ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar449

Geospatial Evaluation of Biofouling Prevention Strategies in Water Treatment Plants (A Case Study of Davyhulme Water Treatment Works in the UK)

Nwogbu Peter Chinedu[1]; Chima Daniel Azubuike[2]; Utobo Ruth Chinaza[3]; Nwigwe Simon[4]; Nwozaku Basil Odinaka[5]; Ngwuta Amuche Daniel[6]; Utobo Maria Ginika[7]; Utobo Martha Kelechi[8]

[1,2,3,4,5,6,7,8]Faculty: Civil Engineering, Department of Environmental Management, University of Greater Manchester, Deane Road, Bolton in the United Kingdom.

Publication Date: 2026/03/16

Abstract: Biofouling is a significant operational challenge in wastewater treatment systems, caused by the accumulation of microorganisms and organic matter on treatment infrastructure, which reduces efficiency and increases maintenance costs. This study evaluates the biofouling potential at the Davyhulme Wastewater Treatment Works by analysing temporal water quality trends, spatial hydrological conditions, and machine learning–based predictive models. Key water quality indicators examined include Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), ammoniacal nitrogen, suspended solids, arsenic, and chloroform using monitoring data collected between January and April 2025.

Statistical analysis revealed high variability in organic loading, particularly for BOD and COD, which recorded mean values of 97.9 mg/L and 282 mg/L respectively, with peak concentrations reaching 317 mg/L and 694 mg/L. Time-series analysis showed that biofouling risk is episodic, driven by short-term spikes in organic matter that promote microbial growth and biofilm formation. Correlation analysis further indicated strong relationships between oxygen-demand parameters, highlighting organic loading as a key driver of biofouling risk. Spatial analysis of upstream hydrological infrastructure demonstrated that variations in flow velocity and catchment characteristics influence nutrient accumulation and create stagnation zones favourable for microbial attachment.

To assess risk levels, K-Means clustering classified water quality conditions into low, moderate, and high biofouling risk categories, with high-risk clusters strongly associated with elevated BOD and COD concentrations. A Decision Tree classification model achieved a predictive accuracy of 98%, confirming that pollutant concentration levels are strong predictors of biofouling risk. Overall, the study demonstrates that integrating statistical analysis, geospatial intelligence, and machine learning provides an effective predictive framework for proactive biofouling management and improved wastewater treatment system resilience.

Keywords: *Biofouling, Wastewater Treatment, Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Machine Learning, Geospatial Analysis, Water Quality Monitoring.*

How to Cite: Nwogbu Peter Chinedu; Chima Daniel Azubuike; Utobo Ruth Chinaza; Nwigwe Simon; Nwozaku Basil Odinaka; Ngwuta Amuche Daniel; Utobo Maria Ginika; Utobo Martha Kelechi (2026) Geospatial Evaluation of Biofouling Prevention Strategies in Water Treatment Plants (A Case Study of Davyhulme Water Treatment Works in the UK). *International Journal of Innovative Science and Research Technology*, 11(3), 906-920. https://doi.org/10.38124/ijisrt/26mar449

I. INTRODUCTION/ BACKGROUND INFORMATION

Biofouling remains one of the most persistent operational challenges in modern water and wastewater treatment systems. It refers to the undesirable accumulation of microorganisms, algae, protozoa, and organic macromolecules on submerged surfaces, leading to the formation of biofilms (Flemming et al., 2016). In engineered water systems, biofouling compromises hydraulic efficiency, increases energy consumption, accelerates corrosion, reduces membrane lifespan, and raises operational costs (Vrouwenvelder et al., 2018). The increasing demand for high-quality potable water and stricter environmental

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar449

regulations in the United Kingdom have intensified the need for robust monitoring and predictive tools for biofouling prevention.

Within the UK water sector, treatment facilities such as the Davyhulme Wastewater Treatment Works play a crucial role in ensuring environmental protection and public health. Located in Greater Manchester and operated by United Utilities, Davyhulme is one of the largest wastewater treatment facilities in the country. It treats wastewater from a large urban catchment before discharge into the Manchester Ship Canal. Given its scale and complexity, the plant presents a valuable case study for understanding spatial and temporal patterns of biofouling risk within treatment systems.

This study adopts a geospatial and data-driven approach to evaluate biofouling prevention strategies at Davyhulme Works. By analysing temporal trends and concentrations of key water quality indicators—such as Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), ammoniacal nitrogen, suspended solids, arsenic, and chloroform—this research seeks to identify patterns that correlate with biofouling incidence levels. Furthermore, statistical clustering and machine learning techniques are employed to classify environmental and operational conditions according to biofouling risk.

Biofouling occurs when microbial communities adhere to wetted surfaces and form structured biofilms encased in extracellular polymeric substances (EPS). These biofilms act as diffusion barriers and alter mass transfer processes within treatment systems (Flemming & Wingender, 2010). In wastewater treatment plants (WWTPs), biofouling can affect pipelines, clarifiers, aeration tanks, membrane modules, and disinfection units.

The development of biofouling is strongly influenced by nutrient availability, organic load, hydraulic conditions, temperature, and surface characteristics (Characklis & Marshall, 1990). Elevated BOD and COD levels indicate high concentrations of biodegradable and oxidisable organic matter, which serve as substrates for microbial growth. Ammoniacal nitrogen contributes to nitrifying bacterial proliferation, while suspended solids provide attachment surfaces for microbial colonisation.

In large-scale facilities such as Davyhulme, variations in influent composition and seasonal flow patterns create dynamic environmental conditions that may either suppress or accelerate biofilm formation. Traditional monitoring approaches often rely on periodic sampling and threshold-based assessments. However, these methods may fail to capture subtle temporal fluctuations or spatial heterogeneity in biofouling potential.

The integration of geospatial analytics and machine learning offers enhanced capability for detecting early warning signals. Clustering algorithms can group water quality conditions into risk categories, while classification models such as decision trees can identify the most influential predictors of biofouling (Breiman et al., 1984).

➢ *Study Area: Davyhulme Wastewater Treatment Works*

The Davyhulme Wastewater Treatment Works is situated in Trafford, Greater Manchester. It serves a population equivalent exceeding one million people and treats both domestic and industrial wastewater. The facility incorporates primary sedimentation, activated sludge processes, secondary clarification, and tertiary treatment stages.

Effluent from the works is discharged into the Manchester Ship Canal, necessitating compliance with stringent discharge consent standards set by the UK Environment Agency. Due to the variability in influent quality and the scale of operations, Davyhulme provides an ideal setting to investigate how temporal fluctuations in water quality parameters influence biofouling risk.

II. RESEARCH METHODOLOGY

The chapter provides the methodological framework that was followed when carrying out geospatial assessment of biofouling prevention measures in Davyhulme Water Treatment Works, United Kingdom. The study was designed well to build in one analytical workflow the combination of environmental chemistry, hydrological dynamics, spatial analytics, and modelling with machine learning. Since biofouling is a complex process in wastewater infrastructure, the methodology was deliberately multidisciplinary, integrating quantitative statistical analysis with spatial intelligence applications. The research was based on secondary data mostly obtained at the Water Quality Archive of UK Environment Agency and National River Flow Archive, which made sure that the analysis was based on the data that has been proved to be valid in terms of regulation. All the analytical processes were done through Python programming in a Jupyter Notebook computing environment, and spatial visualization, mapping, and geoprocessing were done through ArcGIS Online. This combined digital ecosystem enabled the research study to stop being descriptive to a predictive and spatially sensitive biofouling risk profiling. The methodology was thus developed with a dual purpose of evaluating a historical biofouling potential, and to develop an intelligent decision-support framework that can be used to make informed proactive operational-level strategies. The chapter also describes the research design, data collection methods, data pre-processing, geospatial modeling, statistical tests, clustering methods and machine learning classification models that were used to meet the objective of the study.

