):7312–7317, 2016.

F De Angelis, F Gentile, F Mecarini, G Das, M Moretti, P Candeloro, ML Coluccio, G Cojoc, A Accardo, C Liberale, et al. Breaking the diffusion limit with super-hydrophobic delivery of molecules to plasmonic nanofocusing sers structures. *Nature Photonics*, 5(11):682–687, 2011.

Ying Liu, Cheng Zhi Huang, and Yuan Fang Li. Fluorescence assay based on preconcentration by a self-ordered ring using berberine as a model analyte. *Analytical chemistry*, 74(21):5564–5568, 2002.

Cheng Zhi Huang, Ying Liu, and Yuan Fang Li. Microscopic determination of tetracycline based on aluminum-sensitized fluorescence of a self-ordered ring formed by a sessile droplet on glass slide support. *Journal of pharmaceutical and biomedical analysis*, 34(1):103–114, 2004a.

Chuanxiao Yang and Chengzhi Huang. Fluorescent microscopic determination of quinidine sulfate in serum samples with self-ordered ring technique by capillary flow effect. *Chinese Journal of Analytical Chemistry*, 34(2):183–187, 2006.

Ying Liu, YF Li, and CZ Huang. Fluorimetric determination of fluorescein at the femtomole level with a self-ordered ring of a sessile droplet on glass slide support. *Journal of Analytical Chemistry*, 61(7):647–653, 2006.

Lifeng Chen and Julian RG Evans. Drying of colloidal droplets on superhydrophobic surfaces. *Journal of colloid and interface science*, 351(1):283–287, 2010.

Ruoyang Chen, Liyuan Zhang, Xu Li, Lydia Ong, Ye Gaung Soe, Neil Sinsua, Sally L Gras, Rico F Tabor, Xungai Wang, and Wei Shen. Trace analysis and chemical identification on cellulose nanofibers-textured sers substrates using the “coffee ring” effect. *ACS sensors*, 2 (7):1060–1067, 2017.

Abid Hussain, Da-Wen Sun, and Hongbin Pu. Sers detection of urea and ammonium sulfate adulterants in milk with coffee ring effect. *Food Additives & Contaminants: Part A*, 36 (6):851–862, 2019.

Subhavna Juneja and Jaydeep Bhattacharya. Coffee ring effect assisted improved s. aureus screening on a physically restrained gold nanoflower enriched sers substrate. *Colloids and Surfaces B: Biointerfaces*, 182:110349, 2019.

Weiping Zhou, Anming Hu, Shi Bai, Ying Ma, and Quanshuang Su. Surface-enhanced raman spectra of medicines with large-scale self-assembled silver nanoparticle films based

on the modified coffee ring effect. *Nanoscale research letters*, 9(1):1–9, 2014b.

Xiaoyan Li, Alyssa R Sanderson, Selett S Allen, and Rebecca H Lahr. Tap water fingerprinting using a convolutional neural network built from images of the coffee-ring effect. *Analyst*, 145(4):1511–1523, 2020.

Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61: 85–117, 2015.

Yandong Li, Zongbo Hao, and Hang Lei. Survey of convolutional neural network. *Journal of Computer Applications*, 36(9):2508, 2016b.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553): 436–444, 2015.

Falk Schwendicke, Tatiana Golla, Martin Dreher, and Joachim Krois. Convolutional neural networks for dental image diagnostics: A scoping review. *Journal of dentistry*, 91:103226, 2019.

Titus Josef Brinker, Achim Hekler, Jochen Sven Utikal, Niels Grabe, Dirk Schadendorf, Joachim Klode, Carola Berking, Theresa Steeb, Alexander H Enk, and Christof Von Kalle. Skin cancer classification using convolutional neural networks: systematic review. *Journal of medical Internet research*, 20(10):e11936, 2018.

DE RUMBERT. Learning internal representations by error propagation. *Parallel distributed processing*, 1:318–363, 1986.

Y-T Zhou, Rama Chellappa, Aseem Vaid, and B Keith Jenkins. Image restoration using a neural network. *IEEE transactions on acoustics, speech, and signal processing*, 36(7): 1141–1151, 1988.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al. Recent advances in convolutional neural networks. *Pattern recognition*, 77:354–377, 2018.

Yann A LeCun, Léon Bottou, Genevieve B Orr, and Klaus-Robert Müller. Efficient backprop. In *Neural networks: Tricks of the trade*, pages 9–48. Springer, 2012.
Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines vinod nair. 2010.

Tao Wang, David J Wu, Adam Coates, and Andrew Y Ng. End-to-end text recognition with convolutional neural networks. In *Proceedings of the 21st international conference on pattern recognition (ICPR2012)*, pages 3304–3308. IEEE, 2012.

Y-Lan Boureau, Jean Ponce, and Yann LeCun. A theoretical analysis of feature pooling in

visual recognition. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 111–118, 2010.

Naila Murray and Florent Perronnin. Generalized max pooling. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2473–2480, 2014.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.

Geoffrey E Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012.

Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2013.

Shreyas Saxena and Jakob Verbeek. Convolutional neural fabrics. *Advances in neural information processing systems*, 29, 2016.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

Yichuan Tang. Deep learning using linear support vector machines. *arXiv preprint arXiv:1306.0239*, 2013.

Gjorgji Madjarov, Dragi Kocev, Dejan Gjorgjevikj, and Sašo Džeroski. An extensive experimental comparison of methods for multi-label learning. *Pattern recognition*, 45(9):3084–3104, 2012.

Xiao-Xiao Niu and Ching Y Suen. A novel hybrid cnn–svm classifier for recognizing handwritten digits. *Pattern Recognition*, 45(4):1318–1325, 2012.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Frank Rosenblatt. *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory, 1957.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

Ankur P Parikh, Oscar Täckström, Dipanjan Das, and Jakob Uszkoreit. A decomposable attention model for natural language inference. *arXiv preprint arXiv:1606.01933*, 2016.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.

Mark Chen, Alec Radford, Rewon Child, Jeffrey Wu, Heewoo Jun, David Luan, and Ilya Sutskever. Generative pretraining from pixels. In *International conference on machine learning*, pages 1691–1703. PMLR, 2020.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, pages 213–229. Springer, 2020.

Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:2010.04159*, 2020.

Sixiao Zheng, Jiachen Lu, Hengshuang Zhao, Xiatian Zhu, Zekun Luo, Yabiao Wang, Yanwei Fu, Jianfeng Feng, Tao Xiang, Philip HS Torr, et al. Rethinking semantic

segmentation from a sequence-to-sequence perspective with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6881–6890, 2021.

Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12299–12310, 2021.

Luowei Zhou, Yingbo Zhou, Jason J Corso, Richard Socher, and Caiming Xiong. End-to-end dense video captioning with masked transformer. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8739–8748, 2018.

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20: 273–297, 1995.

Kok Seng Chua. Efficient computations for large least square support vector machine classifiers. *Pattern Recognition Letters*, 24(1-3):75–80, 2003.

William S Noble. What is a support vector machine? *Nature biotechnology*, 24(12): 1565–1567, 2006.

Peter D Caie, Neofytos Dimitriou, and Ognjen Arandjelović. Precision medicine in digital pathology via image analysis and machine learning. In *Artificial intelligence and deep learning in pathology*, pages 149–173. Elsevier, 2021.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

Oded Maimon and Lior Rokach. *Data mining and knowledge discovery handbook*. 2005.

Stefano Ceri, Piero Fraternali, Aldo Bongio, Marco Brambilla, Sara Comai, and Maristella Matera. *Morgan Kaufmann series in data management systems: Designing data-intensive Web applications*. Morgan Kaufmann, 2003.

Amanpreet Singh, Narina Thakur, and Aakanksha Sharma. A review of supervised machine learning algorithms. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pages 1310–1315. Ieee, 2016.

Yanli Liu, Yourong Wang, and Jian Zhang. New machine learning algorithm: Random forest. In *Information Computing and Applications: Third International Conference, ICICA 2012, Chengde, China, September 14-16, 2012. Proceedings 3*, pages 246–252. Springer, 2012.

Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

MS Gaya, MU Zango, LA Yusuf, Mamunu Mustapha, Bashir Muhammad, Ashiru Sani,

Aminu Tijjani, NA Wahab, and MTM Khairi. Estimation of turbidity in water treatment plant using hammerstein-wiener and neural network technique. *Indonesian Journal of Electrical Engineering and Computer Science*, 5(3):666–672, 2017.

Yucai Zhu. Estimation of an n – l – n hammerstein–wiener model. *Automatica*, 38(9): 1607–1614, 2002.

SI Abba, Quoc Bao Pham, AG Usman, Nguyen Thi Thuy Linh, DS Aliyu, Quyen Nguyen, and Quang-Vu Bach. Emerging evolutionary algorithm integrated with kernel principal component analysis for modeling the performance of a water treatment plant. *Journal of Water Process Engineering*, 33:101081, 2020.

Adrian Wills, Thomas B Schön, Lennart Ljung, and Brett Ninness. Identification of hammerstein–wiener models. *Automatica*, 49(1):70–81, 2013.

Walid Allafi, Ivan Zajic, Kotub Uddin, and Keith J Burnham. Parameter estimation of the fractional-order hammerstein–wiener model using simplified refined instrumental variable fractional-order continuous time. *IET Control Theory & Applications*, 11(15):2591–2598, 2017.

Claudio Moraga, Enric Trillas, and Sergio Guadarrama. Multiple-valued logic and artificial intelligence fundamentals of fuzzy control revisited. *Artificial intelligence review*, 20: 169–197, 2003.

M Afroozeh, MR Sohrabi, M Davallo, SY Mimezami, F Motlee, and M Khosravi. Application of artificial neural network, fuzzy inference system and adaptive neuro-fuzzy inference system to predict the removal of pb (ii) ions from the aqueous solution by using magnetic graphene/nylon 6. *Chem Sci J*, 9(2):1–7, 2018.

Taesup Moon, Yejin Kim, Hyosu Kim, Myungwon Choi, and Changwon Kim. Fuzzy rule-based inference of reasons for high effluent quality in municipal wastewater treatment plant. *Korean Journal of Chemical Engineering*, 28:817–824, 2011.

Okay Kaynak, Lotfi A Zadeh, Burhan Türksen, and Imre J Rudas. *Computational intelligence: Soft computing and fuzzy-neuro integration with applications*, volume 162. Springer Science & Business Media, 1998.

Lotfi A Zadeh. Roles of soft computing and fuzzy logic in the conception, design and deployment of information/intelligent systems. In *Computational intelligence: soft computing and fuzzy-neuro integration with applications*, pages 1–9. Springer, 1998.

Amir Ali Shahmansouri, Maziar Yazdani, Saeed Ghanbari, Habib Akbarzadeh Bengar, Abouzar Jafari, and Hamid Farrokh Ghatte. Artificial neural network model to predict the compressive strength of eco-friendly geopolymers incorporating silica fume and natural zeolite. *Journal of Cleaner Production*, 279:123697, 2021.

Phil Kim and Phil Kim. Convolutional neural network. *MATLAB deep learning: with machine learning, neural networks and artificial intelligence*, pages 121–147, 2017.

U Rajendra Acharya, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, Muhammad Adam, Arkadiusz Gertych, and Ru San Tan. A deep convolutional neural network model to classify heartbeats. *Computers in biology and medicine*, 89:389–396, 2017.

S Kevin Zhou, Daniel Rueckert, and Gabor Fichtinger. *Handbook of medical image computing and computer assisted intervention*. Academic Press, 2019.

Hongce Zhang, Maxwell Shinn, Aarti Gupta, Arie Gurfinkel, Nham Le, and Nina Narodytska. Verification of recurrent neural networks for cognitive tasks via reachability analysis. In *ECAI 2020*, pages 1690–1697. IOS Press, 2020.

Kamilya Smagulova and Alex Pappachen James. Overview of long short-term memory neural networks. *Deep Learning Classifiers with Memristive Networks: Theory and Applications*, pages 139–153, 2020.

Jitendra Agrawal and Tom V Mathew. Transit route network design using parallel genetic algorithm. *Journal of Computing in Civil Engineering*, 18(3):248–256, 2004.

Xin-She Yang. Nature-inspired optimization algorithms: Challenges and open problems. *Journal of Computational Science*, 46:101104, 2020.

Sourabh Katoch, Sumit Singh Chauhan, and Vijay Kumar. A review on genetic algorithm: past, present, and future. *Multimedia Tools and Applications*, 80:8091–8126, 2021.

N Karimi, S Kazem, D Ahmadian, H Adibi, and LV Ballestra. On a generalized gaussian radial basis function: Analysis and applications. *Engineering analysis with boundary elements*, 112:46–57, 2020.

Michael James David Powell et al. *Approximation theory and methods*. Cambridge university press, 1981.

Kamel Baddari, Tahar Aïfa, Nouredine Djarfour, and Jalal Ferahtia. Application of a radial basis function artificial neural network to seismic data inversion. *Computers & geosciences*, 35(12):2338–2344, 2009.

J Farhoudi, SM Hosseini, and M Sedghi-Asl. Application of neuro-fuzzy model to estimate the characteristics of local scour downstream of stilling basins. *Journal of hydroinformatics*, 12(2):201–211, 2010.

Dervis Karaboga and Ebubekir Kaya. Adaptive network based fuzzy inference system (anfis) training approaches: a comprehensive survey. *Artificial Intelligence Review*, 52:2263–2293, 2019.

PA Adedeji, SO Masebinu, SA Akinlabi, and N Madushele. Adaptive neuro-fuzzy inference system (anfis) modelling in energy system and water resources. In *Optimization Using Evolutionary Algorithms and Metaheuristics*, pages 117–133. CRC Press, 2019.

Qin-Yu Zhu, A Kai Qin, Ponnuthurai N Suganthan, and Guang-Bin Huang. Evolutionary

extreme learning machine. *Pattern recognition*, 38(10):1759–1763, 2005.

Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. Extreme learning machine: a new learning scheme of feedforward neural networks. In *2004 IEEE international joint conference on neural networks (IEEE Cat. No. 04CH37541)*, volume 2, pages 985–990. Ieee, 2004b.

Konstantinos Demertzis, Lazaros Iliadis, Elias Pimenidis, and Panagiotis Kikiras. Variational restricted boltzmann machines to automated anomaly detection. *Neural Computing and Applications*, 34(18):15207–15220, 2022.

Fouzi Harrou, Abdelkader Dairi, Ying Sun, and Mohamed Senouci. Statistical monitoring of a wastewater treatment plant: A case study. *Journal of environmental management*, 223:807–814, 2018.

HY Li, H Osman, CW Kang, and T Ba. Numerical and experimental investigation of uv disinfection for water treatment. *Applied Thermal Engineering*, 111:280–291, 2017.

Chen Xu, GP Rangaiah, and XS Zhao. A computational study of the effect of lamp arrangements on the performance of ultraviolet water disinfection reactors. *Chemical Engineering Science*, 122:299–306, 2015.

Chen Xu, XS Zhao, and GP Rangaiah. Performance analysis of ultraviolet water disinfection reactors using computational fluid dynamics simulation. *Chemical engineering journal*, 221:398–406, 2013.

David L Sedlak and Urs von Gunten. The chlorine dilemma. *Science*, 331(6013):42–43, 2011.

Richard J Bull, LINDA BIRNBAUM, Kenneth P Cantor, Joan B Rose, Byron E Butterworth, REX Pegram, and Juoko Tuomisto. Water chlorination: essential process or cancer hazard? *Toxicological Sciences*, 28(2):155–166, 1995.

André Felipe Librantz, Fábio Cosme Rodrigues dos Santos, and Cleber Gustavo Dias. Artificial neural networks to control chlorine dosing in a water treatment plant. *Acta Scientiarum. Technology*, 40:e37275–e37275, 2018.

Lluís Godo-Pla, Jose Javier Rodríguez, Jordi Suquet, Pere Emiliano, Fernando Valero, Manel Poch, and Hèctor Monclús. Control of primary disinfection in a drinking water treatment plant based on a fuzzy inference system. *Process Safety and Environmental Protection*, 145:63–70, 2021.

Kunwar P Singh and Shikha Gupta. Artificial intelligence based modeling for predicting the disinfection by-products in water. *Chemometrics and Intelligent Laboratory Systems*, 114: 122–131, 2012.

JK Mahato and SK Gupta. Exploring applicability of artificial intelligence and multivariate linear regression model for prediction of trihalomethanes in drinking water. *International Journal of Environmental Science and Technology*, 19(6):5275–5288, 2022.

Jongkwan Park, Chan Ho Lee, Kyung Hwa Cho, Seongho Hong, Young Mo Kim, and Yongeun Park. Modeling trihalomethanes concentrations in water treatment plants using machine learning techniques. *Desalination Water Treat*, 111:125–133, 2018.

Hongjun Lin, Qunyun Dai, Lili Zheng, Huachang Hong, Wenjing Deng, and Fuyong Wu. Radial basis function artificial neural network able to accurately predict disinfection by-product levels in tap water: Taking haloacetic acids as a case study. *Chemosphere*, 248:125999, 2020.

Zeqiong Xu, Jiao Shen, Yuqing Qu, Huangfei Chen, Xiaoling Zhou, Huachang Hong, Hongjie Sun, Hongjun Lin, Wenjing Deng, and Fuyong Wu. Using simple and easy water quality parameters to predict trihalomethane occurrence in tap water. *Chemosphere*, 286:131586, 2022.

Nicolás M Peleato. Application of convolutional neural networks for prediction of disinfection by-products. *Scientific Reports*, 12(1):612, 2022.

Comfort N Okoji, Anthony I Okoji, Musa S Ibrahim, and Okpoko Obinna. Comparative analysis of adaptive neuro-fuzzy inference system (anfis) and rsrm models to predict dbp (trihalomethanes) levels in the water treatment plant. *Arabian Journal of Chemistry*, 15 (6):103794, 2022.

José Andrés Cordero, Kai He, Kanjira Janya, Shinya Echigo, and Sadahiko Itoh. Predicting formation of haloacetic acids by chlorination of organic compounds using machine-learning-assisted quantitative structure-activity relationships. *Journal of Hazardous Materials*, 408:124466, 2021.

Felix Wortmann and Kristina Flüchter. Internet of things: technology and value added. *Business & Information Systems Engineering*, 57:221–224, 2015.

TS Imo, T Oomori, M Toshihiko, and F Tamaki. The comparative study of trihalomethanes in drinking water. *International Journal of Environmental Science & Technology*, 4: 421–426, 2007.

Huachang Hong, Zhiying Zhang, Aidi Guo, Liguang Shen, Hongjie Sun, Yan Liang, Fuyong Wu, and Hongjun Lin. Radial basis function artificial neural network (rbf ann) as well as the hybrid method of rbf ann and grey relational analysis able to well predict trihalomethanes levels in tap water. *Journal of Hydrology*, 591:125574, 2020.

Lauren E Bergman, Jessica M Wilson, Mitchell J Small, and Jeanne M VanBriesen. Application of classification trees for predicting disinfection by-product formation targets from source water characteristics. *Environmental Engineering Science*, 33(7):455–470, 2016.

Rabbi Sikder, Tianyu Zhang, and Tao Ye. Predicting thm formation and revealing its contributors in drinking water treatment using machine learning. *ACS ES&T Water*, 2023.

Haroon R Mian, Guangji Hu, Kasun Hewage, Manuel J Rodriguez, and Rehan Sadiq.

Predicting unregulated disinfection by-products in water distribution networks using generalized regression neural networks. *Urban Water Journal*, 18(9):711–724, 2021.

Guangji Hu, Haroon R Mian, Saeed Mohammadiun, Manuel J Rodriguez, Kasun Hewage, and Rehan Sadiq. Appraisal of machine learning techniques for predicting emerging disinfection byproducts in small water distribution networks. *Journal of Hazardous Materials*, 446:130633, 2023.

Rama Rao Karri, JN Sahu, and BC Meikap. Improving efficacy of cr (vi) adsorption process on sustainable adsorbent derived from waste biomass (sugarcane bagasse) with help of ant colony optimization. *Industrial Crops and Products*, 143:111927, 2020.