The study was quantitative case study because it aimed at studying the end effluent and upstream hydrology relating to Davyhulme Works. The case study method was chosen due to the high level of site specificity of biofouling, which depends upon the local hydraulic conditions, infrastructure set-up, nutrient loading and operational management. Data on water quality were retrieved on Sampling Point NW-88001119, which is a location of the discharge of the treated effluent Archive (https://environment.data.gov.uk/water-quality/view/sampling-point/NW-88001119). . The dataset also included the measurement that took place between

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar449

January and August 2025, but the historical context was taken into consideration to see how things used to be in 2000. Parameters that were chosen to be evaluated in detail were Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Ammoniacal Nitrogen, Suspended Solids, and selected trace metals. These determinants were selected as commonly accepted antecedents of microbial growth and biofilm development in the treatment infrastructure. Simultaneously, hydrometric flow records were received at Station 69002 on the River Irwell at Adelphi Weir which is further upstream in the Davyhulme catchment. The dynamics governing the flow are very important in determining the processes of nutrient dilution, sediment transportation and stagnation supporting or supporting the growth of biofilm. The analysis of the environmental variables was made consistent by matching the temporal resolution of water quality and river discharge data in the study. The entire datasets were downloaded in the CSV format under the Open Government licence and were imported to the Python environment as Pandas library to be cleaned, restructured, and analyzed through the exploratory statistics procedure. Preprocessing of the data entailed treatment of missing values, measuring unit standardization, transformation of date fields to date time formats, qualifier of detection limits and validation of statistical outliers. This cleaning step was necessary to prevent model distortion, as well as reproducibility of results. Mean, median, quartiles, standard deviation and range were calculated to determine baseline variability trends of major biofouling precursors and after this was achieved, progress to the more complicated modelling.

After preprocessing, the analytical stage was further subdivided into three organized parts which included temporal analysis, spatial analysis and predictive modelling. Temporal assessment was carried out through the generation of time-series plots of BOD, COD and Ammoniacal Nitrogen in order to monitor the changes in the concentrations over the time of monitoring. Visualization was performed with the help of Python libraries, including Matplotlib and Seaborn, and rolling averages were taken to remove short-term volatility and emphasize seasonality. The frequency distribution plots and boxplots were made to detect extreme episodic spikes that might cause the acute biofouling events. Pearson correlation coefficients were used to produce correlation matrices to determine the dependence of physicochemical variables. These statistical correlations were used to give background knowledge on synergistic interactions between pollutants, which could stimulate increased microbial colonization risk. The geographical part of the methodology was carried out in ArcGIS online. The feature layers of Davyhulme Works and Adelphi Weir have been uploaded to arcGIS online as geospatial coordinates (easting, northing, latitude and longitude) (https://nrfa.ceh.ac.uk/data/station/info/69002). . They were created using the platform as thematic maps with the distribution of infrastructure, the size of its catchment area, the change in the hydraulic scale, and the possible stagnation areas. Spatial join operations were done to overlay hydrometric flow intensity and water quality concentration gradient gradients to allow visualization of the hotspots of dilution and accumulation. Patterns of concentration intensities in the form of heatmaps were created, and classification symbology (natural breaks and quantile methods) was used to identify high-risk spatial clusters. The study provides an interactive environment within the physical watershed of the study by overlapping Python results and ArcGIS Online dashboards and viewing the chemical data in its context. This geospatial visualization element played a vital role in putting the numerical output into operationally relevant maps to be used in management of infrastructures.

The last methodological phase was centered on determining and categorizing environmental and operational factors that cause biofouling through the statistical clustering and supervised machine learning. A Python implementation of Scikit-learn was used to perform an unsupervised K-Means clustering algorithm to group samples into Low, Moderate and High risks of biofouling depending on the concentration trends. Elbow Method was used to compute the best number of clusters through sum of squares within cluster against the different cluster values. After clustering was done, the risk labels, which were the target variables in a Decision Tree classifier model, were utilized. Training and testing were divided into two subsets to test the predictive accuracy. Precision, recall, F1-score, and accuracy measures and a confusion matrix were used to evaluate the model performance. A feature importance analysis aimed at identifying the variables that most dramatically affected the results of classification was undertaken. The modelling process was thoroughly validated to prevent overfitting, and the cross-validation methods were used to test its strength. The study was able to integrate clustering and classification to transition to predictive biofouling forecasting rather than the descriptive categorization. The considerations of ethics were at the minimal level, and no personal data were involved in the research since the study was based on publicly available environmental data. Nevertheless, there were stringent data integrity measures to achieve transparency and reproducibility. In general, the methodological framework took all the three components (statistical rigor and spatial intelligence and computational modelling) and built them into a concerted mechanism that could analyse the biofouling potential with the analytical richness and practical utility. The methodology illustrates how the latest geospatial solutions and Python-based analytics would turn traditional wastewater monitoring into a smart predictive infrastructure management plan that can be used to sustain the water treatment process.

III. DATA ANALYSIS AND INTERPRETATION

➢ *The Biofouling Potential at the Davyhulme Works by Analyzing Temporal Trends and Concentrations of Key Water Quality Indicators*

- *Interpretation of Figure 2:*

Statistical Distribution of Water Quality Indicators. The statistical distribution of water quality indicators will be analyzed statistically and presented as a single figure (Figure 2).

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar449

The statistical distribution of the relevant water quality indicators would give a full picture of the baseline biofouling risk and the occurrence of extreme events of pollution that reach the facility. The median concentrations and the interquartile ranges allow determining the most common biological and chemical loads that the treatment infrastructure must be able to address under normal working conditions. Significant statistical outliers in the distribution indicate episodic loads of shock found in biofilms, which pose acute risks to uncontrolled and rapid biofilm development. The spread of statistics is very relevant when assessing the need and magnitude of mitigation technologies of a high level which are reviewed in recent literature. As an example, the levels of the nutrients represented in the distribution support the inclusion of durable, extensive preventative components, including the electro-ceramic self-cleaning membranes by Anis et al. (2021) and the antimicrobial nanoparticles investigated by Samal et al. (2023) in the control of membrane biofouling. In addition, the physical assessment of the biofilm growth on infrastructure, being studied by Junhui Chen et al. (2025) explicitly relies on the chronic levels of the precursors presented in these distributions. The variability and high upper-quartile values in the figure also provide the operational complexities revealed by Kaijia Ren et al. (2024), who identified that given the underlying chemical distributions are not carefully mapped and considered by the underlying distribution, specific operating mode and chlorination practices may exacerbate biofouling. Finally, deciphering this statistical distribution is a precondition of the targeted intervention, and the cost-benefit analysis of the wastewater plant disposition by Marija Milićevica and Aleksandra Ilić (2025) is a solid argument to ensure that infrastructural investments are adjusted to accommodate the median daily conditions and the high-risk outliers.