Ramesh Vinayagam, Niyam Dave, Thivaharan Varadavenkatesan, Natarajan Rajamohan, Mika Sillanpää, Ashok Kumar Nadda, Muthusamy Govarthan, and Raja Selvaraj. Artificial neural network and statistical modelling of biosorptive removal of hexavalent chromium using macroalgal spent biomass. *Chemosphere*, 296:133965, 2022.

Suraj Kumar Bhagat, Konstantina Pyrgaki, Sinan Q Salih, Tiyyasha Tiyyasha, Ufuk Beyaztas, Shamsuddin Shahid, and Zaher Mundher Yaseen. Prediction of copper ions adsorption by attapulgite adsorbent using tuned-artificial intelligence model. *Chemosphere*, 276:130162, 2021.

Mohammad Sadegh Mazloom, Farzaneh Rezaei, Abdolhossein Hemmati-Sarapardeh, Maen M Husein, Sohrab Zendehboudi, and Amin Bemani. Artificial intelligence based methods for asphaltene adsorption by nanocomposites: Application of group method of data handling, least squares support vector machine, and artificial neural networks. *Nanomaterials*, 10(5):890, 2020.

Yamin Mesellem, Abdallah Abdallah El Hadj, Maamar Laidi, Salah Hanini, and Mohamed Hentabli. Computational intelligence techniques for modeling of dynamic adsorption of organic pollutants on activated carbon. *Neural Computing and Applications*, 33: 12493–12512, 2021a.

Mohammed Al-Yaari, Theyazn HH Aldhyani, and Sayeed Rushd. Prediction of arsenic removal from contaminated water using artificial neural network model. *Applied Sciences*, 12(3):999, 2022.

H Mazaheri, M Ghaedi, MH Ahmadi Azqhandi, and AJPCCP Asfaram. Application of machine/statistical learning, artificial intelligence and statistical experimental design for the modeling and optimization of methylene blue and cd (ii) removal from a binary aqueous solution by natural walnut carbon. *Physical Chemistry Chemical Physics*, 19 (18):11299–11317, 2017.

Zaki Uddin Ahmad, Lunguang Yao, Qiyu Lian, Fahrin Islam, Mark E Zappi, and Daniel Dianchen Gang. The use of artificial neural network (ann) for modeling adsorption of sunset yellow onto neodymium modified ordered mesoporous carbon. *Chemosphere*, 256:127081, 2020.

Manal Fawzy, Mahmoud Nasr, Samar Adel, Heba Nagy, and Shacker Helmi. Environmental approach and artificial intelligence for ni (ii) and cd (ii) biosorption from aqueous solution using typha domingensis biomass. *Ecological Engineering*, 95:743–752, 2016.

Sami Ullah, Mohammed Ali Assiri, Mohamad Azmi Bustam, Abdullah G Al-Sehemi, Firas A Abdul Kareem, and Ahmad Irfan. Equilibrium, kinetics and artificial intelligence characteristic analysis for zn (ii) ion adsorption on rice husks digested with nitric acid. *Paddy and Water Environment*, 18:455–468, 2020.

Ahmed S Mahmoud, Mohamed K Mostafa, and Mahmoud Nasr. Regression model, artificial intelligence, and cost estimation for phosphate adsorption using encapsulated nanoscale zero-valent iron. *Separation Science and Technology*, 54(1):13–26, 2019.

Yamin Mesellem, Abdallah El Hadj Abdallah, Maamar Laidi, Salah Hanini, and Mohamed Hentabli. Artificial neural network modelling of multi-system dynamic adsorption of organic pollutants on activated carbon. *Kemija u industriji: Časopis kemičara i kemijskih inženjera Hrvatske*, 70(1-2):1–12, 2021b.

Majid Mohammadi, Mehdi Safari, Mostafa Ghasemi, Amin Daryasafar, and Mehdi Sedighi. Asphaltene adsorption using green nanocomposites: Experimental study and adaptive neuro-fuzzy interference system modeling. *Journal of Petroleum Science and Engineering*, 177:1103–1113, 2019.

AK Maurya, M Nagamani, Seung Won Kang, Jong-Taek Yeom, Jae-Keun Hong, Hyokyung Sung, CH Park, Paturi Uma Maheshwera Reddy, and NS Reddy. Development of artificial neural networks software for arsenic adsorption from an aqueous environment. *Environmental Research*, 203:111846, 2022.

Jingxin Liu, Zelin Xu, and Wenjuan Zhang. Unraveling the role of fe in as (iii & v) removal by biochar via machine learning exploration. *Separation and Purification Technology*, page 123245, 2023.

M Ghaedi, N Zeinali, AM Ghaedi, M Teimuori, and J Tashkhourian. Artificial neural network-genetic algorithm based optimization for the adsorption of methylene blue and brilliant green from aqueous solution by graphite oxide nanoparticle. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 125:264–277, 2014.

Kunrong Zeng, Kadda Hachem, Mariya Kuznetsova, Supat Chupradit, Chia-Hung Su, Hoang Chinh Nguyen, and AS El-Shafay. Molecular dynamic simulation and artificial intelligence of lead ions removal from aqueous solution using magnetic-ash-graphene oxide nanocomposite. *Journal of Molecular Liquids*, 347:118290, 2022.

Yajun Wei, Jing Yu, Yonglin Du, Hongxu Li, and Chia-Hung Su. Artificial intelligence simulation of pb (ii) and cd (ii) adsorption using a novel metal organic framework-based nanocomposite adsorbent. *Journal of Molecular Liquids*, 343:117681, 2021.

S Mandal, SS Mahapatra, and RK Patel. Neuro fuzzy approach for arsenic (iii) and

chromium (vi) removal from water. *Journal of Water Process Engineering*, 5:58–75, 2015a.

S Mandal, SS Mahapatra, MK Sahu, and RK Patel. Artificial neural network modelling of as (iii) removal from water by novel hybrid material. *Process Safety and Environmental Protection*, 93:249–264, 2015b.

S Mandal, SS Mahapatra, and RK Patel. Enhanced removal of cr (vi) by cerium oxide polyaniline composite: optimization and modeling approach using response surface methodology and artificial neural networks. *Journal of Environmental Chemical Engineering*, 3(2):870–885, 2015c.

Selina Hube, Majid Eskafi, Kolbrún Fríða Hrafnkelsdóttir, Björg Bjarnadóttir, Margrét Ásta Bjarnadóttir, Snærós Axelsdóttir, and Bing Wu. Direct membrane filtration for wastewater treatment and resource recovery: A review. *Science of the total environment*, 710:136375, 2020.

Wouter Pronk, An Ding, Eberhard Morgenroth, Nicolas Derlon, Peter Desmond, Michael Burkhardt, Bing Wu, and Anthony G Fane. Gravity-driven membrane filtration for water and wastewater treatment: a review. *Water research*, 149:553–565, 2019.

Mohamed Zoubek, Amgad Salama, and Amr Henni. A comprehensive experimental and artificial network investigation of the performance of an ultrafiltration titanium dioxide ceramic membrane: application in produced water treatment. *Water and Environment Journal*, 33(3):459–475, 2019.

Masoud Fetanat, Mohammadali Keshtiara, Ze-Xian Low, Ramazan Keyikoglu, Alireza Khataee, Yasin Orooji, Vicki Chen, Gregory Leslie, and Amir Razmjou. Machine learning for advanced design of nanocomposite ultrafiltration membranes. *Industrial & Engineering Chemistry Research*, 60(14):5236–5250, 2021.

Hammad Khan, Saad Ullah Khan, Sajjad Hussain, and Asmat Ullah. Modelling of transmembrane pressure using slot/pore blocking model, response surface and artificial intelligence approach. *Chemosphere*, 290:133313, 2022.

Zakariah Yusof, Norhaliza Abdul Wahab, Syahira Ibrahim, Shafishuhaza Sahlan, and Mashitah Che Razali. Modeling of submerged membrane filtration processes using recurrent artificial neural networks. *IAES International Journal of Artificial Intelligence*, 9(1):155, 2020.

Sara Nazif, Emad Mirashrafi, Bardia Roghani, and Gholamreza Nabi Bidhendi. Artificial intelligence-based optimization of reverse osmosis systems operation performance. *Journal of Environmental Engineering*, 146(2):04019106, 2020.

Jaegy Shim, Sanghun Park, and Kyung Hwa Cho. Deep learning model for simulating influence of natural organic matter in nanofiltration. *Water Research*, 197:117070, 2021.

Yamina Ammi, Salah Hanini, and Latifa Khaouane. An artificial intelligence approach for modeling the rejection of anti-inflammatory drugs by nanofiltration and reverse osmosis membranes using kernel support vector machine and neural networks. *Comptes Rendus*.

Chimie, 24(2):243–254, 2021a.

V Yangali-Quintanilla, A Verliefde, T-U Kim, A Sadmani, M Kennedy, and G Amy. Artificial neural network models based on qsar for predicting rejection of neutral organic compounds by polyamide nanofiltration and reverse osmosis membranes. *Journal of membrane science*, 342(1-2):251–262, 2009.

Mohamed Zoubeik, Mohamed Echakouri, Amr Henni, and Amgad Salama. Taguchi optimization of operating conditions of a microfiltration alumina ceramic membrane and artificial neural-network modeling. *Journal of Environmental Engineering*, 148(4): 04022001, 2022.

Samira Arefi-Oskoui, Alireza Khataee, and Vahid Vatanpour. Modeling and optimization of nldh/pvdf ultrafiltration nanocomposite membrane using artificial neural network-genetic algorithm hybrid. *ACS Combinatorial Science*, 19(7):464–477, 2017.

Nurazizah Mahmud, Norhaliza Abdul Wahab, and Muhammad Sani Gaya. Modelling and control of fouling in submerged membrane bioreactor using neural network internal model control. *IAES International Journal of Artificial Intelligence*, 9(1):100, 2020.

Çağla Odabaşı, Pelin Dologlu, Fatih Gülmez, Gizem Kuşoğlu, and Ömer Çağlar. Investigation of the factors affecting reverse osmosis membrane performance using machine-learning techniques. *Computers & Chemical Engineering*, 159:107669, 2022.

Chen Wang, Li Wang, Allan Soo, Nirenkumar Bansidhar Pathak, and Ho Kyong Shon. Machine learning based prediction and optimization of thin film nanocomposite membranes for organic solvent nanofiltration. *Separation and Purification Technology*, 304:122328, 2023.

Sung Ju Im, Viet Duc Nguyen, and Am Jang. Prediction of forward osmosis membrane engineering factors using artificial intelligence approach. *Journal of Environmental Management*, 318:115544, 2022.

Yamina Ammi, Latifa Khaouane, and Salah Hanini. Stacked neural networks for predicting the membranes performance by treating the pharmaceutical active compounds. *Neural Computing and Applications*, pages 1–16, 2021b.

Latifa Khaouane, Yamina Ammi, and Salah Hanini. Modeling the retention of organic compounds by nanofiltration and reverse osmosis membranes using bootstrap aggregated neural networks. *Arabian Journal for Science and Engineering*, 42:1443–1453, 2017.

Yamina Ammi, Latifa Khaouane, and Salah Hanini. Prediction of the rejection of organic compounds (neutral and ionic) by nanofiltration and reverse osmosis membranes using neural networks. *Korean Journal of Chemical Engineering*, 32:2300–2310, 2015.

Yamina Ammi, Latifa Khaouane, and Salah Hanini. A model based on bootstrapped neural networks for modeling the removal of organic compounds by nanofiltration and reverse osmosis membranes. *Arabian Journal for Science and Engineering*, 43:6271–6284, 2018.

Sangsuk Lee and Jooho Kim. Prediction of nanofiltration and reverse-osmosis-membrane rejection of organic compounds using random forest model. *Journal of Environmental Engineering*, 146(11):04020127, 2020.

Robert D Deegan, Olgica Bakajin, Todd F Dupont, Greb Huber, Sidney R Nagel, and Thomas A Witten. Capillary flow as the cause of ring stains from dried liquid drops. *Nature*, 389(6653):827–829, 1997.

Yanan Li, Qiang Yang, Mingzhu Li, and Yanlin Song. Rate-dependent interface capture beyond the coffee-ring effect. *Scientific reports*, 6(1):1–8, 2016c.

Noushine Shahidzadeh, Marthe FL Schut, Julie Desarnaud, Marc Prat, and Daniel Bonn. Salt stains from evaporating droplets. *Scientific reports*, 5(1):1–9, 2015.

Yasunari Matsuzaka and Yoshihiro Uesawa. Optimization of a deep-learning method based on the classification of images generated by parameterized deep snap a novel molecular-image-input technique for quantitative structure–activity relationship (qsar) analysis. *Frontiers in bioengineering and biotechnology*, page 65, 2019.

Jesse G Meyer, Shengchao Liu, Ian J Miller, Joshua J Coon, and Anthony Gitter. Learning drug functions from chemical structures with convolutional neural networks and random forests. *Journal of chemical information and modeling*, 59(10):4438–4449, 2019.

Félix Lussier, Dimitris Missirlis, Joachim P Spatz, and Jean-François Masson. Machine-learning-driven surface-enhanced raman scattering optophysiology reveals multiplexed metabolite gradients near cells. *ACS nano*, 13(2):1403–1411, 2019.

William John Thrift and Regina Ragan. Quantification of analyte concentration in the single molecule regime using convolutional neural networks. *Analytical chemistry*, 91(21): 13337–13342, 2019.

Ling Liu and M Tamer Özsu. *Encyclopedia of database systems*, volume 6. Springer, 2009. Dongmao Zhang, Yong Xie, Melissa F Mrozek, Corasi Ortiz, V Jo Davisson, and Dor Ben-Amotz. Raman detection of proteomic analytes. *Analytical chemistry*, 75(21): 5703–5709, 2003.

Corasi Ortiz, Dongmao Zhang, Yong Xie, V Jo Davisson, and Dor Ben-Amotz. Identification of insulin variants using raman spectroscopy. *Analytical Biochemistry*, 332(2):245–252, 2004.

Viral H Chhasatia, Abhijit S Joshi, and Ying Sun. Effect of relative humidity on contact angle and particle deposition morphology of an evaporating colloidal drop. *Applied Physics Letters*, 97(23):231909, 2010.

Corasi Ortiz, Dongmao Zhang, Yong Xie, Alexander E Ribbe, and Dor Ben-Amotz. Validation of the drop coating deposition raman method for protein analysis. *Analytical biochemistry*, 353(2):157–166, 2006.

Vijayakumar Kadappa and Atul Negi. A theoretical investigation of feature partitioning principal component analysis methods. *Pattern Analysis and Applications*, 19(1):79–91, 2016.

Mark M Benjamin. *Water chemistry*. Waveland Press, 2014.

William M Haynes, David R Lide, and Thomas J Bruno. *CRC handbook of chemistry and physics*. CRC press, 2016.

Enric Junqué de Fortuny, David Martens, and Foster Provost. Predictive modeling with big data: is bigger really better? *Big data*, 1(4):215–226, 2013.

David Martens, Foster Provost, Jessica Clark, and Enric Junqué de Fortuny. Mining massive fine-grained behavior data to improve predictive analytics. *MIS quarterly*, 40(4):869–888, 2016.

Nathalie Japkowicz and Shaju Stephen. The class imbalance problem: A systematic study. *Intelligent data analysis*, 6(5):429–449, 2002.

Bartosz Krawczyk. Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, 5(4):221–232, 2016.

Steven J Burian, Stephan J Nix, Robert E Pitt, and S Rocky Durrans. Urban wastewater management in the united states: Past, present, and future. *Journal of Urban Technology*, 7(3):33–62, 2000.

Martin R Coghill, Gary A Eaton, and Nathan D Faber. A long-term commitment to pipeline infrastructure: Implementing, funding, and delivering the san diego county water authority’s asset management program. In *Pipelines 2014: From Underground to the Forefront of Innovation and Sustainability*, pages 1187–1197. 2014.

Steven Folkman. Water main break rates in the usa and canada: A comprehensive study. 2018.

PE Darlene Garcia and PE Susan Funchion. How to select and prioritize water main replacement. *Opflow*, 41(10):10–14, 2015.

David A Cornwell, Richard A Brown, and Steve H Via. National survey of lead service line occurrence. *Journal-American Water Works Association*, 108(4):E182–E191, 2016.

Orazio Giustolisi and Luigi Berardi. Prioritizing pipe replacement: From multiobjective genetic algorithms to operational decision support. *Journal of Water Resources Planning and Management*, 135(6):484–492, 2009.

Peter D Rogers and Neil S Grigg. Failure assessment model to prioritize pipe replacement in water utility asset management. In *Water Distribution Systems Analysis Symposium 2006*, pages 1–17, 2008.

Y Tlili and A Nafi. A practical decision scheme for the prioritization of water pipe replacement. *Water Science and Technology: Water Supply*, 12(6):895–917, 2012.

Go Bong Choi, Jong Woo Kim, Jung Chul Suh, Kwang Ho Jang, and Jong Min Lee. A prioritization method for replacement of water mains using rank aggregation. *Korean Journal of Chemical Engineering*, 34(10):2584–2590, 2017.

Mohamed Marzouk, Said Abdel Hamid, and Moheeb El-Said. A methodology for prioritizing water mains rehabilitation in egypt. *HBRC Journal*, 11(1):114–128, 2015.

Cheng-I Ho, Min-Der Lin, and Shang-Lien Lo. Prioritizing pipe replacement in a water distribution system using a seismic-based artificial neural network model. *Environmental engineering science*, 26(4):745–752, 2009.

Gregory J Kirmeyer. *Guidance manual for monitoring distribution system water quality*. American Water Works Association, 2002.

Mohsin J Qazi, Rinse W Liefferink, Simon J Schlegel, Ellen HG Backus, Daniel Bonn, and Noushine Shahidzadeh. Influence of surfactants on sodium chloride crystallization in confinement. *Langmuir*, 33(17):4260–4268, 2017.

Xiaoxiao Wei, Jian Yang, Zhiyong Li, Yunlan Su, and Dujin Wang. Comparison investigation of the effects of ionic surfactants on the crystallization behavior of calcium oxalate: From cationic to anionic surfactant. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 401:107–115, 2012.

Maria Sammalkorpi, Mikko Karttunen, and Mikko Haataja. Ionic surfactant aggregates in saline solutions: sodium dodecyl sulfate (sds) in the presence of excess sodium chloride (nacl) or calcium chloride (cacl₂). *The Journal of Physical Chemistry B*, 113(17): 5863–5870, 2009.

Julie Desarnaud, Hannelore Derluyn, Jan Carmeliet, Daniel Bonn, and Noushine Shahidzadeh. Metastability limit for the nucleation of nacl crystals in confinement. *The journal of physical chemistry letters*, 5(5):890–895, 2014.

Fiona C Meldrum and Cedrick O’Shaughnessy. Crystallization in confinement. *Advanced Materials*, 32(31):2001068, 2020.

Xin Zhong, Alexandru Crivoi, and Fei Duan. Sessile nanofluid droplet drying. *Advances in colloid and interface science*, 217:13–30, 2015.

Huicheng Feng, Karen Siew-Ling Chong, Kian-Soo Ong, and Fei Duan. Octagon to square wetting area transition of water–ethanol droplets on a micropylramid substrate by increasing ethanol concentration. *Langmuir*, 33(5):1147–1154, 2017.

Xin Zhong and Fei Duan. Flow regime and deposition pattern of evaporating binary mixture droplet suspended with particles. *The European Physical Journal E*, 39(2):1–6, 2016.

Manos Anyfantakis, Zheng Geng, Mathieu Morel, Sergii Rudiuk, and Damien Baigl. Modulation of the coffee-ring effect in particle/surfactant mixtures: the importance of particle–interface interactions. *Langmuir*, 31(14):4113–4120, 2015.

Bo Zhang, Xuemei Chen, Jure Dobnikar, Zuankai Wang, and Xianren Zhang. Spontaneous wenzel to cassie dewetting transition on structured surfaces. *Physical review fluids*, 1(7): 073904, 2016.

Chenglong Xu, Shuhua Peng, Greg Qiao, and Xuehua Zhang. Effects of the molecular structure of a self-assembled monolayer on the formation and morphology of surface nanodroplets. *Langmuir*, 32(43):11197–11202, 2016.

Leila Bahmani, Mahdi Neysari, and Maniya Maleki. The study of drying and pattern formation of whole human blood drops and the effect of thalassaemia and neonatal jaundice on the patterns. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 513: 66–75, 2017.

Hau Him Lee, Sau Chung Fu, Chi Yan Tso, and Christopher YH Chao. Study of residue patterns of aqueous nanofluid droplets with different particle sizes and concentrations on different substrates. *International Journal of Heat and Mass Transfer*, 105:230–236, 2017.

Nainsi Saxena, Tapaswinee Naik, and Santanu Paria. Organization of sio₂ and tio₂ nanoparticles into fractal patterns on glass surface for the generation of superhydrophilicity. *The Journal of Physical Chemistry C*, 121(4):2428–2436, 2017.

Hui Li, Hao Luo, Zhen Zhang, Yongjun Li, Bin Xiong, Chunyan Qiao, Xuan Cao, Tie Wang, Yan He, and Guangyin Jing. Direct observation of nanoparticle multiple-ring pattern formation during droplet evaporation with dark-field microscopy. *Physical Chemistry Chemical Physics*, 18(18):13018–13025, 2016d.

Xuemei Chen, Ruiyuan Ma, Jintao Li, Chonglei Hao, Wei Guo, Bing Lam Luk, Shuai Cheng Li, Shuhuai Yao, and Zuankai Wang. Evaporation of droplets on superhydrophobic surfaces: Surface roughness and small droplet size effects. *Physical review letters*, 109 (11):116101, 2012.

Niranjan A Malvadkar, Matthew J Hancock, Koray Sekeroglu, Walter J Dressick, and Melik C Demirel. An engineered anisotropic nanofilm with unidirectional wetting properties. *Nature materials*, 9(12):1023–1028, 2010.

Eleanor R Townsend, Willem JP van Enckevort, Jan AM Meijer, and Elias Vlieg. Additive enhanced creeping of sodium chloride crystals. *Crystal Growth & Design*, 17(6):3107–3115, 2017.

Subra Suresh. Colloid model for atoms. *Nature materials*, 5(4):253–254, 2006.

William G Walter. Standard methods for the examination of water and wastewater, 1961.

Carlos Rodriguez-Navarro and Eric Doehne. Salt weathering: influence of evaporation rate, supersaturation and crystallization pattern. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group*, 24(3):191–209, 1999.

Alvaro G Marin, Hanneke Gelderblom, Detlef Lohse, and Jacco H Snoeijer. Order-to-disorder transition in ring-shaped colloidal stains. *Physical review letters*, 107(8):085502, 2011.

Marti J Anderson. Permutational multivariate analysis of variance (permanova). *Wiley statsref: statistics reference online*, pages 1–15, 2014.

Ralph G O’Brien and Mary K Kaiser. Manova method for analyzing repeated measures designs: an extensive primer. *Psychological bulletin*, 97(2):316, 1985.

Adery CA Hope. A simplified monte carlo significance test procedure. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(3):582–598, 1968.

Frank Nielsen. On a variational definition for the jensen-shannon symmetrization of distances based on the information radius. *Entropy*, 23(4):464, 2021.

Christopher Manning and Hinrich Schutze. *Foundations of statistical natural language processing*. MIT press, 1999.

Ido Dagan, Lillian Lee, and Fernando Pereira. Similarity-based methods for word sense disambiguation. *arXiv preprint cmp-lg/9708010*, 1997.

Dominik Maria Endres and Johannes E Schindelin. A new metric for probability distributions. *IEEE Transactions on Information theory*, 49(7):1858–1860, 2003.

Ferdinand Osterreicher and Igor Vajda. A new class of metric divergences on probability spaces and its applicability in statistics. *Annals of the Institute of Statistical Mathematics*, 55(3):639–653, 2003.

Bent Fuglede and Flemming Topsøe. Jensen-shannon divergence and hilbert space embedding. In *International Symposium on Information Theory, 2004. ISIT 2004. Proceedings.*, page 31. IEEE, 2004.

Joseph B Kruskal and Myron Wish. *Multidimensional scaling*. Sage, 1978.

Andreas Buja, Deborah F Swayne, Michael L Littman, Nathaniel Dean, Heike Hofmann, and Lisha Chen. Data visualization with multidimensional scaling. *Journal of computational and graphical statistics*, 17(2):444–472, 2008.

Alisha Faherty. Tapped out: How newark, new jersey’s lead drinking water crisis illuminates the inadequacy of the federal drinking water regulatory scheme and fuels environmental injustice throughout the nation. *Environmental Claims Journal*, 33(4):304–327, 2021.

Avraham Ebenstein. The consequences of industrialization: evidence from water pollution and digestive cancers in china. *Review of Economics and Statistics*, 94(1):186–201, 2012.

Guilherme Lages Barbosa, Francisca Daiane Almeida Gadelha, Natalya Kublik, Alan Proctor, Lucas Reichelm, Emily Weissinger, Gregory M Wohlleb, and Rolf U Halden. Comparison of land, water, and energy requirements of lettuce grown using hydroponic vs. conventional agricultural methods. *International journal of environmental research and public health*, 12(6):6879–6891, 2015.

Lin Yang, Yuantao Yang, Haodong Lv, Dong Wang, Yiming Li, and Weijun He. Water usage for energy production and supply in china: Decoupled from industrial growth? *Science of the Total Environment*, 719:137278, 2020.

Chi Thanh Vu and Tingting Wu. Recent progress in adsorptive removal of per-and poly-fluoroalkyl substances (pfas) from water/wastewater. *Critical Reviews in Environmental Science and Technology*, 52(1):90–129, 2022.

Aditi Podder, AHM Anwar Sadmani, Debra Reinhart, Ni-Bin Chang, and Ramesh Goel. Per and poly-fluoroalkyl substances (pfas) as a contaminant of emerging concern in surface water: a transboundary review of their occurrences and toxicity effects. *Journal of hazardous materials*, 419:126361, 2021.

Gulzar Alam, Ihsanullah Ihsanullah, Mu Naushad, and Mika Sillanpää. Applications of artificial intelligence in water treatment for optimization and automation of adsorption processes: Recent advances and prospects. *Chemical Engineering Journal*, 427:130011, 2022.

Nawal Taoufik, Wafaa Boumya, Mounia Achak, Hamid Chennouk, Raf Dewil, and Nouredine Barka. The state of art on the prediction of efficiency and modeling of the processes of pollutants removal based on machine learning. *Science of The Total Environment*, 807:150554, 2022.

Ariya Gordanshekan, Shakiba Arabian, Ali Reza Solaimany Nazar, Mehrdad Farhadian, and Shahram Tangestaninejad. A comprehensive comparison of green $\text{bi}_2\text{wo}_6/\text{g-c}_3\text{n}_4$ and $\text{bi}_2\text{wo}_6/\text{tio}_2$ s-scheme heterojunctions for photocatalytic adsorption/degradation of cefixime: Artificial neural network, degradation pathway, and toxicity estimation. *Chemical Engineering Journal*, 451:139067, 2023.

Yifan Xie, Yongqi Chen, Qing Lian, Hailong Yin, Jian Peng, Meng Sheng, and Yimeng Wang. Enhancing real-time prediction of effluent water quality of wastewater treatment plant based on improved feedforward neural network coupled with optimization algorithm. *Water*, 14(7):1053, 2022.

Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. Ieee, 2016.

Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. In *Proceedings of the twentieth annual symposium on Computational geometry*, pages 347–356, 2004.

Tamal K Dey, Facundo Mémoli, and Yusu Wang. Topological analysis of nerves, reeb

spaces, mappers, and multiscale mappers. *arXiv preprint arXiv:1703.07387*, 2017.

Gunnar Carlsson. Topological methods for data modelling. *Nature Reviews Physics*, 2(12): 697–708, 2020.

Ross Geoghegan. *Topological methods in group theory*, volume 243. Springer Science & Business Media, 2007.

APPENDIX

Table 6.1: Measured water chemistry data from tap water samples collected across Michigan and treatment information from annual municipal water quality reports and system operators. Averages and standard deviations are listed for values conducted in replicate.

City	F^- (mM)	NO_3^- (mM)	Zn (mM)	TOC (ppm)
MSU - academic hall	0.04	BD	2.3×10^{-3}	3.1
Durand	0.03	0.02	BD	1.3
Kalamazoo	0.04	0.03	BD	BD
Portland	0.03	BD	BD	2.1
Battle Creek site A	0.05	BD	3.9×10^{-3}	0.79
Battle Creek site B	0.05	0.02	BD	1.2
Charlotte	0.02	0.01	1.1×10^{-4}	1.4
Fowlerville	0.04	BD	BD	1.4
Lansing site A	0.01	0.01	1.2×10^{-4}	1.5
Lansing site B	0.03	0.01	9.1×10^{-4}	1.5
East Lansing	0.02	0.03	6.3×10^{-4}	1.3
Howell	0.03	BD	1.5×10^{-4}	BD
MSU - residence hall	0.05	BD	4.2×10^{-4}	3.2
Williamston	0.03	BD	3.5×10^{-4}	2.2
Genoa Twp soft	BD	BD	1.7×10^{-4}	2.2
Genoa Twp	BD	BD	1.1×10^{-4}	2.0
Rest stop Okemos	0.03	BD	BD	1.1
Rest stop Zeeland	0.04	BD	2.8×10^{-3}	1.0
Rest stop I96/M66	0.03	BD	4.9×10^{-4}	3.3
Rest stop Fenton	0.06	0.02	1.0×10^{-3}	1.0
Allegan	0.03	BD	5.5×10^{-4}	BD
Genoa Twp	BD	BD	1.2×10^{-4}	BD
Detroit	0.03	0.07	3.2×10^{-3}	1.6
Flint	0.04	0.03	7.5×10^{-4}	BD
hline Swartz Creek	0.03	0.03	1.2×10^{-4}	BD
Grand Rapids	0.03	0.03	2.3×10^{-4}	BD
Holland	0.04	0.03	8.9×10^{-4}	BD
Wyoming	0.03	0.03	BD	BD

Table 6.2: Composition of synthetic tap water solutions.

Chemicals (mM)	Detroit	Lansing	MSU hard water
$NaHCO_3$	0.23	0.50	0.55
Na_2SO_4	-	1.20	-
$MgCl_2(H_2O)_6$	0.25	0.53	0.40
$MgSO_4(H_2O)_7$	0.10	-	0.80
$MgCO_3$	-	-	0.50
$CaCl_2$	-	0.56	-
$CaSO_4$	0.16	-	-
$CaCO_3$	0.50	-	2.60
KCl	-	0.100	0.027
KH_2PO_4	0.0152	0.0100	0.0113
$NaNO_3$	0.0725	0.0140	-
$KF (H_2O)_2$	0.0325	0.0270	0.0430
$FeCl_3$	0.0016	-	0.0190
$CuCl_2(H_2O)_2$	0.0006	0.0005	0.0020

Table 6.3: Examples of raw and pre-processed images used for the convolutional neural network (CNN) model.

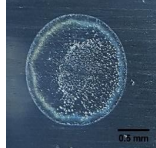
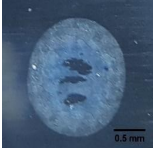
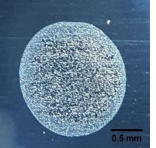
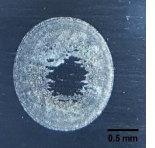
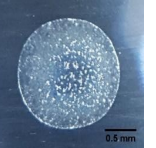
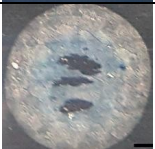
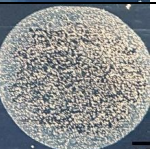
Water sample	Detroit	Genoa Township well RO	Genoa Township private well untreated	Howell	Williamston
Raw image					
Pre-processed image					

Table 6.4: Five replicates of each freshly collected water sample (stored less than one week).
The lab temperature was 24-25 °C and relative humidity 52% for this experiment.

MINIMALLY TREATED GROUNDWATER									
MSU academic hall					Durand				
									
Kalamazoo					Portland				
									
Battle Creek Site					Battle Creek Site B				
									
Charlotte					Fowlerville				
									
LIME SOFTENED									
Lansing Site A					Lansing Site B				
									
East Lansing					Howell				
									
ION EXCHANGE									
MSU residence hall					Williamston				
									
Genoa Township private well softened									
									
UNTREATED GROUNDWATER									
Genoa Township private well untreated					Rest stop A - Okemos				
									
Rest stop C - Zeeland A					Rest stop D - M66/I96 East				
									
REVERSE OSMOSIS									
Allegan					Genoa Township private well RO				
									

Table 6.4: (cont'd)
SURFACE WATER

SURFACE WATER									
Detroit					Flint				
									
Swartz Creek					Grand Rapids				
									
Holland					Wyoming				
									

Table 6.5: Consistency of tap water residue patterns on different mirrored aluminum slides prepared by different researchers, with nanopure water and synthetic hard freshwater controls. The lab temperature was 24 °C and relative humidity 47%.

	Analyst 1 Experienced 1 year			Analyst 2 Moderate Experience 0.5 month			Analyst 3 Lea t Experienced 1 week		
	MSU academic hall								
slide	7	8	9	7	8	9	1	2	3
Replicate 1									
Replicate 2									
Replicate 3									
Blank									
Synthetic									
	East Lansing								
slide	1	2	3	1	2	NA	1	2	3
Replicate 1									
Replicate 2									
Replicate 3									
Blank									
Synthetic									
	Rest Stop (M66)								
slide	4	5	6	4	5	6	1	2	3
Replicate 1									
Replicate 2									
Replicate 3									
Blank									
Synthetic									

Table 6.5: (cont'd)

	Analyst 1 Experienced 1 year			Analyst 2 Moderate Experience 0.5 month			Analyst 3 Least Experienced 1 week		
	Detroit								
slide	1	2	3	1	2	NA	1	2	3
Replicate 1									
Replicate 2									
Replicate 3									
Blank									
Synthetic									
	Grand Rapids								
slide	4	5	6	4	5	6	1	2	3
Replicate 1									
Replicate 2									
Replicate 3									
Blank									
Synthetic									

Table 6.6: Nanochromatography patterns of Michigan tap waters (stored for two months at 4°C) dried on slides cut from the same sheet of aluminum. Nanopure water synthetic hard water served as controls. The lab temperature was 24 °C and relative humidity was 47-48% for this experiment.

Minimally treated groundwater						
MSU academic building					nanopure	Synthetic
Durand						

Table 6.6: (cont'd)

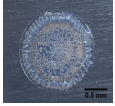
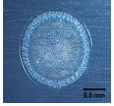
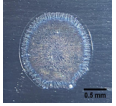
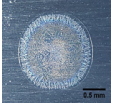
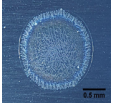
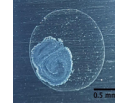
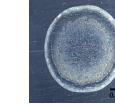
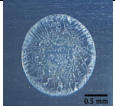
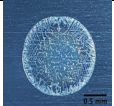
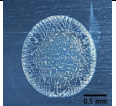
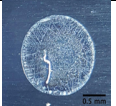
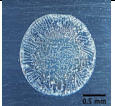
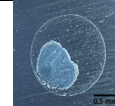
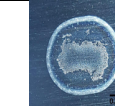
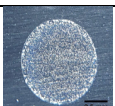
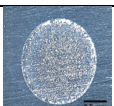
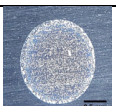
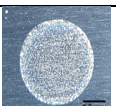
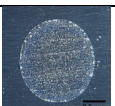
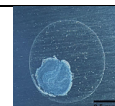
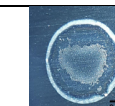
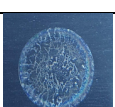
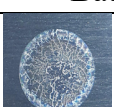
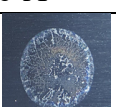
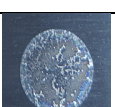
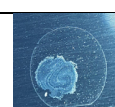
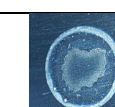
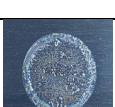
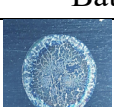
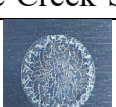
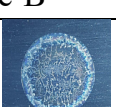
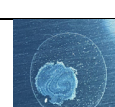
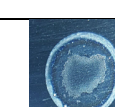
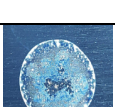
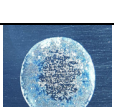
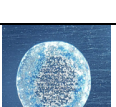
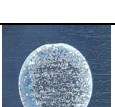
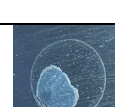
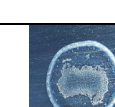
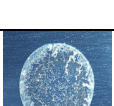
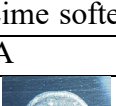
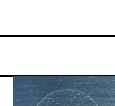
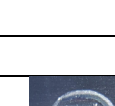
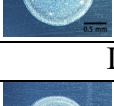
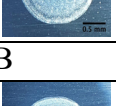
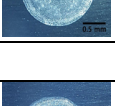
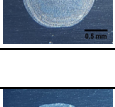
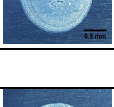
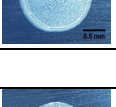
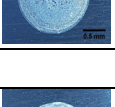
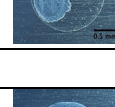
						
Kalamazoo						
						
Portland						
						
Battle Creek Site A						
						
Battle Creek Site B						
						
Fowlerville						
						
Charlotte						
						
Lime softened						
Lansing Site A						
						
Lansing Site B						
						
Howell						
						
East Lansing						

Table 6.6: (cont'd)

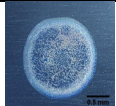
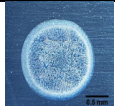
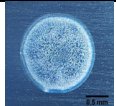
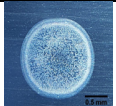
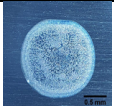
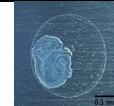
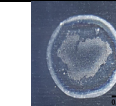
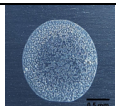
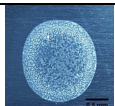
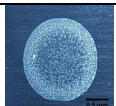
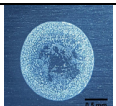
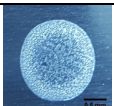
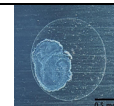
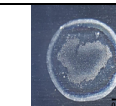
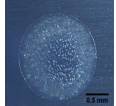
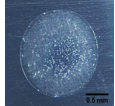
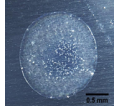
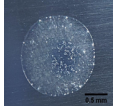
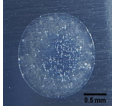
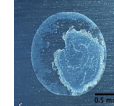
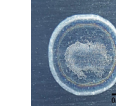
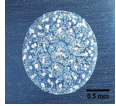
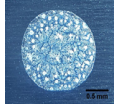
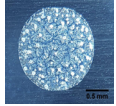
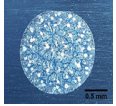
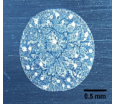
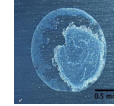
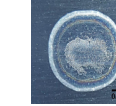
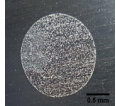
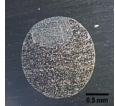
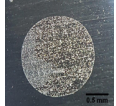
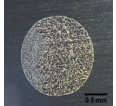
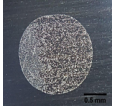
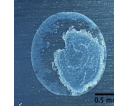
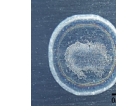
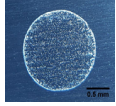
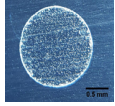
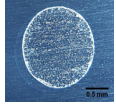
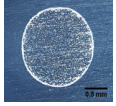
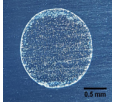
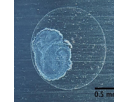
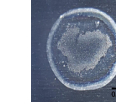
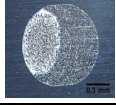
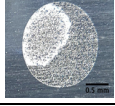
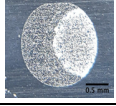
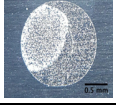
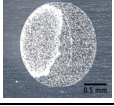
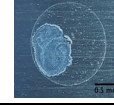
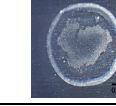
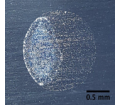
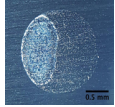
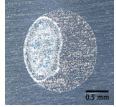
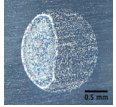
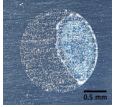
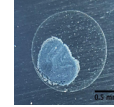
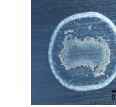
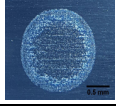
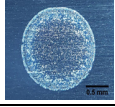
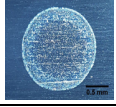
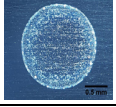
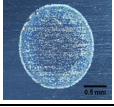
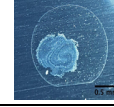
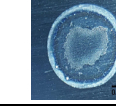
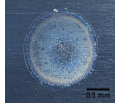
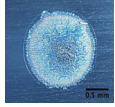
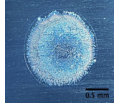
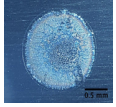
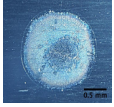
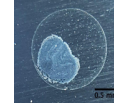
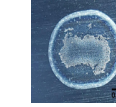
						