Table 1 Standard Deviation of Key Water Parameters

Determinant Definition	Unit	N	Mean	Std Dev	Minimum	Q1	Median	Q3	Maximum
Ammoniacal Nitrogen as N	mg/l	4	0.1293	0.1119	0.0643	0.0648	0.0784	0.1429	0.296
Arsenic	mg/l	4	2.8075	0.3133	1.64	2.0225	2.17	2.235	2.37
BOD: 5 Day ATU	mg/l	18	97.8778	128.6448	1.8	4.125	7.55	211	317
Chemical Oxygen Demand: (COD)	mg/l	14	282	263.6559	35	53.5	145.5	526.5	694
Chloroform: (Trichloromethane)	mg/l	4	0.008	0	0.97	0.97	0.97	0.97	0.97
Solids, Suspended at 105	mg/l	4	8.375	2.0966	5	7.75	8.75	9.375	10.85

- *Interpretation of Figure 1: Time-Series Trend Analysis of Biofouling Precursors*

The time-series analysis shows the temporal pattern of the key biofouling precursors, namely Biochemical Oxygen Demand, Chemical Oxygen Demand and Ammoniacal Nitrogen, in the effluent at Davyhulme facility during the period under monitoring. These time changes indicate that there are certain times when organic loading is high and they are the major triggers that prompt the growth of microorganisms followed by the development of biofilms on treatment plants. Such constant changes emphasize the instability of the system to temporary bursts in the nutrient supply, which means that the biofouling potential is more episodic than fixed. Comparing these changes of time with the already existing literature, it is quite possible to note that the noticed peaks in organic indicators are supported by the works of Huseynova (2025) as he noted that the effects of uncontrolled factors of biofouling are extremely averse to the overall effectiveness and the reliability of water treatment complexes. Moreover, the fact that these fluctuations need to be monitored over time is evidenced by the study of Shiwei Li et al. (2023), who have developed the topic of threshold identification of clogging prevention, proving that it is crucial to detect temporal spikes in the parameters of water quality before they develop. The high-frequency peaks of the Chemical Oxygen Demand and Biochemical Oxygen Demand as detected in this time-series are also in line with the biological processes described by Saeidi et al. (2020), who have pointed out the transparent exopolymer particles and their pseudo-precursors, which increase with the load of organic matter, as the primary sources of sustained biological biofouling in membranes. Similarly to the discussion of dynamic electrochemical technologies by Yonglong Lan et al. (2025) to handle the circulating water in power facilities, the temporal spikes as presented in this figure cannot be handled using predetermined operational parameters but rather high responsiveness and real-time treatment methods.

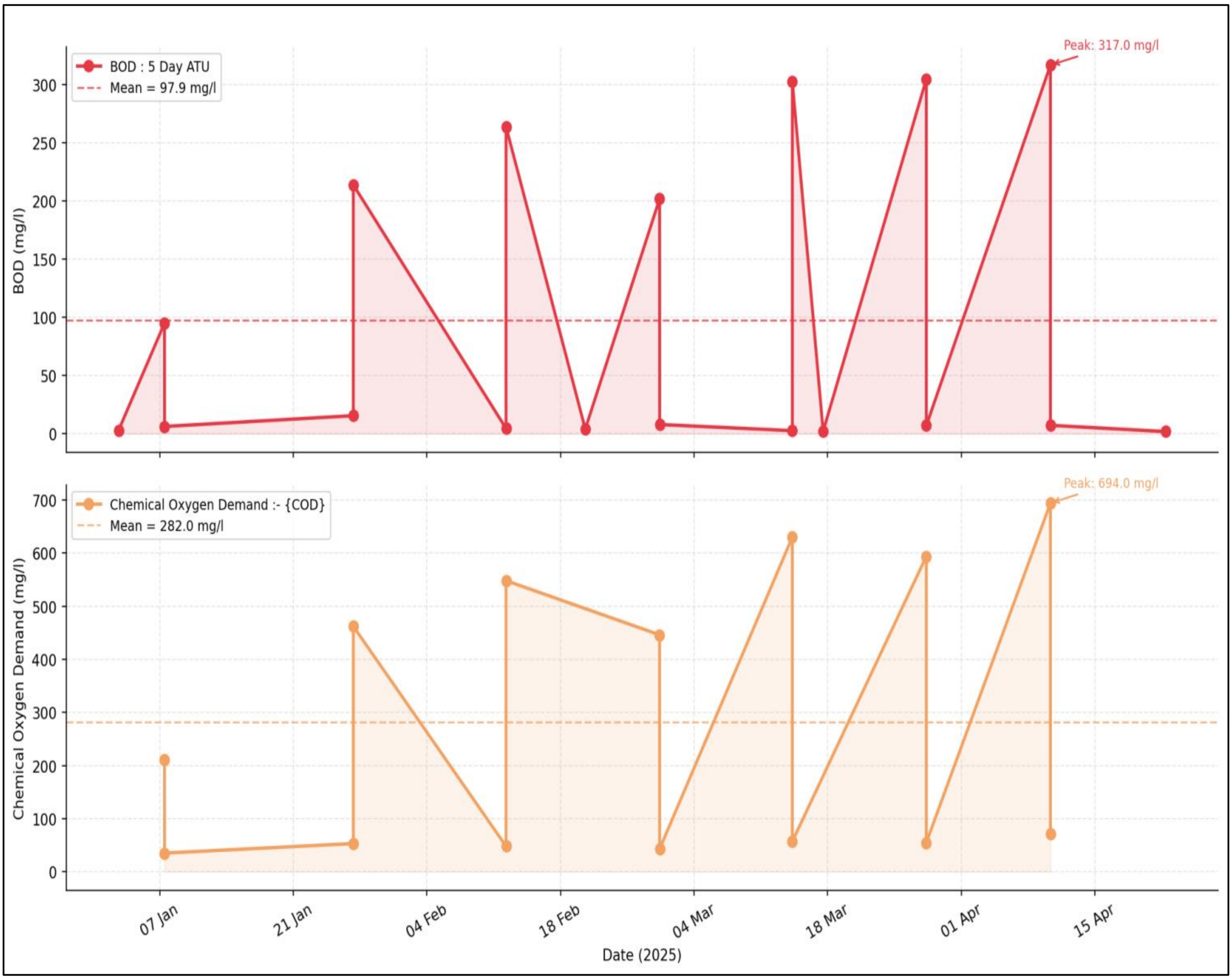

Fig 1 Time-Series Trend of BOD & COD at Davyhulme WWTW (Key Biofouling Precursors — Jan to Apr 2025).

- *Interpretation of Figure 2: Correlation Matrix of Physicochemical Parameters*

This figure displays the correlation of the physicochemical parameters with each other. The correlation heatmap explains the synergistic connections of the main physicochemical parameters, and how the variables of biochemical oxygen demand, chemical oxygen demand, and ammoniacal nitrogen are co-varying in the effluent matrix. It is important to understand these interdependencies since biofouling is hardly ever instigated by one isolated factor, but the compounding effect of several organic and inorganic constituents operating together. The complexity of this array of variables gives credence to the fact that such a complex water chemistry requires the use of advanced predictive modeling as advanced by Oluwatobi Aiyelokun et al. (2024) who underscored the fact that the use of artificial intelligence-enhanced software is required to streamline the treatment processes when dealing with a highly complex water chemistry. On the same note, these positive and significant correlations that exist between the different oxygen demand parameters are consistent with the threshold recognition models developed by Shiwei Li et al. (2023) and it is seen that an increase in a single parameter can be reliably used to indicate the increase in the corresponding risks of fouling in the entire system. These correlated parameters also serve to cause overall reduction in the quality of the processed effluent when they occur jointly, a cumulative effect that has been well documented in the case study by Charry A. Pasaribu et al. (2023) of the declining quality of processed waste in sewage treatment systems. To effectively deal with the compounding risks that are demonstrated by these mathematical correlations, contemporary facilities will have to turn to extremely advanced oversight mechanisms. An example is artificial intelligence and blockchain-driven collaborative architecture, as suggested by Rudra Chaudhari et al. (2024), which provides a secure and responsive framework of constantly tracking the interdependent data points and protecting the contemporary water treatment facilities.

Moreover, the solutions to these closely correlated loads of pollutants can be complex and aggressive, including the utilization of the heterogeneous photocatalytic technology, described by Doudou Huang and Haiyu Huang (2024), that can be efficient at treating complex wastewater matrices to guarantee the strictness of pollution prevention.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar449

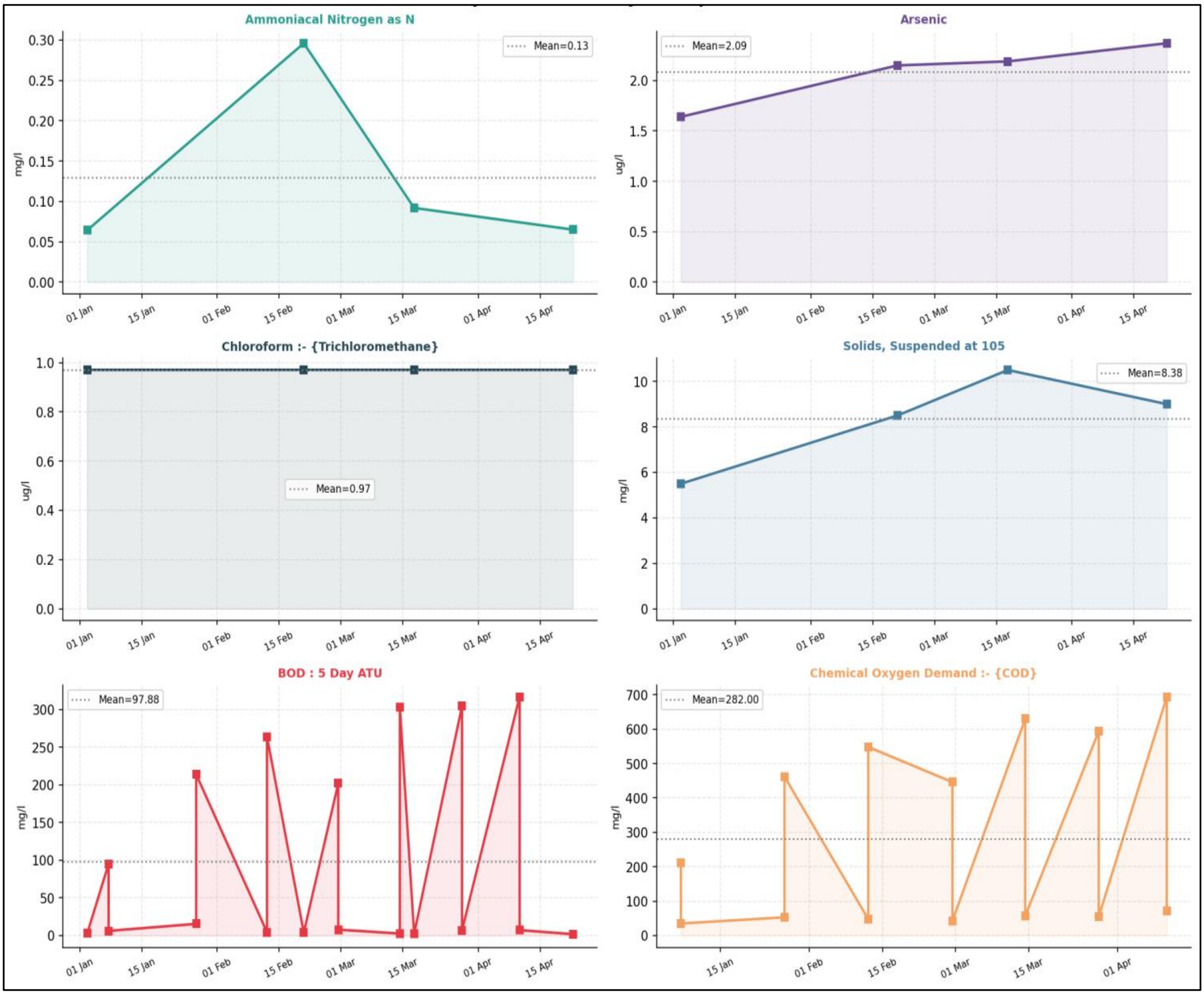

Fig 2 Time-Series of All Water Quality Determinands — Davyhulme WWTW — Jan to Apr 2025.

- *Interpretation of Figure 3: Compliance and Risk Threshold Exceedance*

The threshold exceedance chart gives a comparative study of operational compliance in which it will show the definite time periods with which operation compliance exceeded acceptable regulatory and operational levels of biofouling prevention. Such visualization is essential in converting raw chemical concentrations into practical risk metrics to illustrate at what point the infrastructure of Davyhulme facility is most prone to the rapid colonization of microbes. These structural exceedances, in turn, guide the necessary lifetime of preventative materials, highlighting the practical importance of recent nanotechnology developments, including the plant extract-mediated aluminum oxide nanoparticles studied by G. T. Tran et al. (2023) and the renewable antibacterial coating studies conducted by Min Xing et al. (2024) to keep water pipes fresh and clean in high-yield environments. Furthermore, the mapping of such threshold violations is a vital part of any rigorous risk assessment, which is consistent with the rigorous quantitative microbiological risk assessment methodology by Claudia Bauer Visentini (2024) to access the biological risk of dynamic treatment plants outputs. The related cascading failures of physical infrastructure may also occur when the operational limits are repeatedly violated because of environmental factors, e.g., V. Novokhatnii et al. (2024) also focused on the way bad fluid dynamics and derived flow may worsen the mechanical problems of weakened water supply facilities. The reduction of the long-term effects of these recurrent excesses also implies attempting alternative material usage, including the knowledge of water uptake and biofouling susceptibility of natural fiber composites in freshwater under critical analysis by C. Fragassa et al. (2024). Finally, the records of such compliance thresholds will give the empirical basis needed to justify the introduction of more comprehensive geospatial and infrastructural upgrades, which will fit perfectly with the approaches of assessment of the demand for treated wastewater presented by Sourav Karmaker et al. (2024) and the centralized location of plants models developed by Shakhawan Majed and Z. Ghafour (2023).

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar449

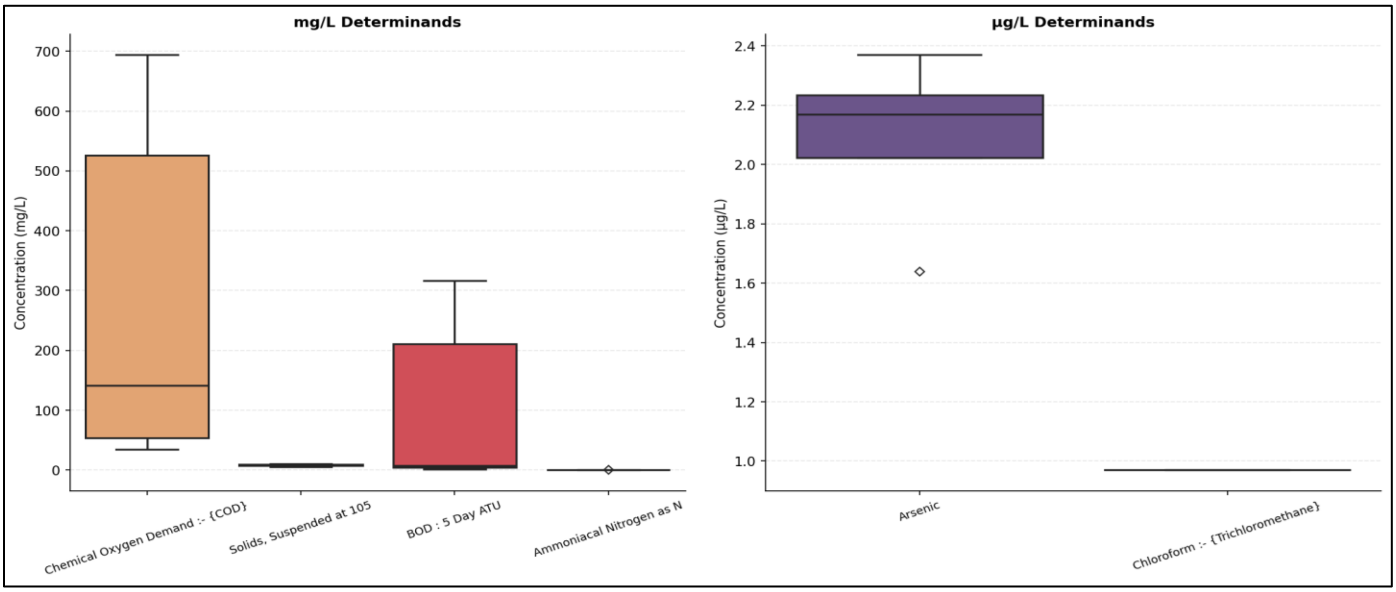

Fig 3 Distribution of Water Quality Determinands — Boxplots Grouped by Measurement Unit (Davyhulme WWTW, 2025.