Ion exchange						
MSU residence hall						
						
Williamston						
						
Genoa Township private well softened						
						
Untreated groundwater						
Genoa Township well untreated						
						
Rest stop A Okemos						
						
Rest stop D M66/I96 East						
						
Lansing Site C Zeeland						
						
Rest stop B Fenton						
						
Reverse osmosis						
Allegan						
						
Genoa Township well RO						

Table 6.6: (cont'd)

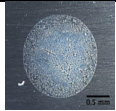
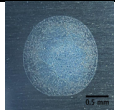
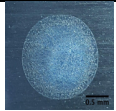
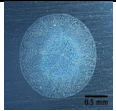
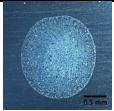
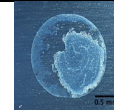
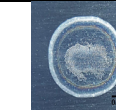
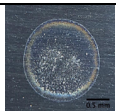
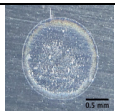
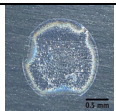
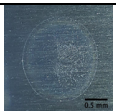
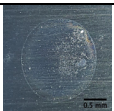
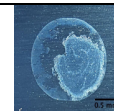
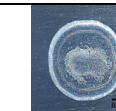
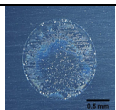
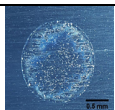
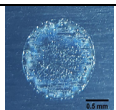
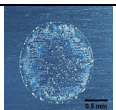
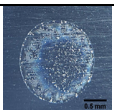
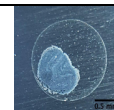
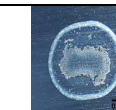
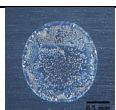
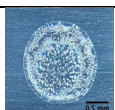
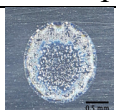
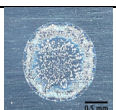
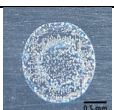
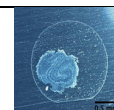
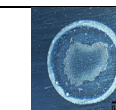
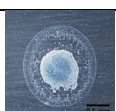
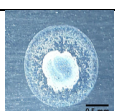
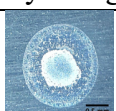
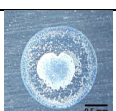
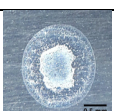
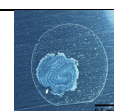
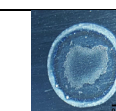
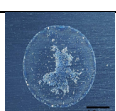
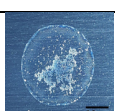
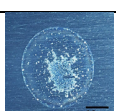
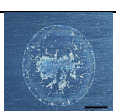
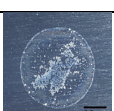
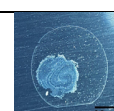
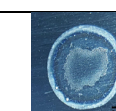
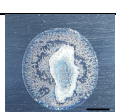
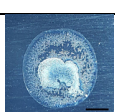
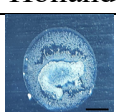
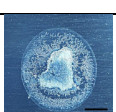
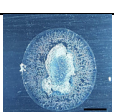
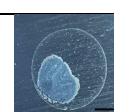
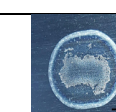
						
Surface waters						
Detroit						
						
Flint						
						
Grand Rapids						
						
Gyoming						
						
Swartz Creek						
						
Holland						
						

Table 6.7: Temperature and humidity effect on residue pattern for four salt mixtures.

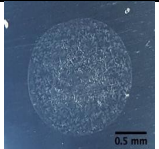
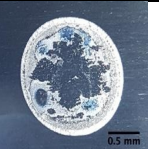
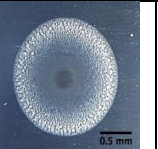
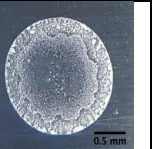
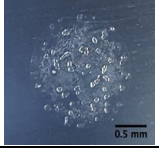
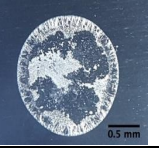
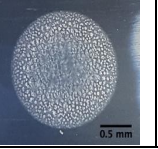
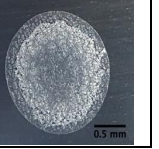
Temperature and relative humidity	Drying time (min)	3.0 mM $CaCl_2$, 1.5 mM $MgCl_2$, 10 mM $NaCl$;	0.5 mM $CaSO_4$, 0.25 mM $MgSO_4$, 5.0 mM Na_2SO_4 ;	0.5 mM $CaSO_4$, 0.25 mM $MgSO_4$, 10 mM $NaHCO_3$;	0.5 mM $CaCl_2$, 0.25 mM $MgCl_2$, 10 mM $NaHCO_3$;
24 °C <20% RH	20				
24 °C 46-48% RH	25				

Table 6.8: Residue patterns of synthetic tap water solutions compared to real tap water at 24 °C and relative humidity of 47%.

	Collected tap water			Simplified synthetic, Calcium, magnesium, sodium, chloride, sulfate, bicarbonate			Complex synthetic, simplified synthetic water sample plus iron, copper, nitrate, fluoride, phosphate		
MSU									
Detroit									
Lansing									
									

Table 6.9: Simple synthetic mixtures on a separate slides analyzed at 24 °C and 48% relative humidity. The low concentration mixtures that are not the same as the previous table are indicated by bold font.

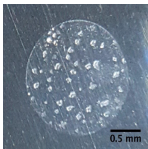
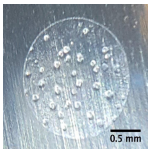
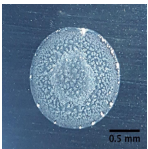
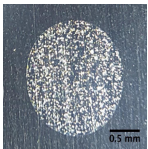
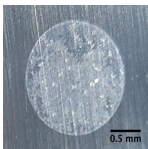
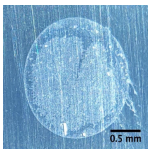
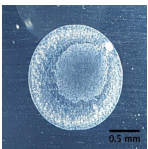
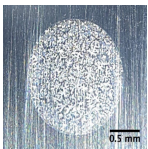
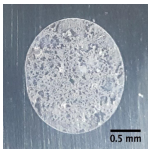
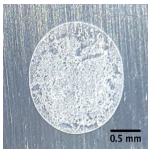
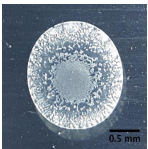
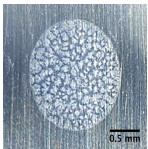
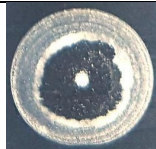
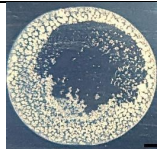
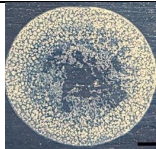
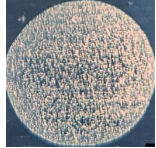
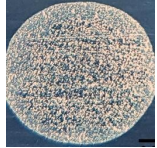
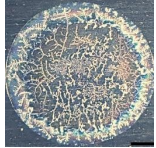
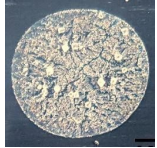
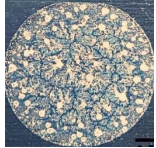
	<i>NaCl</i> 10 mM	3 mM <i>NaCl</i> 5.0 mM	<i>NaHCO₃</i> 10 mM	<i>NaHCO₃</i> 5.0 mM
3 mM <i>Cal₂</i> 1.5 mM <i>MgCl₂</i>				
1 mM <i>Cal₂</i> 0.5 mM <i>MgCl₂</i>				
0.1 mM <i>Cal₂</i> 0.05 mM <i>MgCl₂</i>				

Table 6.10: Images with mis-classification percentage over 70%.

Image is different from other replicates				
Lansing site B	MSU residence hall	MSU residence hall	Portland	Portland
				
Reason not clear			Image in class two	
Genoa Township private well untreated	Genoa Township private well untreated	Battle Creek site B	Genoa Township private well softened	Genoa Township private well softened
				

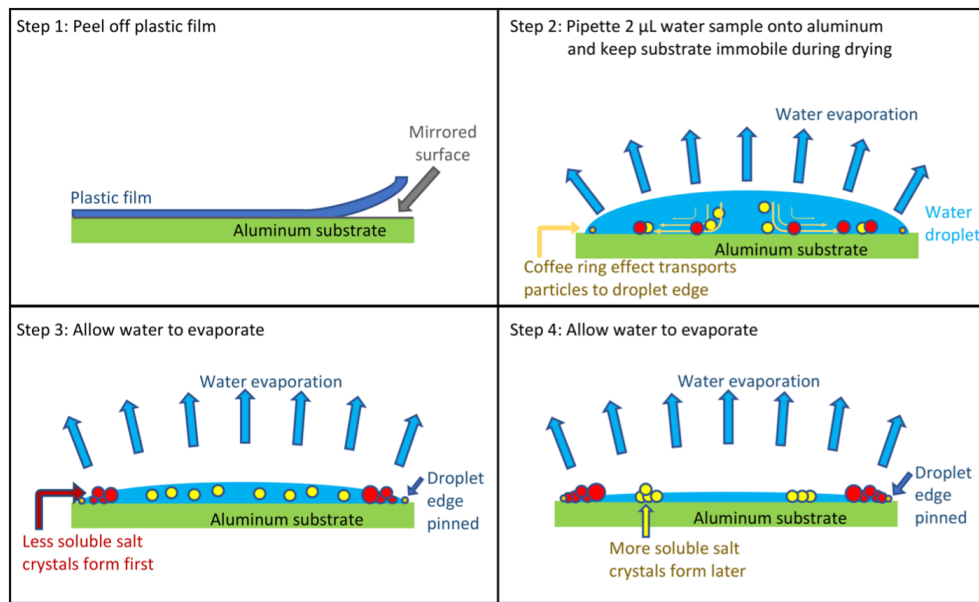


Figure 6.1: The experimental procedure includes depositing two microliter droplets of an aqueous solution onto an aluminum substrate and allowing it to dry without movement.

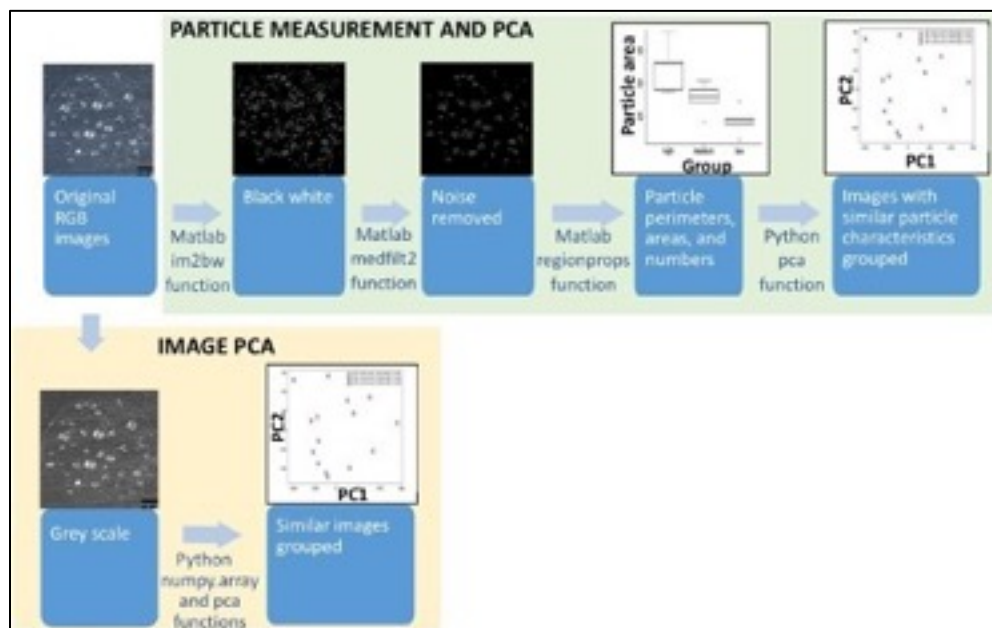


Figure 6.2: Image analysis pipeline in MATLAB and Python.

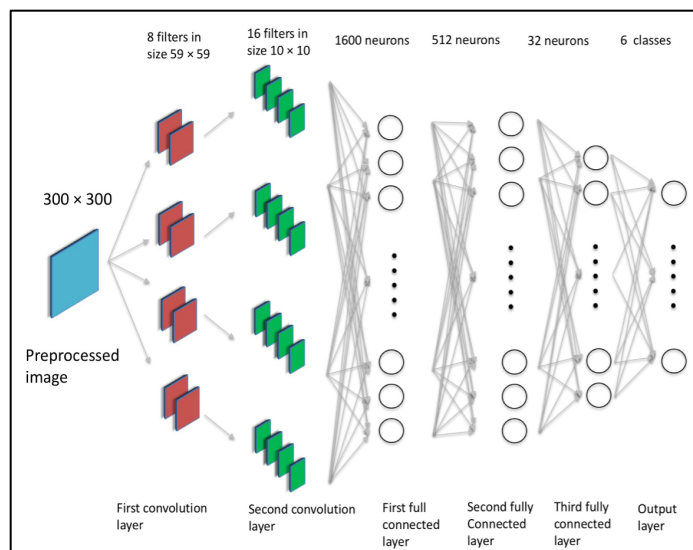


Figure 6.3: A schematic of the convolutional neural network (CNN) model.

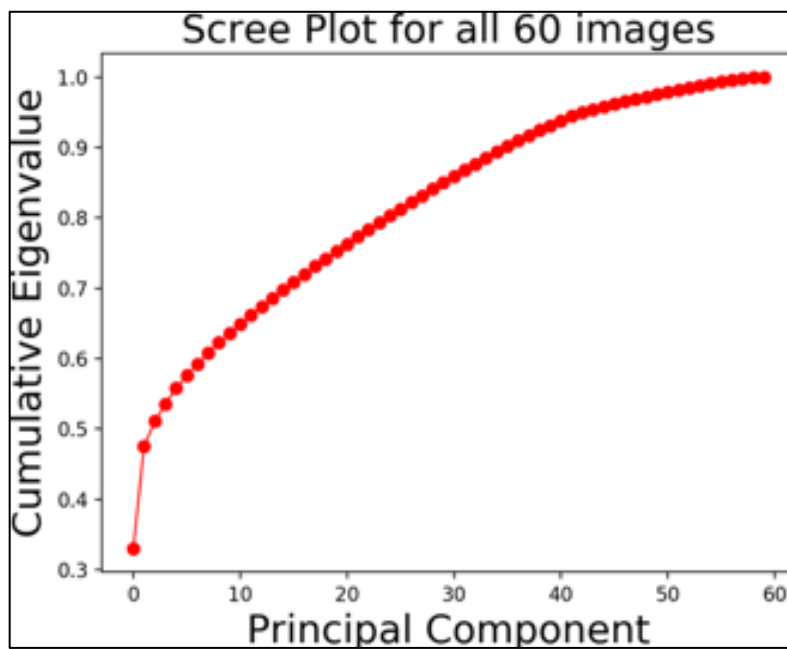


Figure 6.4: PCA on the nanochromatography image files for simplified synthetic waters (five replicates of twelve mixtures of salts).

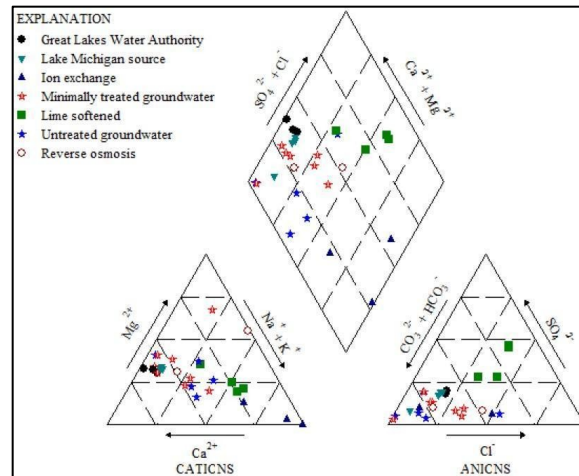


Figure 6.5: Trilinear classification of tap water samples organized by treatment technology.

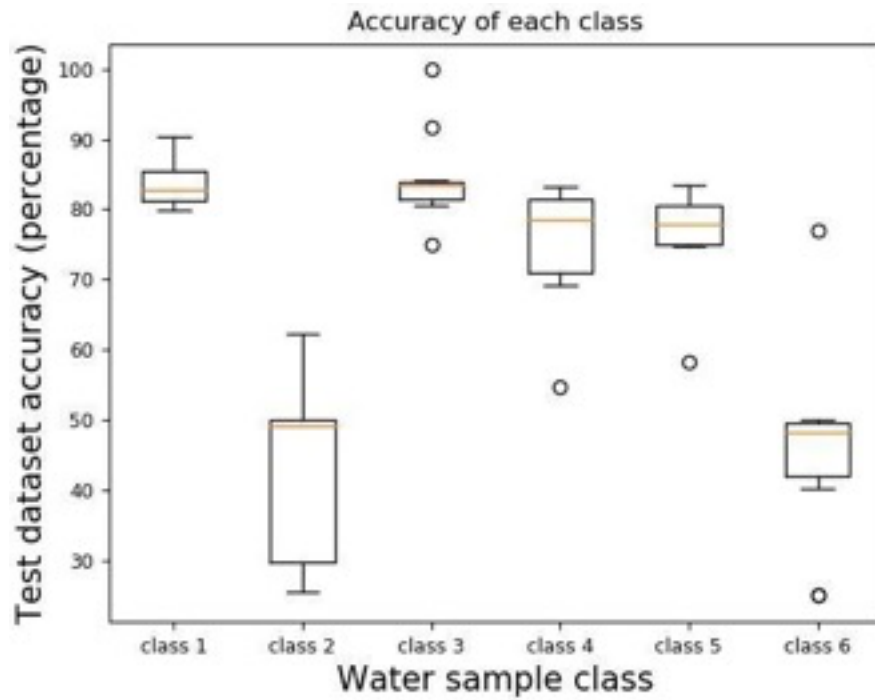


Figure 6.6: Test dataset accuracies by class.

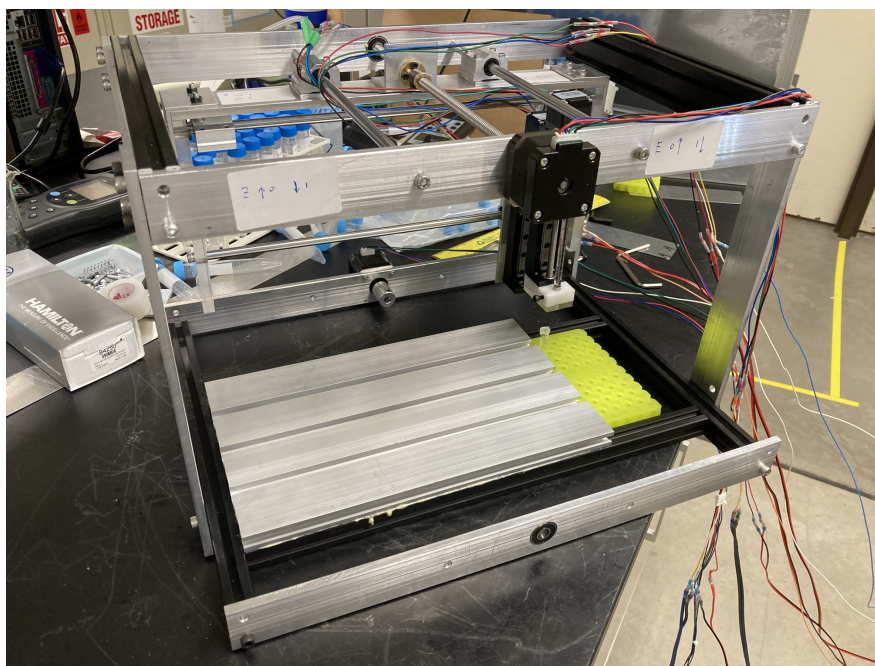


Figure 6.7: Autosampler for coffee-ring effect nanochromatography experiment.