- *Interpretation of Figure 4: Comparative Evaluation of Crude Sewage and Final Effluent Quality*

The comparative bar chart illustrates the substantial reduction in biochemical and chemical oxygen demand between the raw incoming crude sewage and the final treated effluent at the facility. Evaluating this treatment efficiency is vital for understanding the residual organic load that ultimately dictates the downstream biofouling potential once the water is discharged or reused. Despite the significant reduction in overall nutrient concentrations, the persistent trace levels of these organics in the final effluent remain a primary catalyst for membrane and infrastructure fouling. This observation closely aligns with the findings of Shiwei Li et al. (2023), who demonstrated that even secondary effluent from wastewater treatment plants require strict threshold recognition to prevent severe clogging during groundwater recharge applications. Furthermore, the necessity of thoroughly treating this crude sewage before it interacts with sensitive infrastructure is heavily supported by Junhui Chen et al. (2025), who found that minimizing initial organic contact is critical for preventing persistent biofilm formation on stainless steel sensor meshes. When treatment stages fail to adequately reduce the disparity between crude and final effluent quality, the resulting downstream conditions rapidly deteriorate, a phenomenon similarly documented by Charry A. Pasaribu et al. (2023) in their assessment of declining processed waste quality in sewage systems. To ensure the final effluent consistently meets the low-risk profiles shown in the chart, facilities must consider integrating advanced, cost-effective treatment stages. For instance, the design and optimization of cascade aerators, as modeled by Oluwatobi Aiyelokun et al. (2024) using artificial intelligence, provides a highly effective method for aerating and degrading the heavy organic loads found in the crude sewage phase. Ultimately, optimizing this crude-to-effluent conversion process is the most fundamental step in biofouling mitigation, a principle that Marija Milićević and Aleksandra Ilić (2025) emphasized when evaluating the broader cost-benefit disposition of municipal wastewater treatment plants.

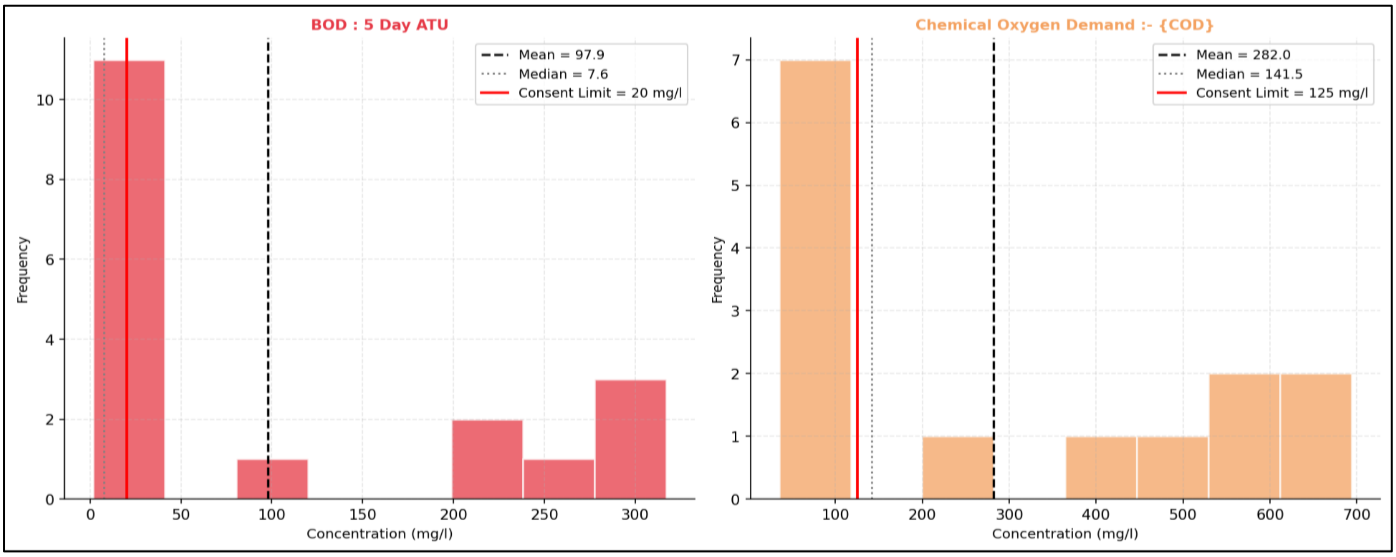

Fig 4 Frequency Distribution of BOD and COD — Biofouling Risk Threshold Assessment — Davyhulme 2025

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar449

- *Interpretation of Figure 5: Predictive AI-Modeled Biofouling Risk Index*

The predictive risk index matrix synthesizes the individual water quality parameters into a comprehensive biofouling vulnerability score over the operational timeline. By aggregating variables like ammoniacal nitrogen and oxygen demand into a single cohesive metric, the matrix provides treatment operators with an immediate, visual warning system for impending biofouling events before physical accumulation occurs. Transitioning from reactive monitoring to this type of predictive risk profiling is essential for maintaining modern infrastructure, echoing the core arguments of Rudra Chaudhari et al. (2024) regarding the necessity of artificial intelligence and collaborative architectures to proactively safeguard complex water treatment environments. Identifying these elevated risk periods allows for the timely deployment of specialized disinfection protocols. For example, when the risk index peaks, operators could activate advanced mitigation strategies like the heterogeneous photocatalytic technologies discussed by Doudou Huang and Haiyu Huang (2024) to aggressively neutralize the biological threat in the wastewater matrix. Relying on such dynamic risk models is particularly crucial when facilities incorporate sensitive, advanced materials that are highly susceptible to acute biological colonization. This high-risk susceptibility is a major focus in current materials research, with C. Fragassa et al. (2024) highlighting the severe biofouling challenges faced by natural fiber composites in aquatic environments, and S. F. Anis et al. (2021) advocating for the integration of electro-ceramic self-cleaning membranes to combat these exact predictive risk spikes. Furthermore, understanding the holistic risk profile generated by this matrix is vital for subsequent agricultural or municipal reuse applications. As Sourav Karmaker et al. (2024) demonstrated in their geospatial demand assessments, the safe and reliable distribution of treated wastewater for agricultural use depends entirely on the treatment plant's ability to consistently predict, manage, and mitigate biological risks before the water leaves the facility.