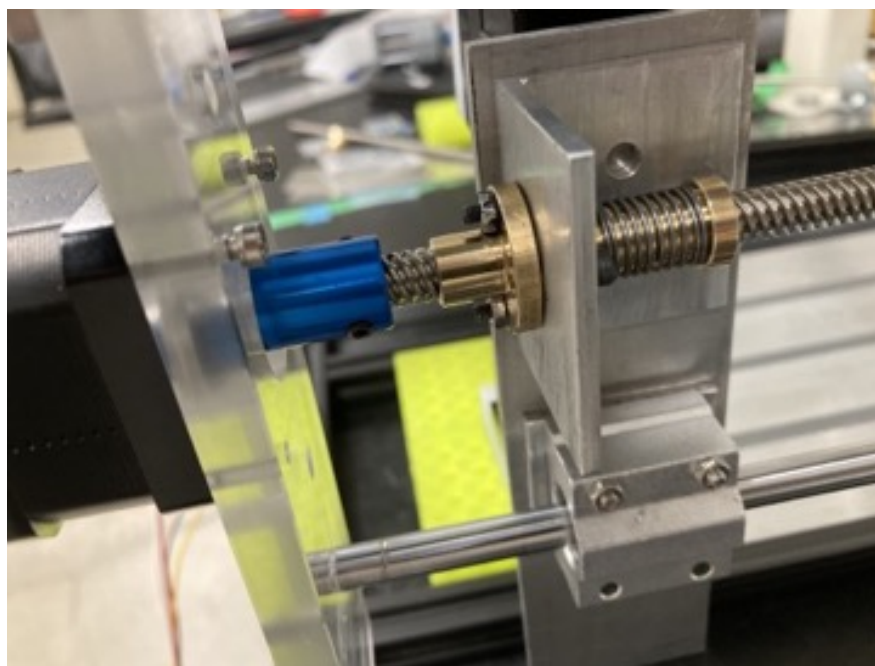


Figure 6.8: Autosampler for coffee-ring effect nanochromatography experiment.

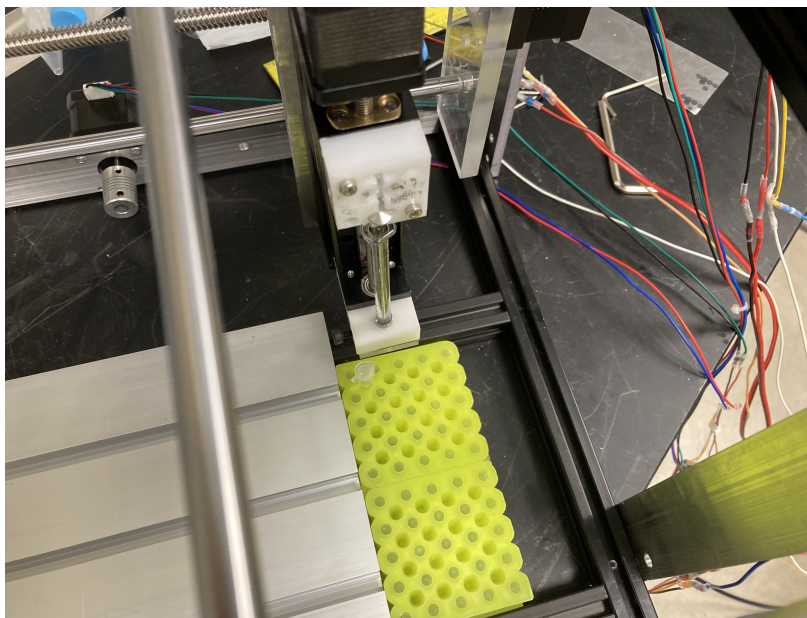


Figure 6.9: Autosampler for coffee-ring effect nanochromatography experiment.

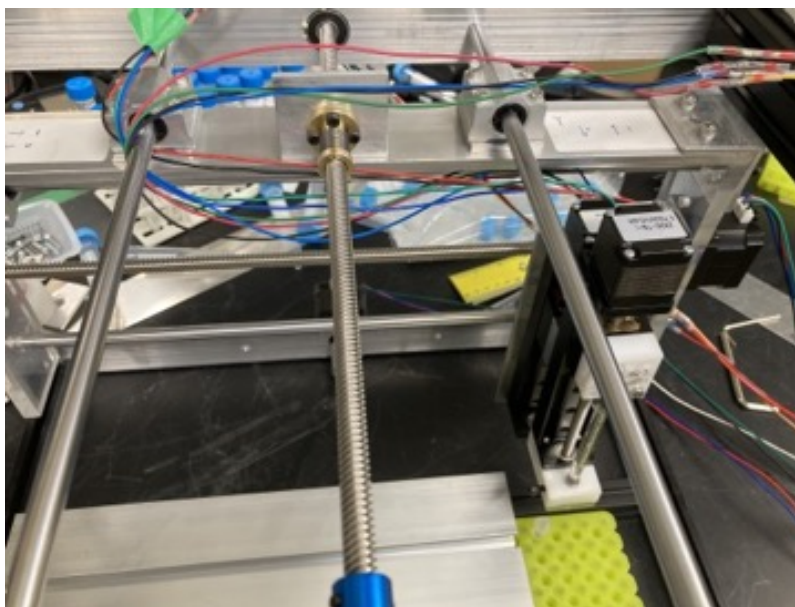


Figure 6.10: Autosampler for coffee-ring effect nanochromatography experiment.

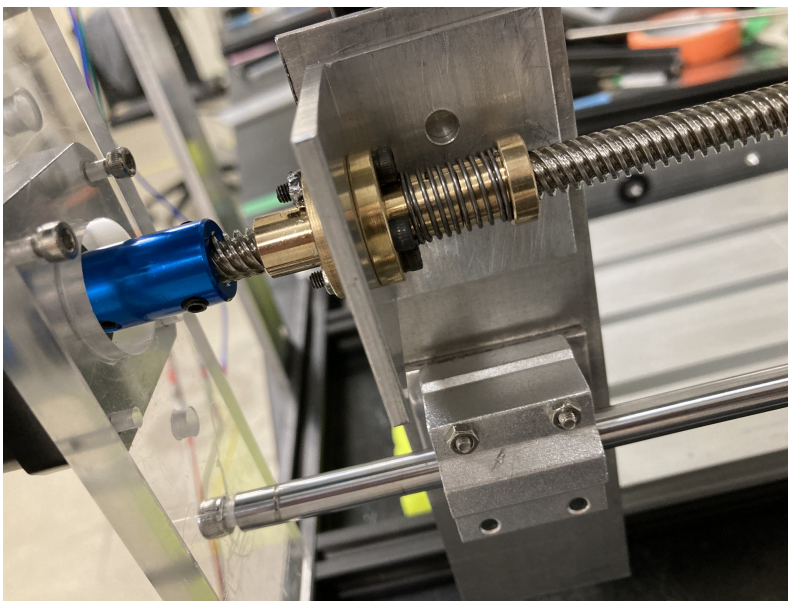


Figure 6.11: Autosampler for coffee-ring effect nanochromatography experiment.

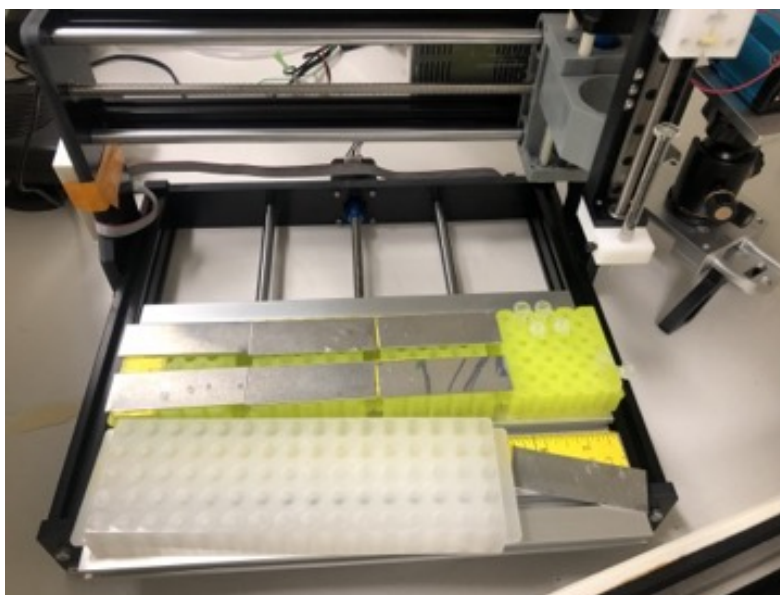


Figure 6.12: Autosampler for coffee-ring effect nanochromatography experiment.



Figure 6.13: Temperature humidity control chamber.

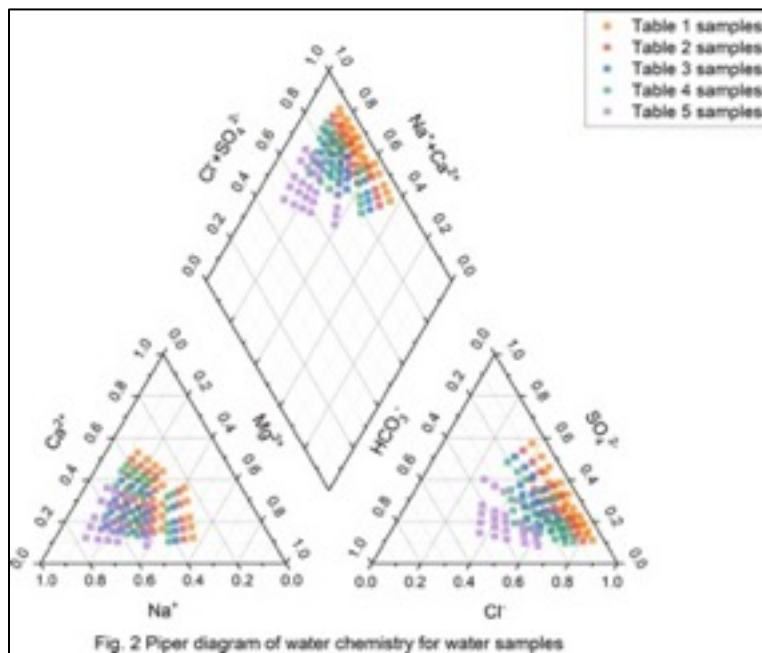


Figure 6.14: Trilinear plot for water samples.

Table 6.11: PERMANOVA clustering result

Experiment condition Temperature, Relative Humidity	20-23 °C	23-26 °C	26-29 °C
35%-40%			
40%-45%			
45%-50%			

Table 6.12: ANOSIM of particles CRE residue features

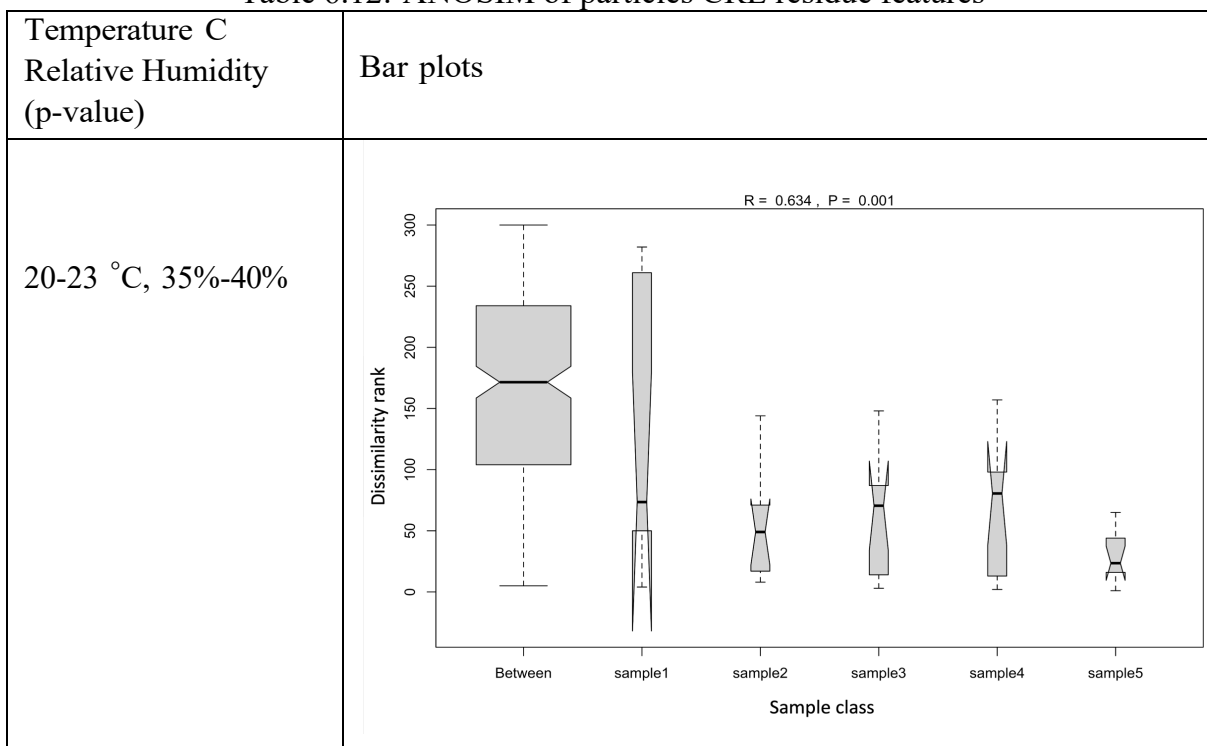


Table 6.12: (cont'd)

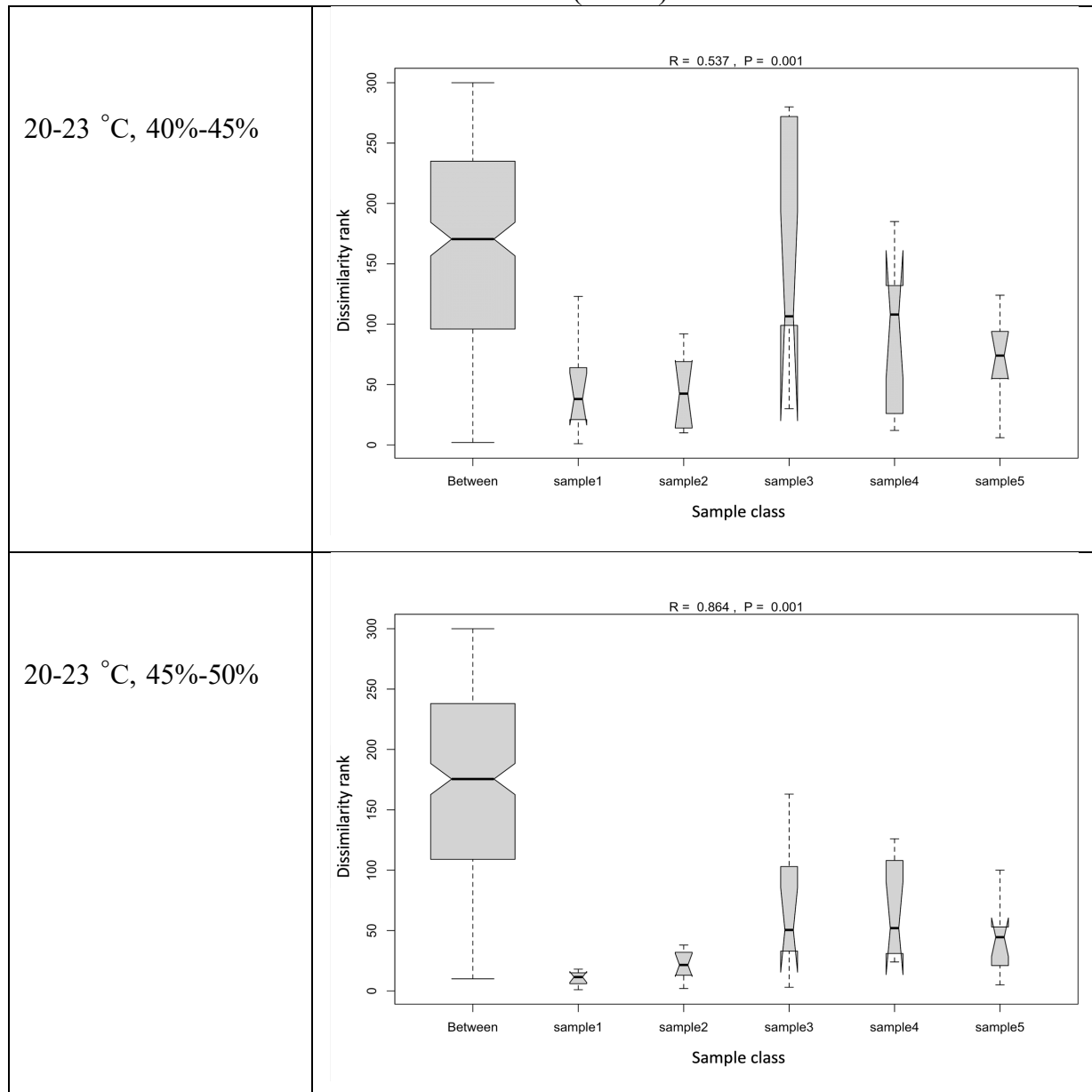
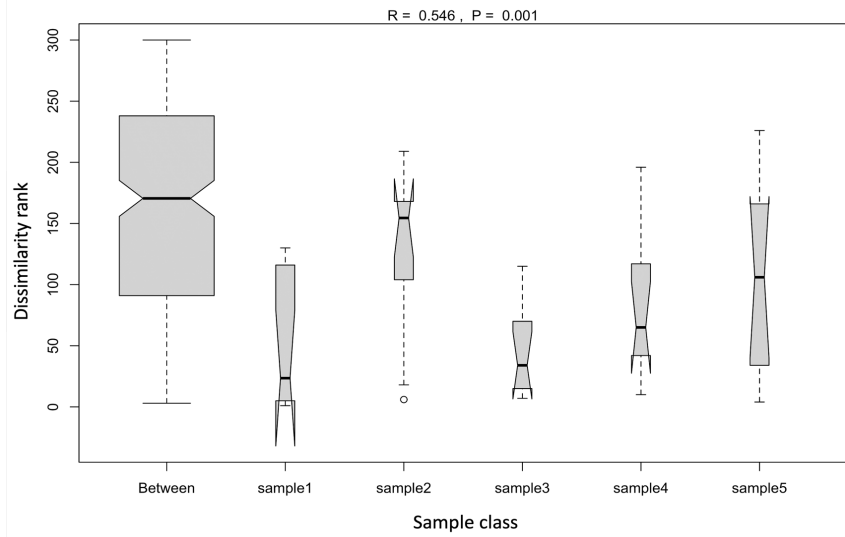


Table 6.12: (cont'd)

23-26 °C, 35%-40%



23-26 °C, 40%-45%

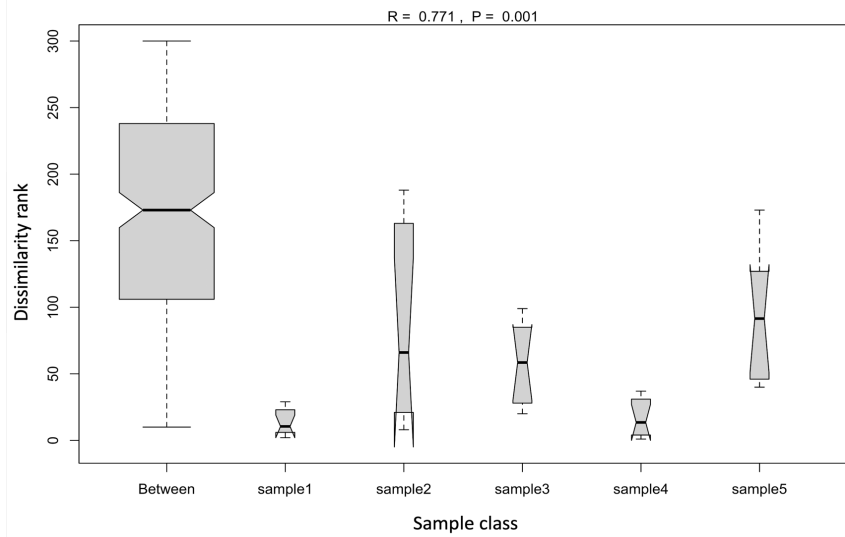
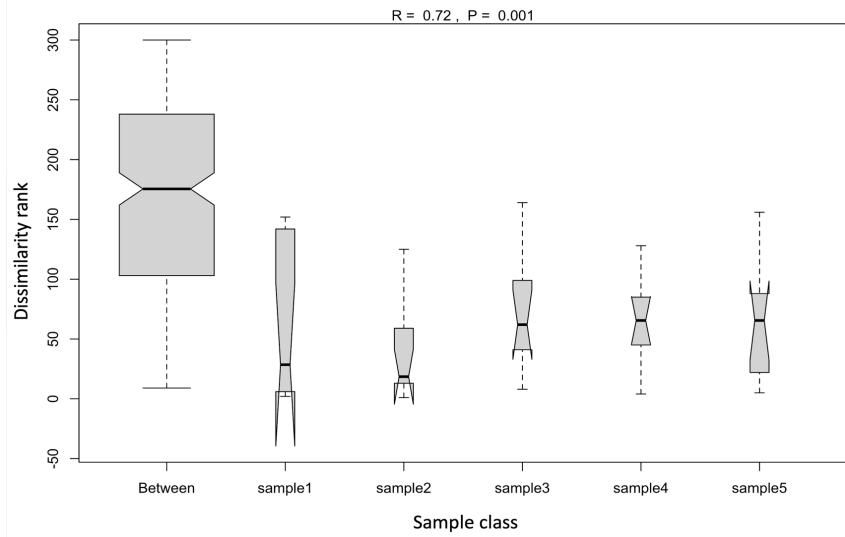


Table 6.12: (cont'd)

23-26 °C, 45%-50%



26-29 °C, 35%-40%

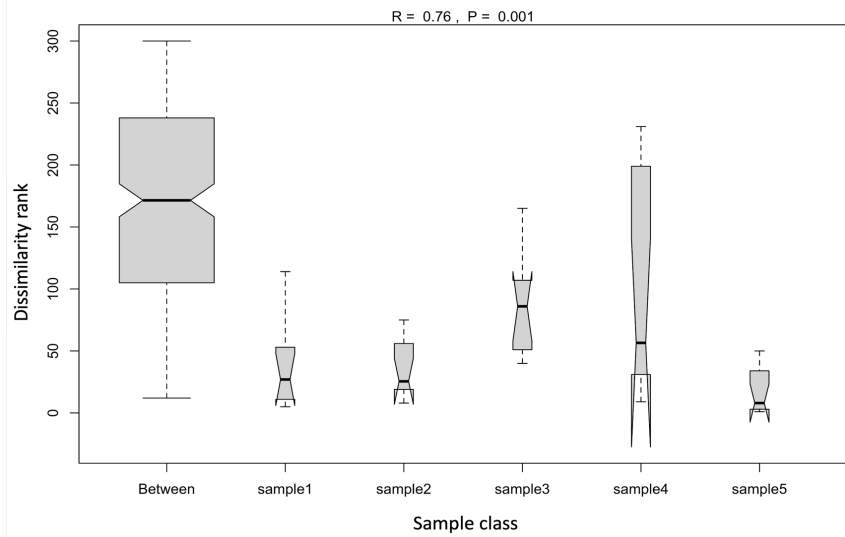


Table 6.12: (cont'd)

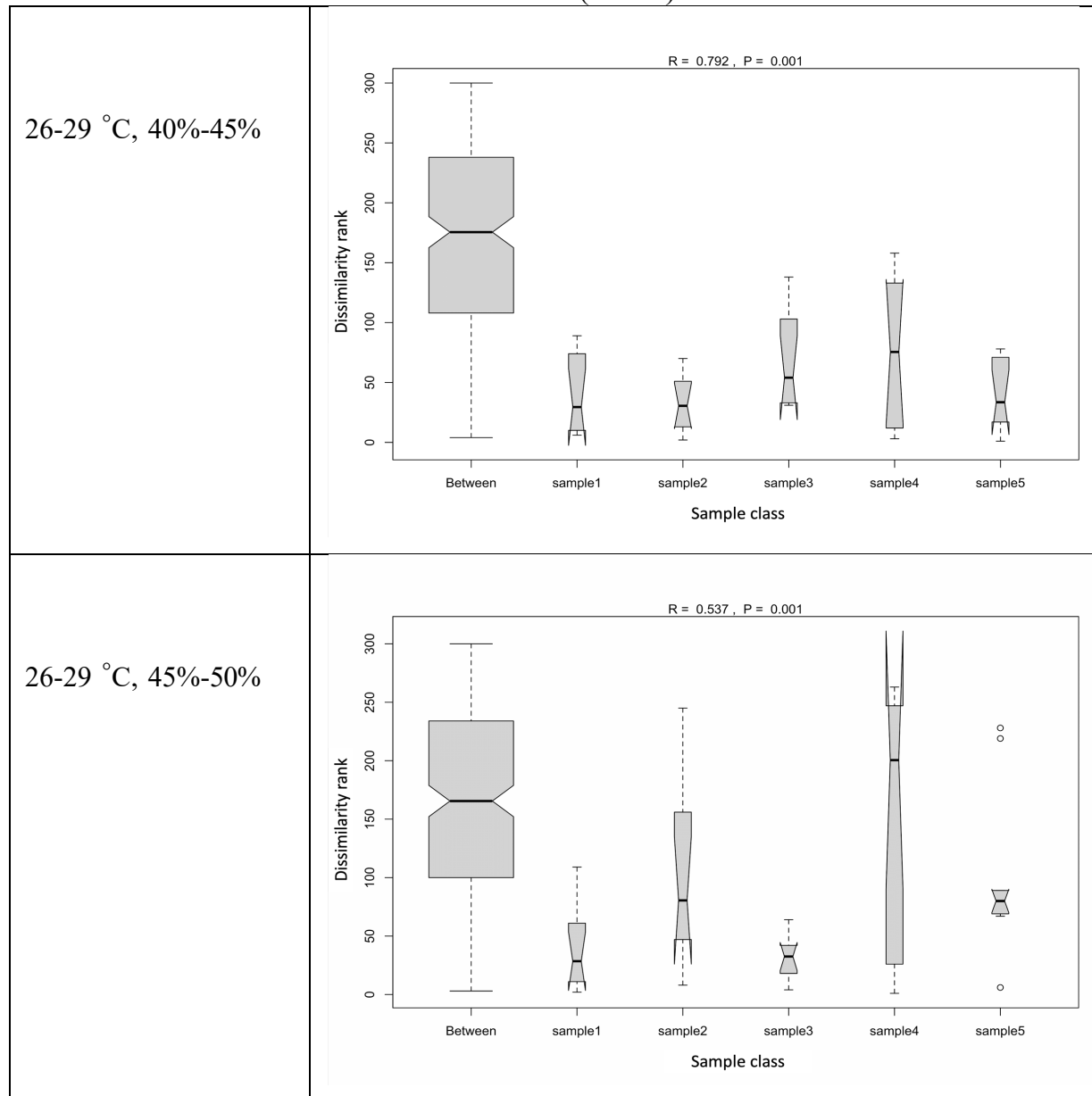


Table 6.13: ANOSIM of CRE residue pattern area. Images are arranged in two orientations: from left to right across the top row, numbered 1 to 25, and from top to bottom along the left column, also numbered 1 to 25.

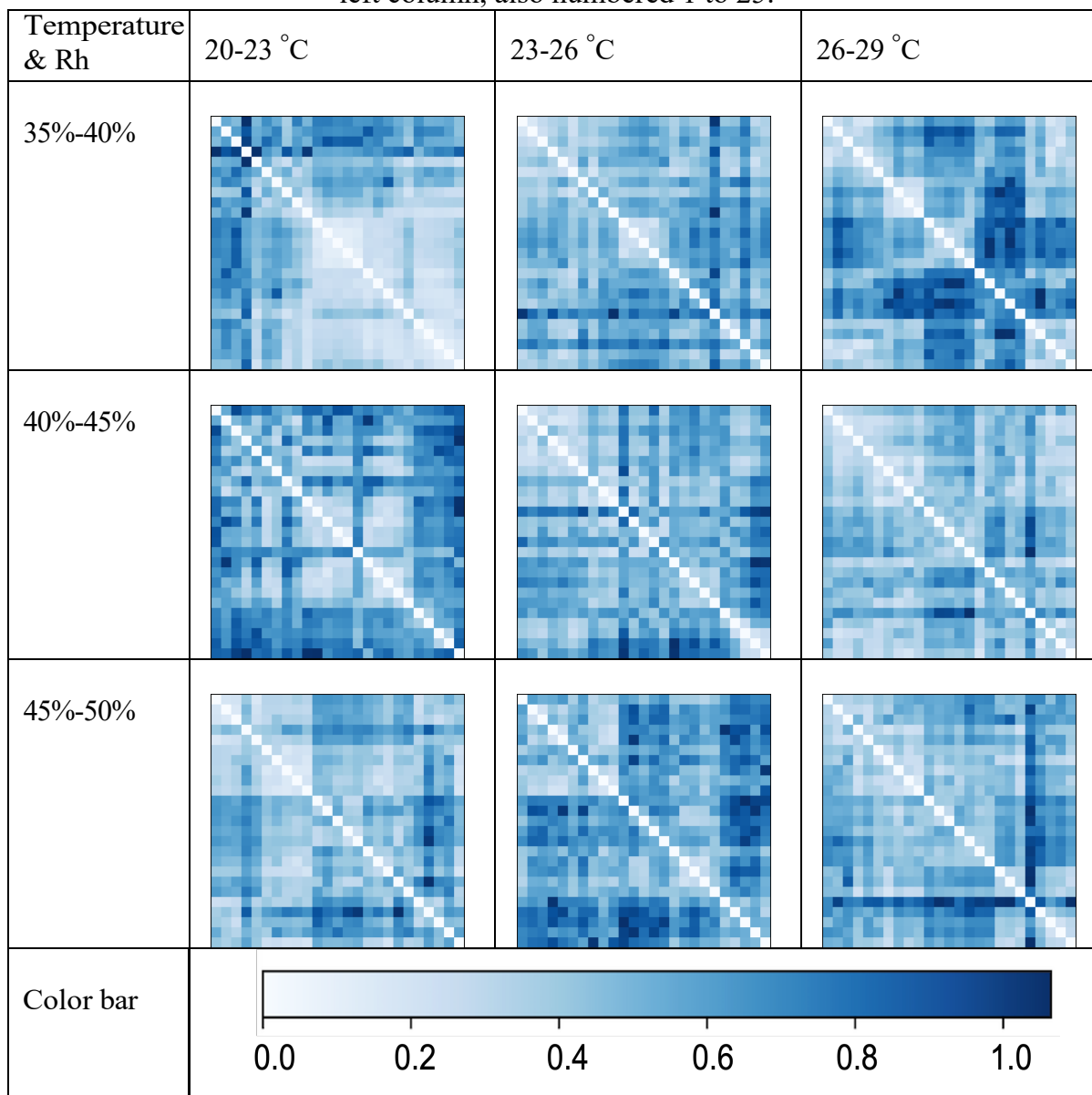


Table 6.14: CMDS of CRE residue pattern area. Red circle represents water sample A; green circle represents water sample B; blue circle represents water sample C; yellow circle represents water sample D; purple circle represents water sample E. The three axes are labeled as Dimension 1, Dimension 2, and Dimension 3.

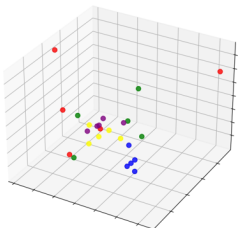
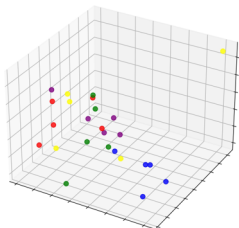
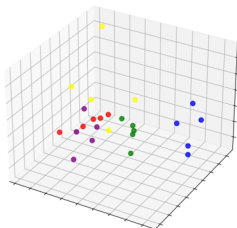
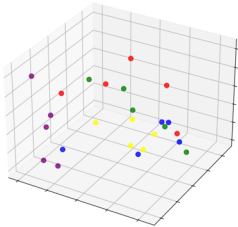
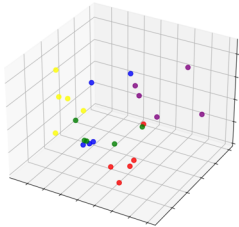
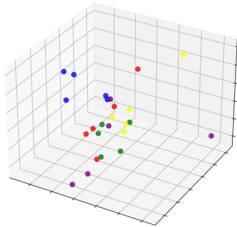
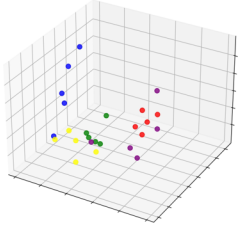
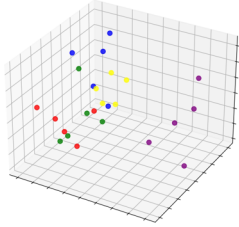
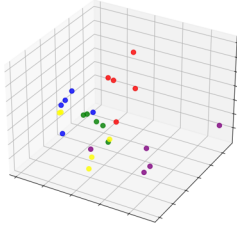
Temperature C Relative Humidity (p-value)	20-23 °C	23-26 °C	26-29 °C
35%-40%			
40%-45%			
45%-50%			
Color bar	 water sample 1  water sample 2  water sample 3  water sample 4  water sample 5		

Table 6.15: ANOSIM of CRE residue pattern perimeter. Images are arranged in two orientations: from left to right across the top row, numbered 1 to 25, and from top to bottom along the left column, also numbered 1 to 25.

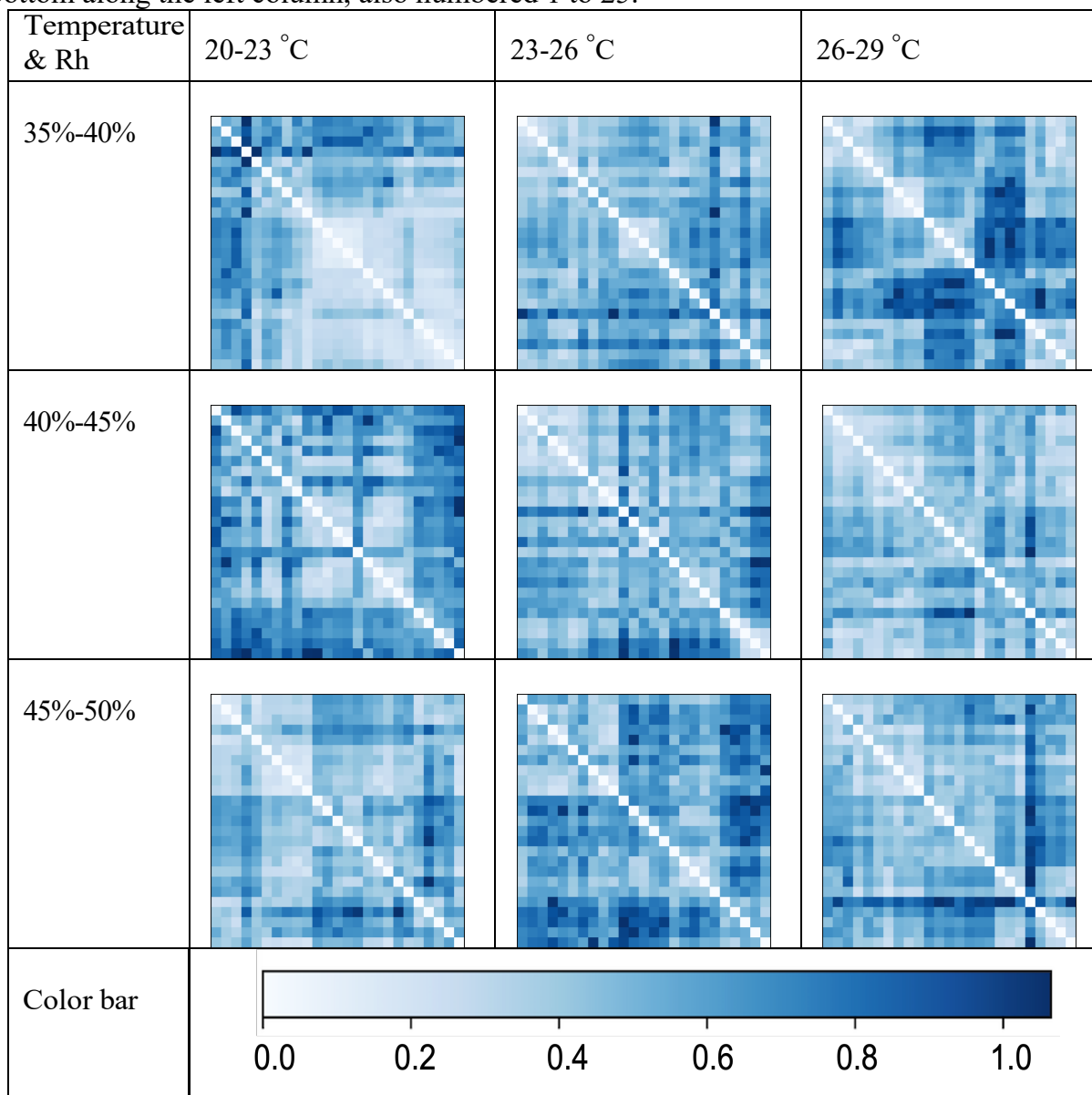


Table 6.16: CMDS of CRE residue pattern centroid perimeter. Red circle represents water sample A; green circle represents water sample B; blue circle represents water sample C; yellow circle represents water sample D; purple circle represents water sample E. The three axes are labeled as Dimension 1, Dimension 2, and Dimension 3.

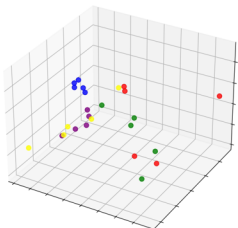
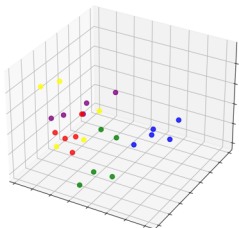
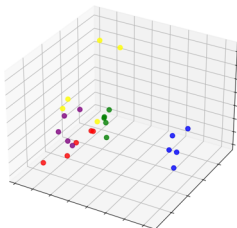
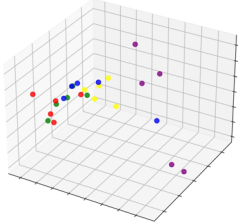
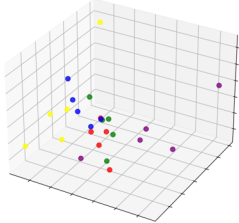
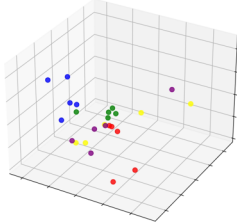
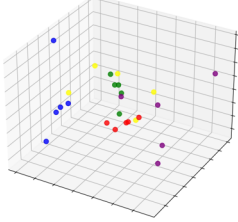
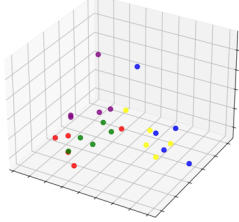
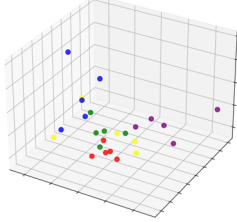
Temperature C Relative Humidity (p-value)	20-23 °C	23-26 °C	26-29 °C
35%-40%			
40%-45%			
45%-50%			
Color bar	 water sample 1  water sample 2  water sample 3  water sample 4  water sample 5		

Table 6.17: ANOSIM of CRE residue pattern centroid. Images are arranged in two orientations: from left to right across the top row, numbered 1 to 25, and from top to bottom along the left column, also numbered 1 to 25.