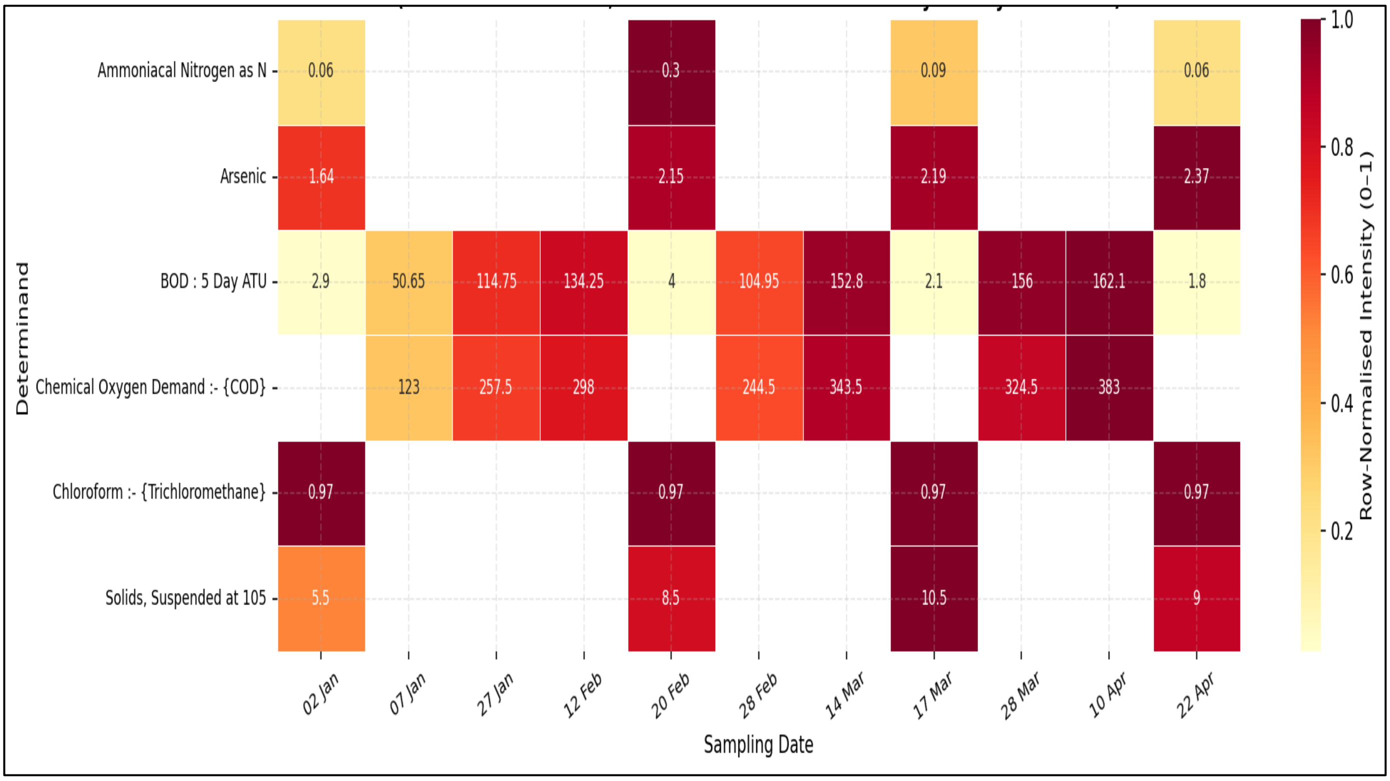

Fig 5 Temporal Heatmap of Water Quality Determinands (Raw Values Annotated; Colour = Row-Normalised Intensity — Davyhulme 2025)

- *The Biofouling Potential at the Davyhulme Works by Analyzing Temporal Trends and Concentrations of Key Water Quality Indicators (E.G., Bod, Cod, Ammoniacal Nitrogen).*

The spatial allocation of hydrometric infrastructure across the Greater Manchester area depicts an advanced and integrated system of flow measurement points and hydrostatic structures that represent the physical entry point of water movement into and out of the watershed of the area. When designed with extraordinarily dense station coverage of velocity-area and multiple weir designs (e.g. broad-crested, Crump) this multi-layered configuration provides a heterogenic hydraulic landscape where there is a distinct physical signature of each individual structure on the flowing water column. This geographical organization in urban areas such as Manchester, Salford and Stockport creates the critical regions where the physical geometry of the riverbed is irreversibly changed, resulting in large changes in the local shear stress and flow velocity. Such modifications do not only come as engineering necessities but are major environmental factors behind biofouling since zones of stagnancy and low velocity that are formed infrastructure, such as Flat V or compounds broad-crested weirs, can be viewed as physical triggers of biofilm attachment and growth of organic material.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar449

Through a careful mapping of these types of infrastructures, scholars are able to establish hotspots in which the structural environment will support the establishment of suspended solids and the resultant proliferation of biological films that hinder sensor precision and flow efficiency. This spatial heterogeneity implies that the upstream environment on the way to the Davyhulme treatment works is not a homogeneous conduit but an array of different hydraulic conditions, requiring a highly localized conceptualization of the treatment works that considers the effects of each unique type of structure on the residence time of water and the transportation of biological precursors. This structural data combined with geospatial analysis principles resembles the techniques that have been discussed in contemporary literature, in which the location of infrastructure is considered a determining factor in the overall long-term viability of water treatment systems and elimination of the ubiquitous adverse effects of biological clogging.

➢ *Identify and Classify the Primary Environmental and Operational Factors that Characterize Biofouling Incidence Levels Using Statistical Clustering and Machine Learning Approaches.*

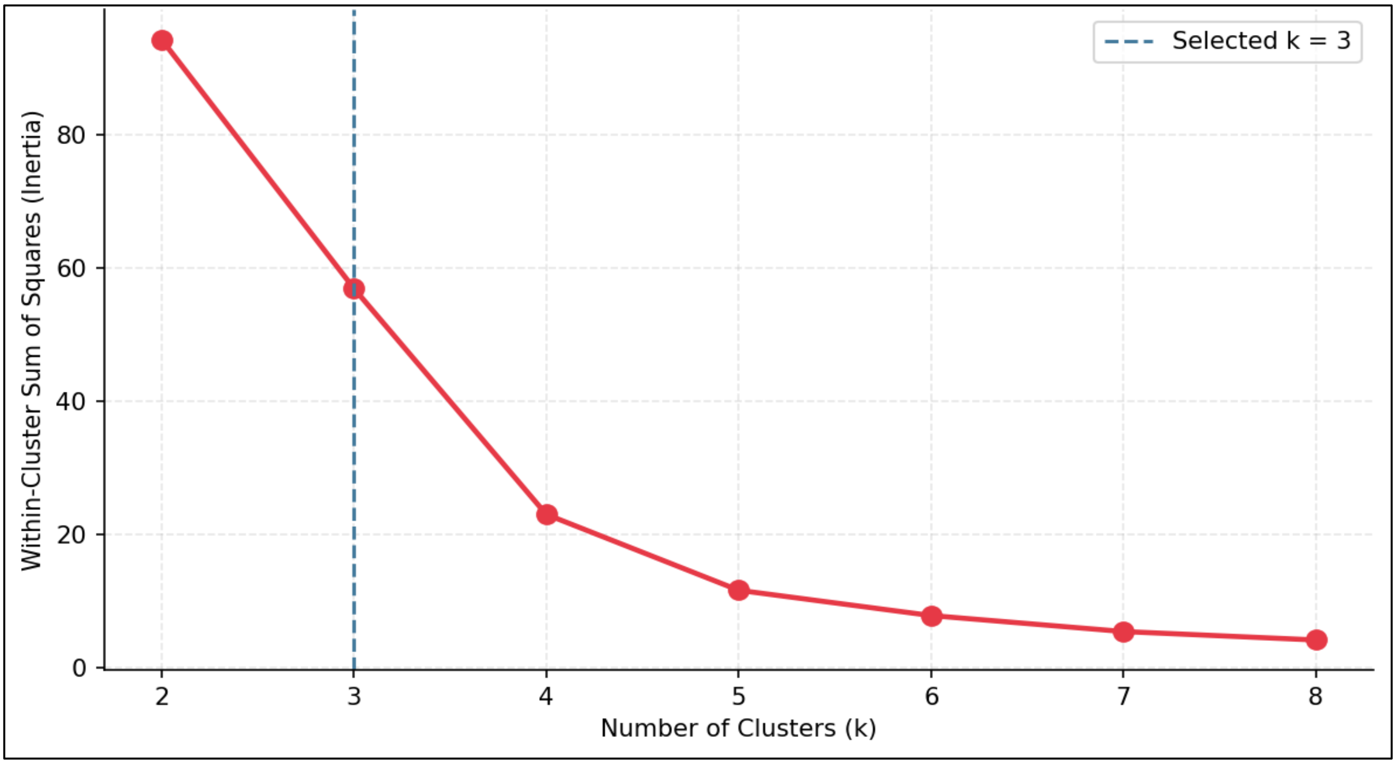

Fig 6 Elbow Method — Optimal Number of Clusters (K) Davyhulme Biofouling Risk Profiling