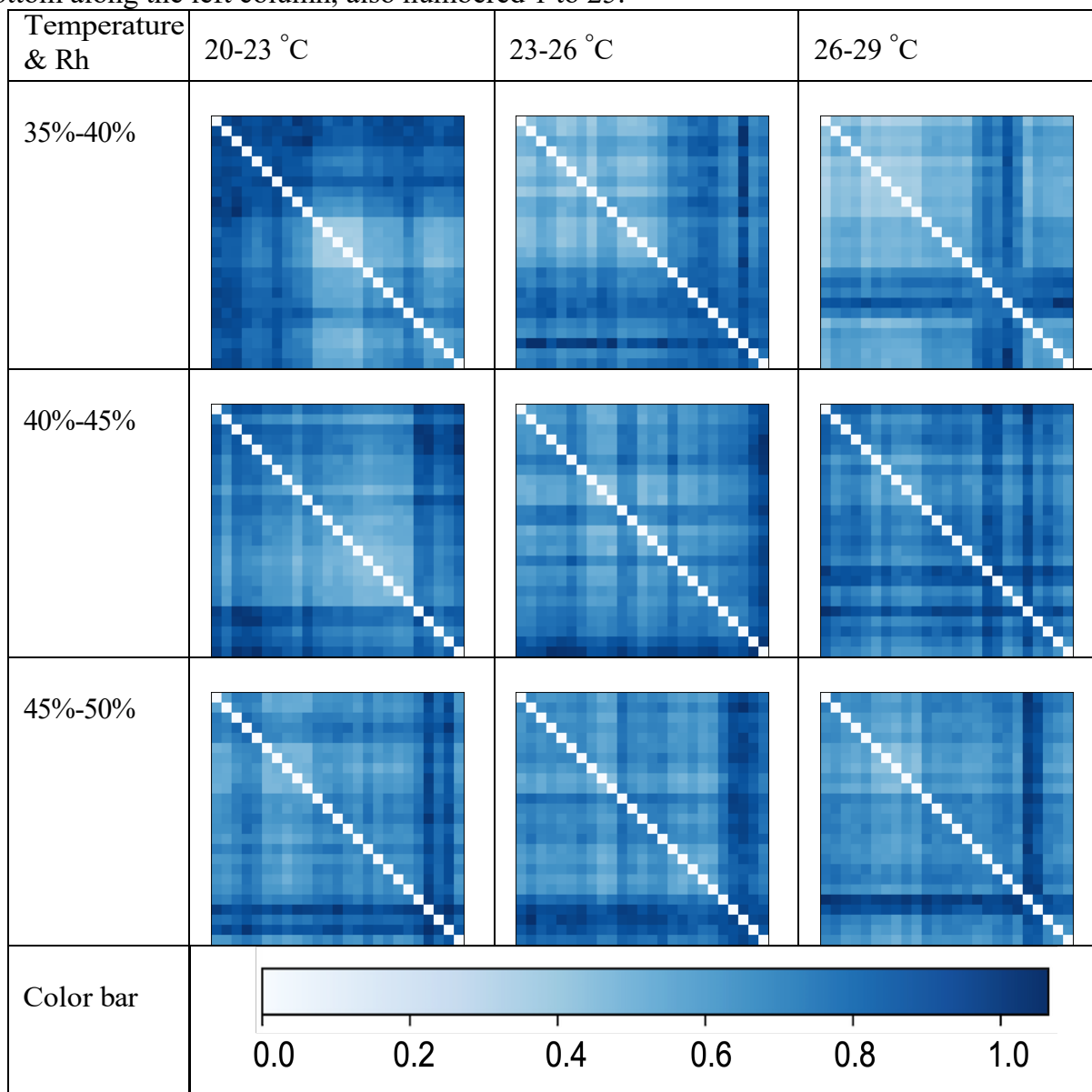


Table 6.18: CMDS of CRE residue pattern centroid. Red circle represents water sample A; green circle represents water sample B; blue circle represents water sample C; yellow circle represents water sample D; purple circle represents water sample E. The three axes are labeled as Dimension 1, Dimension 2, and Dimension 3.

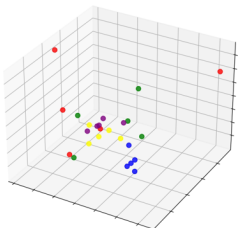
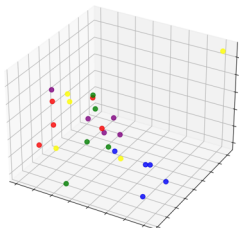
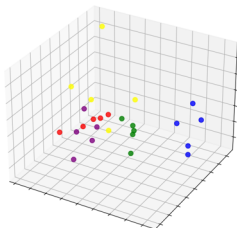
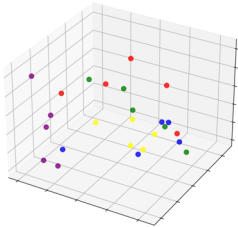
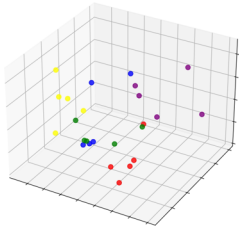
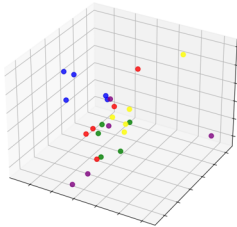
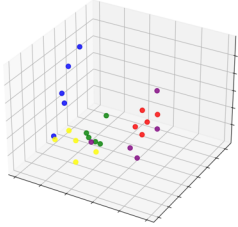
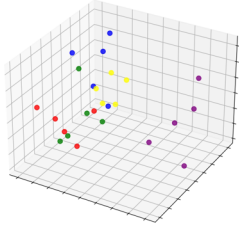
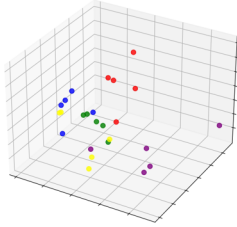
Temperature C Relative Humidity (p-value)	20-23 °C	23-26 °C	26-29 °C
35%-40%			
40%-45%			
45%-50%			
Color bar	 water sample 1  water sample 2  water sample 3  water sample 4  water sample 5		

Table 6.19: ANOSIM of CRE residue pattern eccentricity. Images are arranged in two orientations: from left to right across the top row, numbered 1 to 25, and from top to bottom along the left column, also numbered 1 to 25.

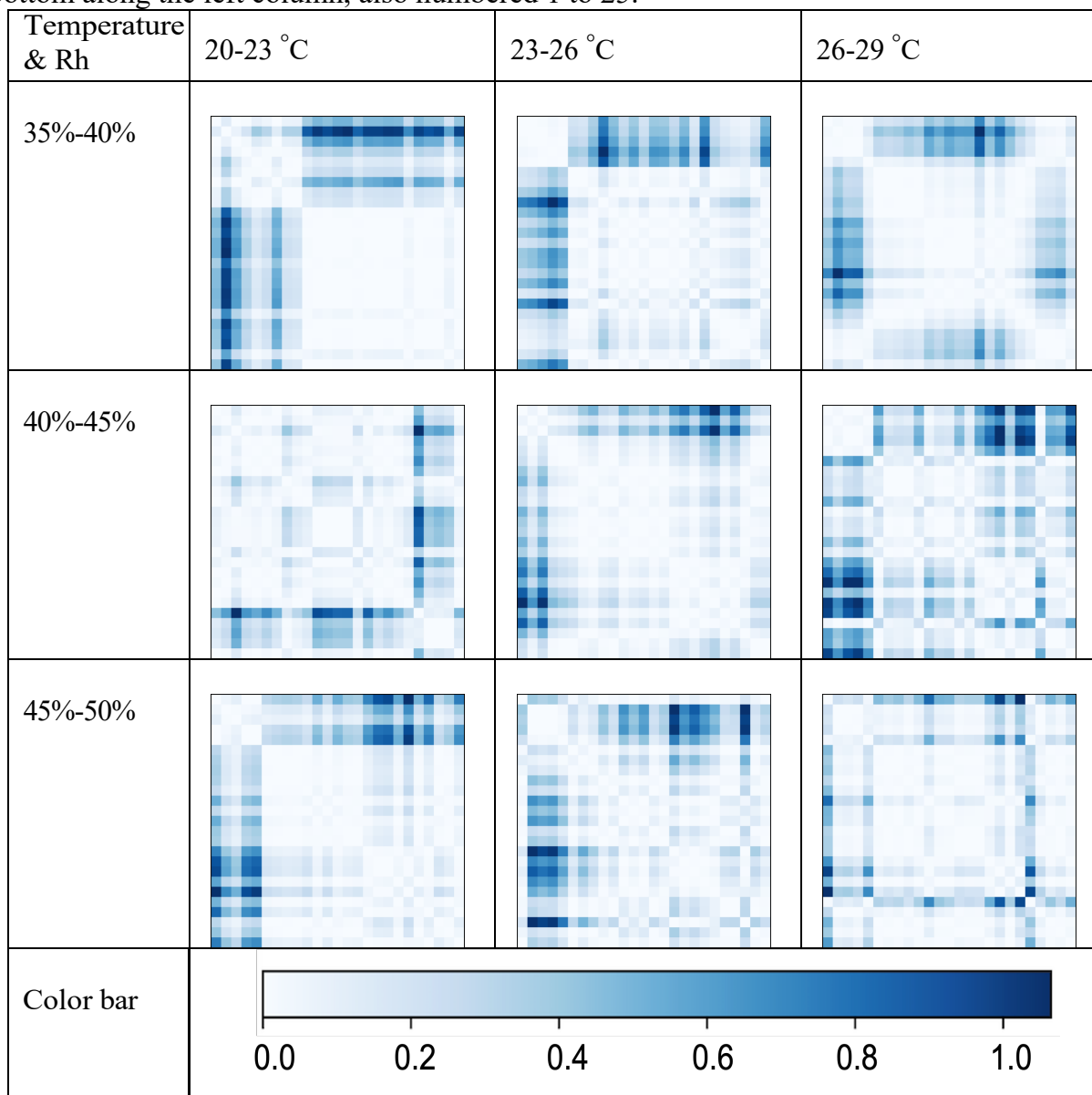


Table 6.20: CMDS of CRE residue pattern eccentricity. Red circle represents water sample A; green circle represents water sample B; blue circle represents water sample C; yellow circle represents water sample D; purple circle represents water sample E. The three axes are labeled as Dimension 1, Dimension 2, and Dimension 3.

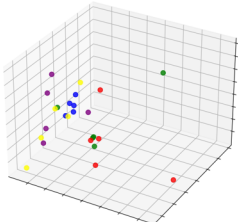
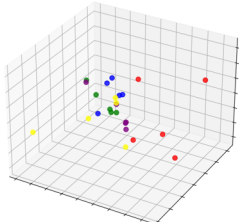
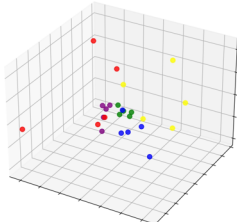
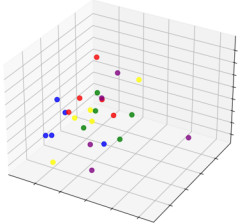
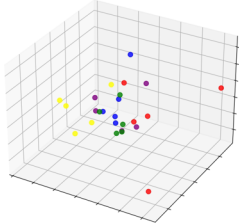
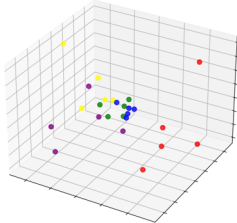
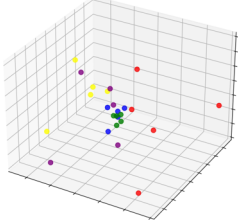
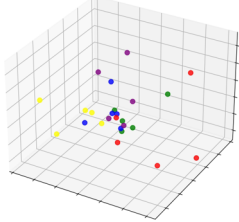
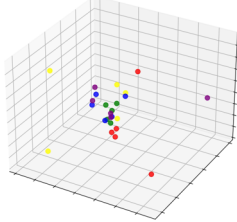
Temperature C Relative Humidity (p-value)	20-23 °C	23-26 °C	26-29 °C
35%-40%			
40%-45%			
45%-50%			
Color bar	 water sample 1  water sample 2  water sample 3  water sample 4  water sample 5		

Table 6.21: Two-way ANOVA for Carbon, Chlorine and Sulfur elements

Condition A						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
Class	4	4.05×10^6	1.01×10^6	151	$< 2 \times 10^{-16}$	***
Element	2	7.72×10^7	3.86×10^7	5751	$< 2 \times 10^{-16}$	***
Class:Element	8	1.42×10^7	1.78×10^6	256.8	$< 2 \times 10^{-16}$	
Residuals	1.18×10^7	7.93×10^{10}	6713			
Condition B						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
Class	4	6.42×10^6	1.6×10^6	245.8	$< 2 \times 10^{-16}$	***
Element	2	4.64×10^7	2.32×10^7	3546	$< 2 \times 10^{-16}$	***
Class:Element	8	1.25×10^7	1.56×10^6	239.7	$< 2 \times 10^{-16}$	
Residuals	1.24×10^7	8.12×10^{10}	6537			
Condition C						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
Class	4	1.25×10^7	3.13×10^6	467.1	$< 2 \times 10^{-16}$	***
Element	2	7.58×10^7	3.79×10^7	5645	$< 2 \times 10^{-16}$	***
Class:Element	8	2.34×10^7	2.92×10^6	434.8	$< 2 \times 10^{-16}$	
Residuals	1.17×10^7	7.83×10^{10}	6714			
Condition D						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
Class	4	1.18×10^7	2.96×10^6	442.8	$< 2 \times 10^{-16}$	***
Element	2	1.16×10^8	5.80×10^7	8677	$< 2 \times 10^{-16}$	***
Class:Element	8	3.17×10^7	3.96×10^6	592.2	$< 2 \times 10^{-16}$	
Residuals	1.19×10^7	7.92×10^{10}	6686			
Condition E						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
Class	4	3.87×10^6	9.68×10^5	148.3	$< 2 \times 10^{-16}$	***
Element	2	3.36×10^7	1.68×10^7	2568	$< 2 \times 10^{-16}$	***
Class:Element	8	8.06×10^6	1.00×10^6	154.3	$< 2 \times 10^{-16}$	

Table 6.21: (cont'd)

Residuals	1.26×10^7	8.26×10^{10}	6532			
Condition F						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
Class	4	1.04×10^7	2.61×10^6	387.3	$< 2 \times 10^{-16}$	***
Element	2	5.92×10^7	2.96×10^7	4398	$< 2 \times 10^{-16}$	***
Class:Element	8	2.13×10^7	2.67×10^6	396.9	$< 2 \times 10^{-16}$	
Residuals	1.20×10^7	8.13×10^{10}	6733			
Condition G						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
Class	4	1.07×10^7	2.67×10^6	400.1	$< 2 \times 10^{-16}$	***
Element	2	6.12×10^7	3.06×10^7	4575	$< 2 \times 10^{-16}$	***
Class:Element	8	2.54×10^7	3.18×10^6	475.4	$< 2 \times 10^{-16}$	
Residuals	1.22×10^7	8.19×10^{10}	6688			
Condition H						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
Class	4	6.66×10^6	1.67×10^6	245.8	$< 2 \times 10^{-16}$	***
Element	2	6.25×10^7	3.12×10^7	4609	$< 2 \times 10^{-16}$	***
Class:Element	8	1.32×10^7	1.65×10^6	244.1	$< 2 \times 10^{-16}$	
Residuals	1.19×10^7	8.09×10^{10}	6778			
Condition I						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
Class	4	5.07×10^6	1.27×10^6	187.7	$< 2 \times 10^{-16}$	***
Element	2	4.87×10^7	2.44×10^7	3605	$< 2 \times 10^{-16}$	***
Class:Element	8	1.81×10^7	2.26×10^6	334.9	$< 2 \times 10^{-16}$	
Residuals	1.19×10^7	8.01×10^{10}	6757			

Table 6.22: Two-way ANOVA for Calcium, Magnesium and Sodium elements

Condition A						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
Class	4	8.47×10^4	2.11×10^4	3.16	0.0132	***
Element	2	3.70×10^7	1.85×10^7	2760	$< 2 \times 10^{-16}$	***
Class:Element	8	1.35×10^6	1.69×10^5	25.21	$< 2 \times 10^{-16}$	
Residuals	1.35×10 ⁷	9.07×10 ¹⁰	6701			
Condition B						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
Class	4	3.85×10^6	9.63×10^5	146.1	$< 2 \times 10^{-16}$	***
Element	2	3.80×10^7	1.90×10^7	2887	$< 2 \times 10^{-16}$	***
Class:Element	8	5.97×10^6	7.46×10^5	113.2	$< 2 \times 10^{-16}$	
Residuals	1.3×10 ⁷	9.07×10 ¹⁰	6593			
Condition C						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
Class	4	1.38×10^6	3.45×10^5	51.45	$< 2 \times 10^{-16}$	***
Element	2	3.87×10^7	1.93×10^7	2882	$< 2 \times 10^{-16}$	***
Class:Element	8	6.36×10^6	7.95×10^5	118.52	$< 2 \times 10^{-16}$	
Residuals	1.34×10 ⁷	9.00×10 ¹⁰	6708			
Condition D						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
Class	4	2.17×10^6	5.42×10^5	81.05	$< 2 \times 10^{-16}$	***
Element	2	5.11×10^7	2.56×10^7	3817	$< 2 \times 10^{-16}$	***
Class:Element	8	5.95×10^6	7.43×10^5	111	$< 2 \times 10^{-16}$	
Residuals	1.36×10 ⁷	9.12×10 ¹⁰	6699			
Condition E						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
Class	4	1.42×10^6	3.56×10^5	54.01	$< 2 \times 10^{-16}$	***
Element	2	2.36×10^7	1.18×10^7	1791	$< 2 \times 10^{-16}$	***
Class:Element	8	9.27×10^6	1.15×10^6	175.8	$< 2 \times 10^{-16}$	

Table 6.22: (cont'd)

Residuals	1.41×10^7	9.30×10^{10}	6587			
Condition F						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
Class	4	6.67×10^5	1.67×10^5	24.82	$< 2 \times 10^{-16}$	***
Element	2	3.16×10^7	1.58×10^7	2354	$< 2 \times 10^{-16}$	***
Class:Element	8	2.74×10^6	3.42×10^5	51.05	$< 2 \times 10^{-16}$	
Residuals	1.38×10^7	9.25×10^{10}	6714			
Condition G						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
Class	4	6.40×10^5	1.60×10^5	23.87	$< 2 \times 10^{-16}$	***
Element	2	2.72×10^7	1.35×10^7	2017	$< 2 \times 10^{-16}$	***
Class:Element	8	2.98×10^6	3.72×10^5	55.66	$< 2 \times 10^{-16}$	
Residuals	1.39×10^7	9.31×10^{10}	6698			
Condition H						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
Class	4	2.74×10^5	6.85×10^4	10.18	$< 2 \times 10^{-16}$	***
Element	2	3.11×10^7	1.55×10^7	2311	$< 2 \times 10^{-16}$	***
Class:Element	8	2.02×10^7	2.52×10^5	37.47	$< 2 \times 10^{-16}$	
Residuals	1.37×10^7	9.28×10^{10}	6732			
Condition I						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	sig.
Class	4	5.05×10^5	1.26×10^5	18.78	$< 1.88 \times 10^{-15}$	***
Element	2	4.21×10^7	2.10×10^7	3132	$< 2 \times 10^{-16}$	***
Class:Element	8	2.26×10^6	2.83×10^5	42.07	$< 2 \times 10^{-16}$	
Residuals	1.36×10^7	9.19×10^{10}	6728			