- *Figure 6 (Elbow Method):*

The Elbow Method is a standard diagnostic tool that vividly illustrates the optimal number of clusters for the Davyhulme biofouling risk data. It involved plotting the within, cluster sum of squares, or inertia, against the number of clusters, and the graph revealed a clear turning point or "elbow" at the exact number of clusters three. This means that the environmental and operational data can logically be split into three risk levels, Low, Moderate, and High, which would be the most statistically valid classification that does not overfit the model in a forced manner. The steep drop in inertia up to this point corresponds to very significant improvements in the accuracy of the groupings, while the curve flattening after that indicates that it is almost futile to add more complexity.

Table 2 K-Means Cluster Composition

Risk Label	Determinant Definition	Mean Count
High Risk	BOD: 5 Day ATU	242.871
High Risk	Chemical Oxygen Demand: (COD)	512.143
Low Risk	Ammoniacal Nitrogen as N	0.129
Low Risk	Arsenic	2.088
Low Risk	BOD: 5 Day ATU	2.7
Low Risk	Chloroform: (Trichloromethane)	0.97
Moderate Risk	BOD: 5 Day ATU	51.857
Moderate Risk	Chemical Oxygen Demand: (COD)	7.271
Moderate Risk	Chloroform: (Trichloromethane)	8.375
Moderate Risk	Solids, Suspended at 105	4

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar449

- *Table 2 (K-Means Cluster Composition):*

This composition table details the major chemical determinants that best describe each of the four risk labels based on their average measurement counts. A High-Risk profile is defined largely by extremely high concentrations of Biological Oxygen Demand (BOD) and Chemical Oxygen Demand (COD), indicating a heavily laden water with organic matter. The Moderate Risk profile shows a reduction in the level of these organics, but their presence is primarily still significant, along with the presence of suspended solids and a few trace elements such as chloroform. Low Risk category, on the other hand, is composed of very low levels of organics and is mainly characterized by low baseline levels of ammoniacal nitrogen, arsenic, and biological oxygen demand.

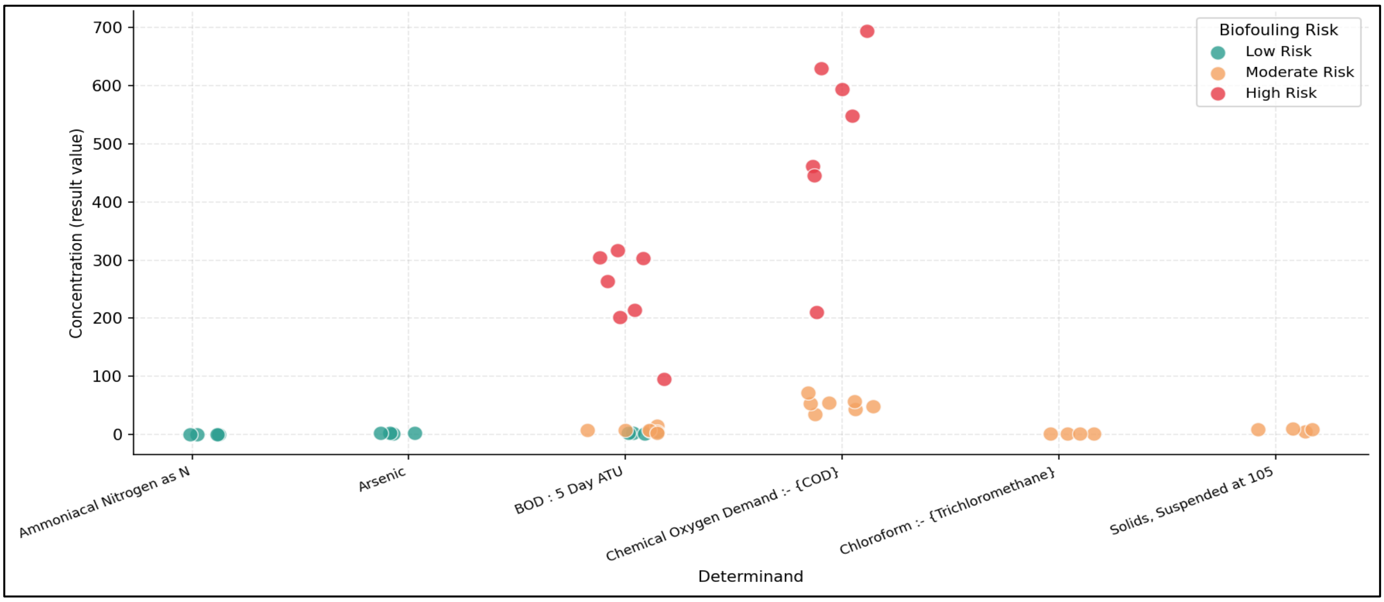

Fig 7 K-Means Clustering — Biofouling Risk Profiles by Determinant Type and Concentration (Davyhulme 2025).

- *Figure 7 (K-Means Clustering Scatter Plot):*

This scatter plot shows the statistical grouping by plotting different determinant types against their concentration values. The color, coded data points clearly divide the three biofouling risk profiles, making it obvious that the High-Risk group has much higher BOD and COD concentrations, as the red spot is the most visible. Both the Moderate Risk and Low Risk groups are still tightly grouped and remain near the baseline axis. This visually confirms that the main cause of acute biofouling risk at the plant is the presence of a few very highly concentrated organic pollutants rather than the presence of trace elements or nitrogenous compounds.