Table 6.23: Heat map of particle area, eccentricity and element compositions

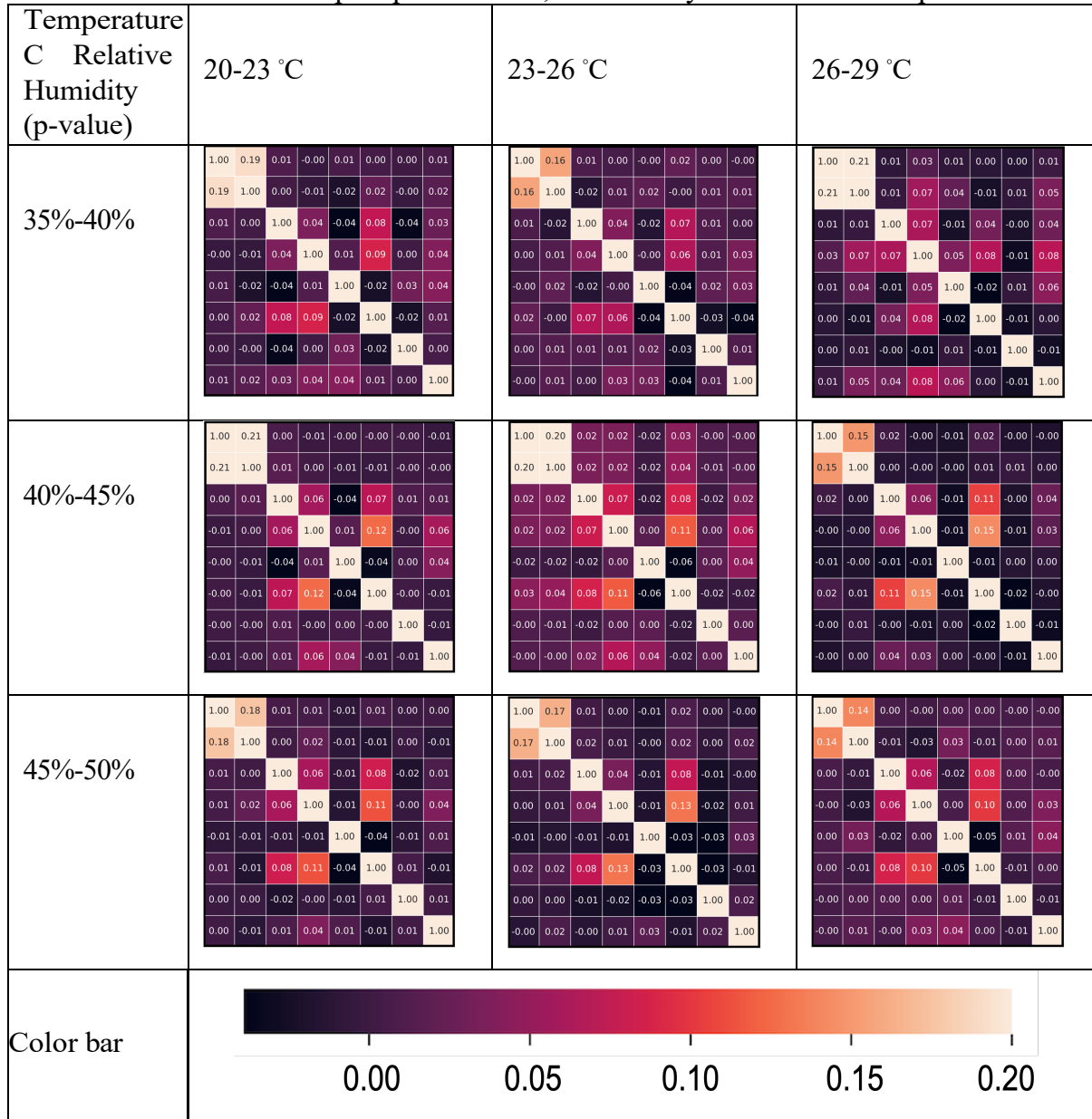


Table 6.24: Nanochromatography images under condition A, 20-23 °C, 35%-40%

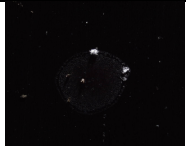
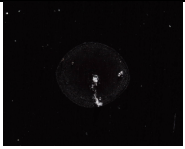
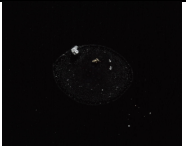
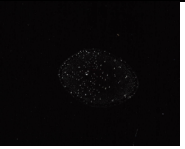
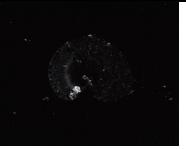
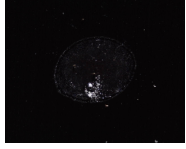
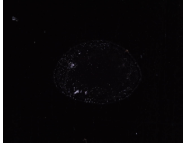
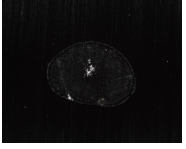
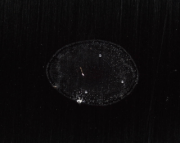
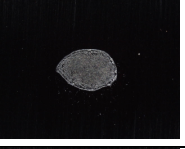
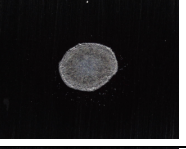
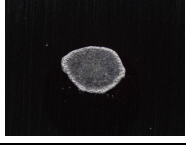
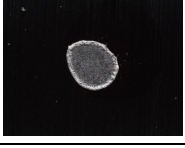
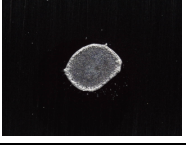
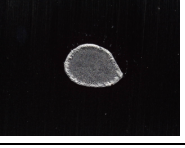
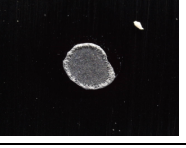
Replicate 1	Replicate 2	Replicate 3	Replicate 4	Replicate 5
				
				
				
				
				

Table 6.25: Nanochromatography images under condition B, 20-23 °C, 40%-45%

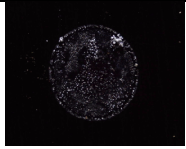
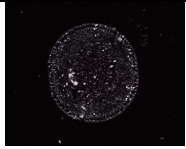
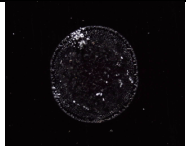
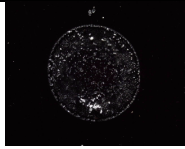
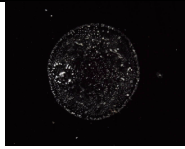
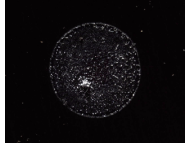
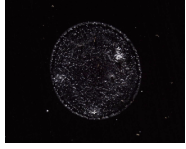
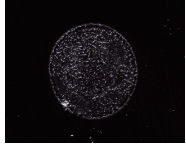
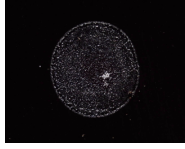
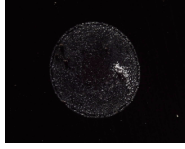
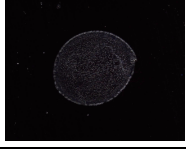
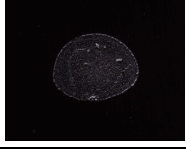
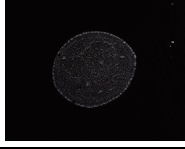
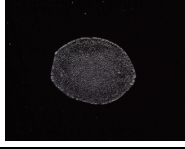
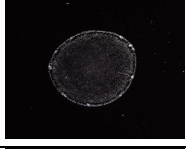
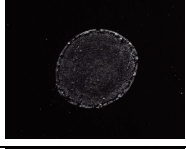
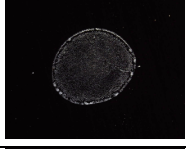
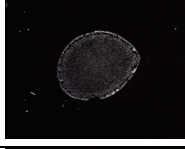
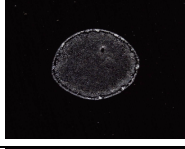
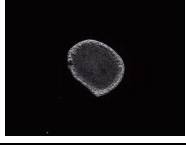
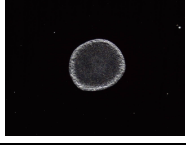
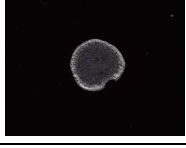
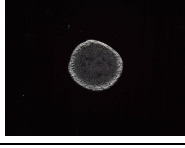
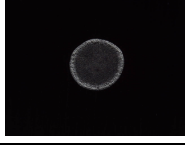
Replicate 1	Replicate 2	Replicate 3	Replicate 4	Replicate 5
				
				
				
				
				

Table 6.26: Nanochromatography images under condition C, 20-23 °C, 45%-50%

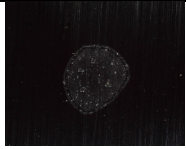
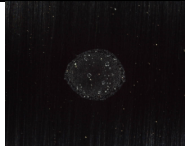
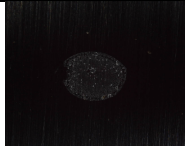
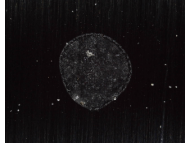
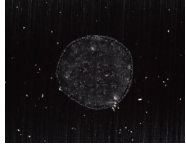
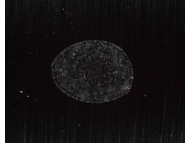
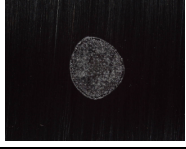
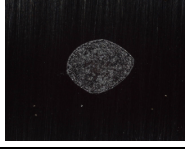
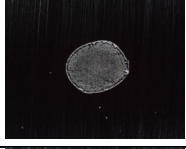
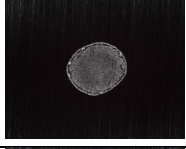
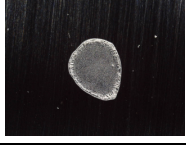
Replicate 1	Replicate 2	Replicate 3	Replicate 4	Replicate 5
				
				
				
				
				

Table 6.27: Nanochromatography images under condition F 23-26 °C, 45%-50%

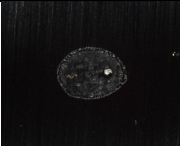
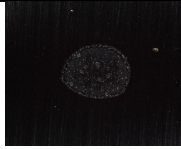
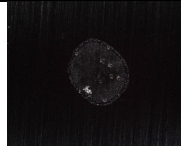
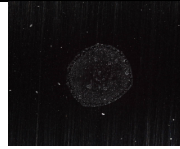
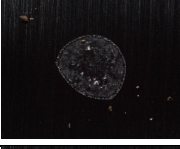
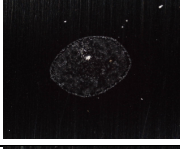
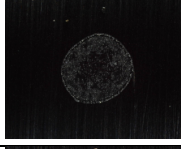
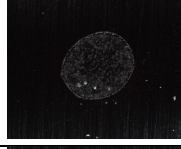
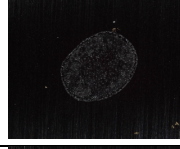
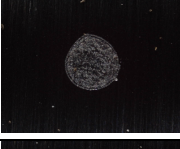
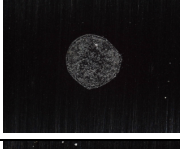
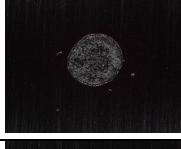
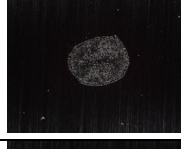
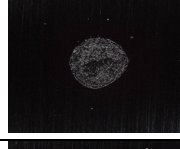
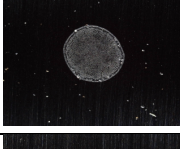
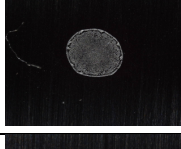
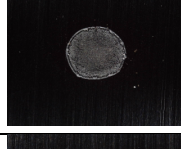
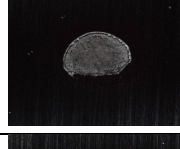
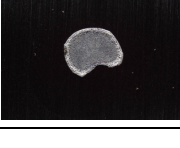
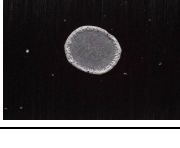
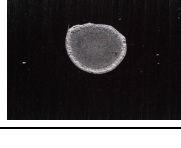
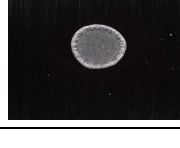
Replicate 1	Replicate 2	Replicate 3	Replicate 4	Replicate 5
				
				
				
				
				

Table 6.28: Water samples recipe of table A for stage 2

Table 1						
$MgCl_2$ 0.45 mM, $NaHCO_3$ 0.25 mM	$CaCl_2$ mM	0.5	0.75	1.0	1.5	2
$MgSO_4$ (mM)						
0.25		1	2	3	4	5
0.5		6	7	8	9	10
0.75		11	12	13	14	15
1.0		16	17	18	19	20
2.0		21	22	23	24	25

Table 6.29: Water samples recipe of table B for stage 2

Table 2						
$MgCl_2$ 0.45 mM, $NaHCO_3$ 0.5 mM	$CaCl_2$ mM	0.5	0.75	1.0	1.5	2
$MgSO_4$ (mM)						
0.25		1	2	3	4	5
0.5		6	7	8	9	10
0.75		11	12	13	14	15
1.0		16	17	18	19	20
2.0		21	22	23	24	25

Table 6.30: Water samples recipe of table C for stage 2

Table 3						
$MgCl_2$ 0.45 mM, $NaHCO_3$ 0.75 mM	$CaCl_2$ mM	0.5	0.75	1.0	1.5	2
$MgSO_4$ (mM)						
0.25		1	2	3	4	5
0.5		6	7	8	9	10
0.75		11	12	13	14	15
1.0		16	17	18	19	20
2.0		21	22	23	24	25

Table 6.31: Water samples recipe of table D for stage 2

Table 4						
$MgCl_2$ 0.45 mM, $NaHCO_3$ 1.0 mM	$CaCl_2$ mM	0.5	0.75	1.0	1.5	2
$MgSO_4$ (mM)						
0.25		1	2	3	4	5
0.5		6	7	8	9	10
0.75		11	12	13	14	15
1.0		16	17	18	19	20
2.0		21	22	23	24	25

Table 6.32: Water samples recipe of table E for stage 2

Table 5						
$MgCl_2$ 0.45 mM, $NaHCO_3$ 2.0 mM	$CaCl_2$ mM	0.5	0.75	1.0	1.5	2
$MgSO_4$ (mM)						
0.25		1	2	3	4	5
0.5		6	7	8	9	10
0.75		11	12	13	14	15
1.0		16	17	18	19	20
2.0		21	22	23	24	25

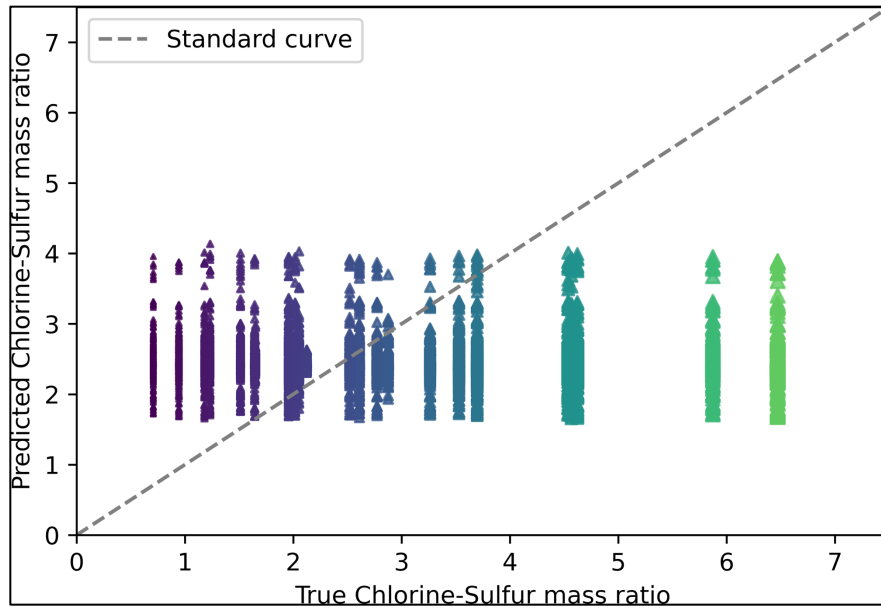


Figure 6.15: TwoVtMoM Chlorine-Sulfur mass ratio. Targets Chlorine-Sulfur mass ratio vs predictions Chlorine-Sulfur mass ratio. Marker colors relates target Chlorine-Sulfur ratio value.

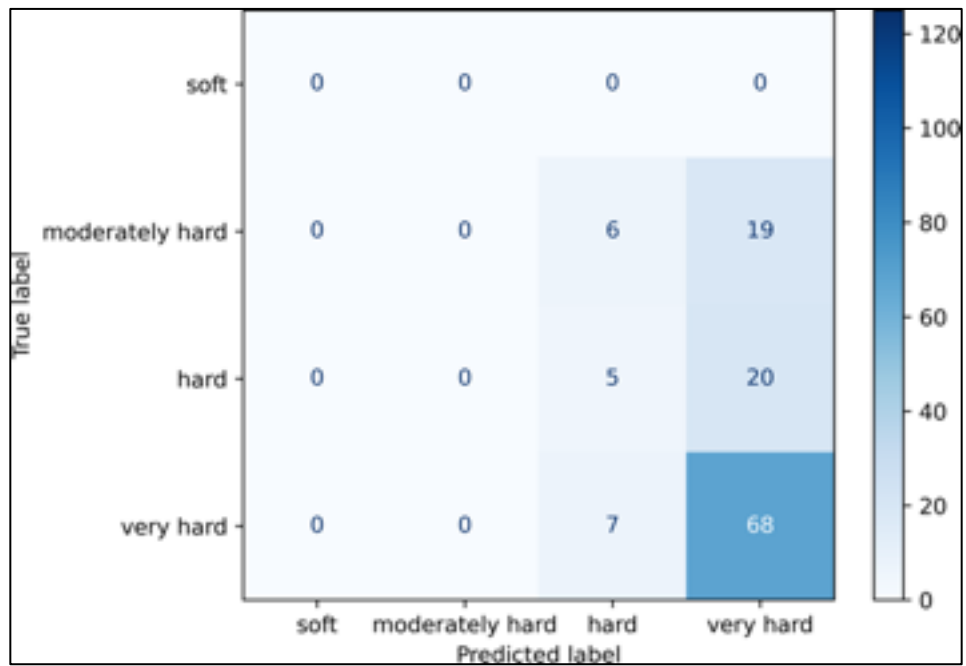


Figure 6.16: TwoVtMoM of water samples hardness category classification results

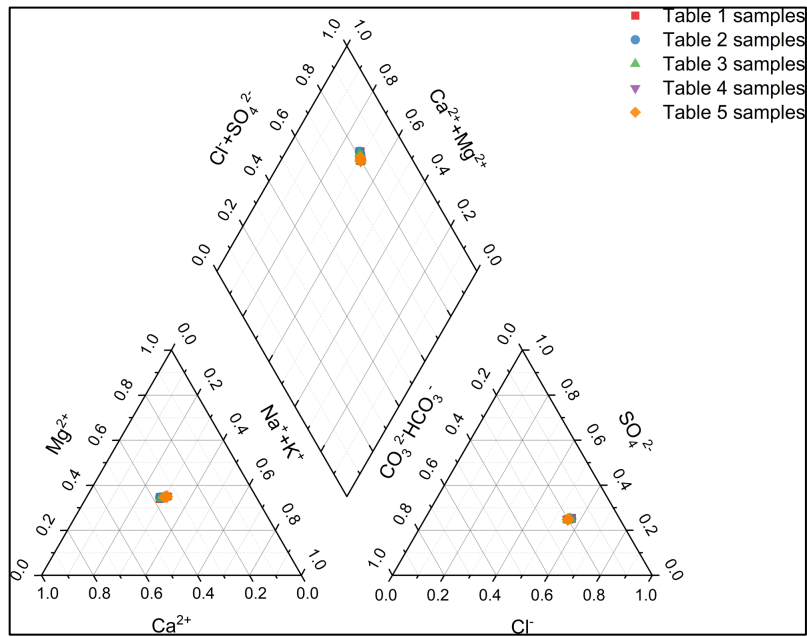


Figure 6.17: TwoVtMoM of water samples trilinear plot.