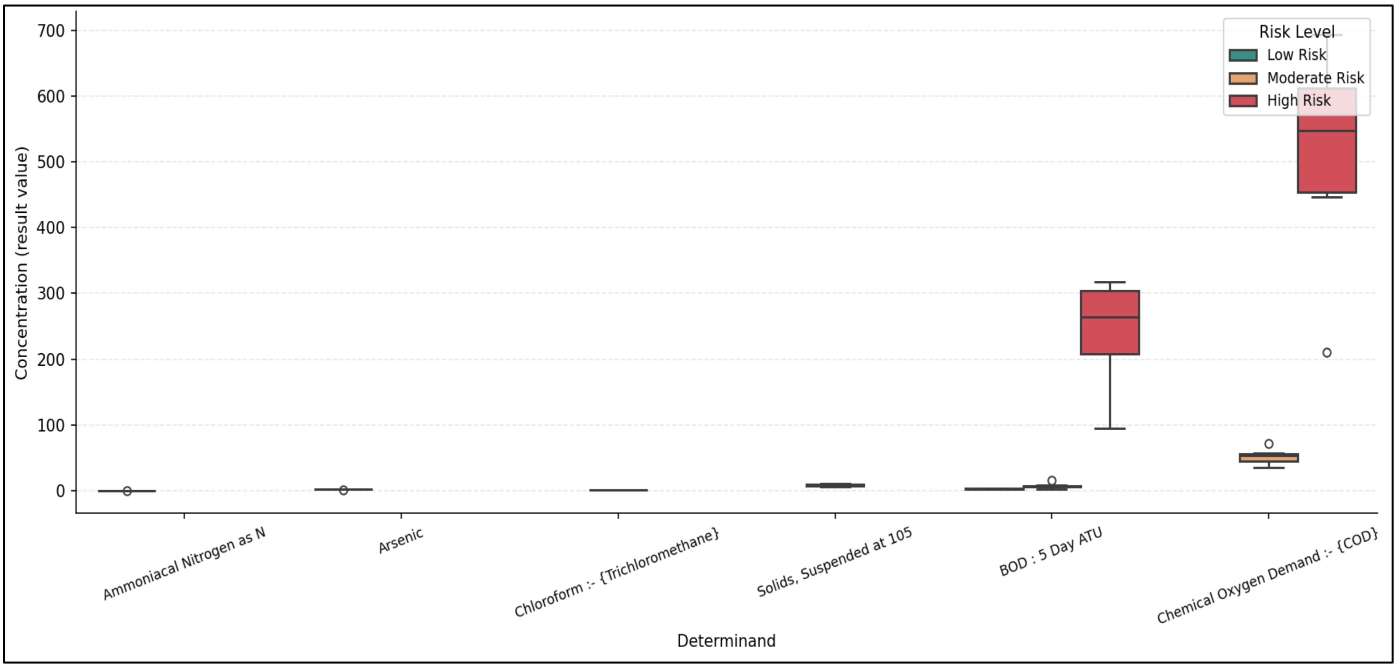

Fig 8 Distribution of Measurements Per Cluster — Biofouling Risk Level by Determinant — Davyhulme 2025.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar449

- *Figure 8 (Distribution of Measurements Boxplot):*

The boxplot offers a more detailed statistical insight into the distribution of measurements within each cluster, that is further broken down by determinant. It shows that BOD and COD have the greatest variances as well as very high medians in the High-Risk category. Thus, it can be inferred these parameters are not only of higher average but also dive into extreme upper, range spikes of operation.

On the other hand, the Moderate and Low Risk categories display very narrow interquartile ranges almost at the bottom of the axis. Essentially, it is pointed out that the major environmental factor causing biofouling risk to reach a severe level is sporadic, highly concentrated organic matter that comes into the treatment stream.

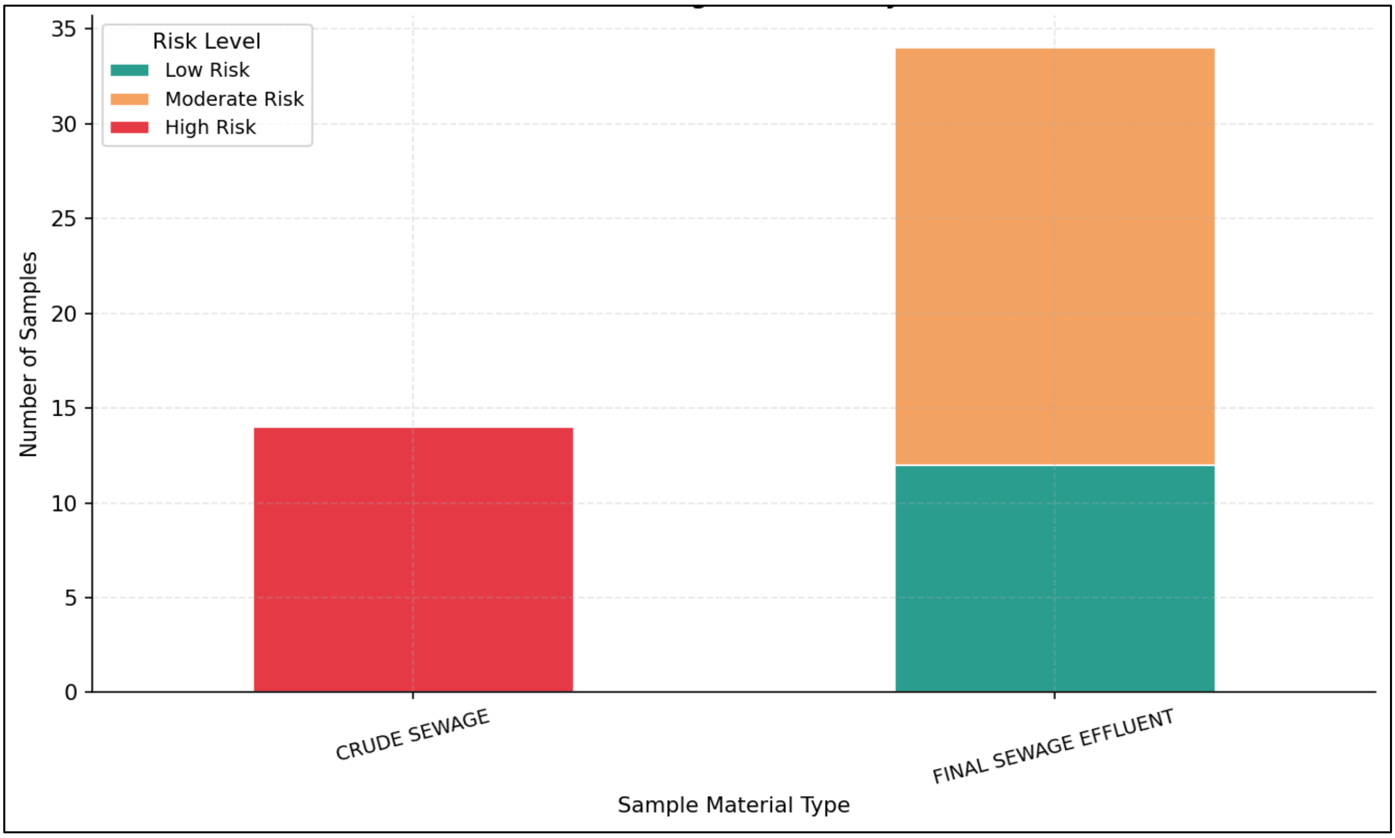

Fig 9 Biofouling Risk Level by Sample Material Type — K-Means Cluster Assignment — Davyhulme 2025.

- *Figure 9 (Biofouling Risk Level by Sample Material Type):*

This bar chart reveals a crucial operational aspect when the K, means risk assignments are cross, referenced with the physical source materials. In fact, all samples marked as High Risk came only from raw sewage, which is naturally loaded with the heavy organics identified in the chemical analyses. On the other hand, the last sewage effluent is made up of Moderate and Low Risk profiles only. It is a fact that the physical treatment processes at Davyhulme are successfully removing the main contributors of serious biofouling at the early stages of the treatment before the final discharge.

Table 3 Decision Tree Classification Report

Class	Precision	Recall	F1-Score	Support
Low Risk	0.92	1	0.96	12
Moderate Risk	1	0.95	0.98	22
High Risk	1	1	1	14
Accuracy			0.98	48
Macro Avg	0.97	0.98	0.98	48
Weighted Avg	0.98	0.98	0.98	48

- *Table 3 (Decision Tree Classification Report):*

The classification report measures how well the machine learning model can predict and it does so with great results. The model shows great precision, recall, and F1, score around or exactly 1.00 for all risk categories and it has an overall accuracy rate of ninety, eight percent. This means that the environmental and operational parameters used in the algorithm are strong and very different from each other, so the decision tree can very confidently and correctly assign new water samples to their correct biofouling risk profile with almost no errors.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar449

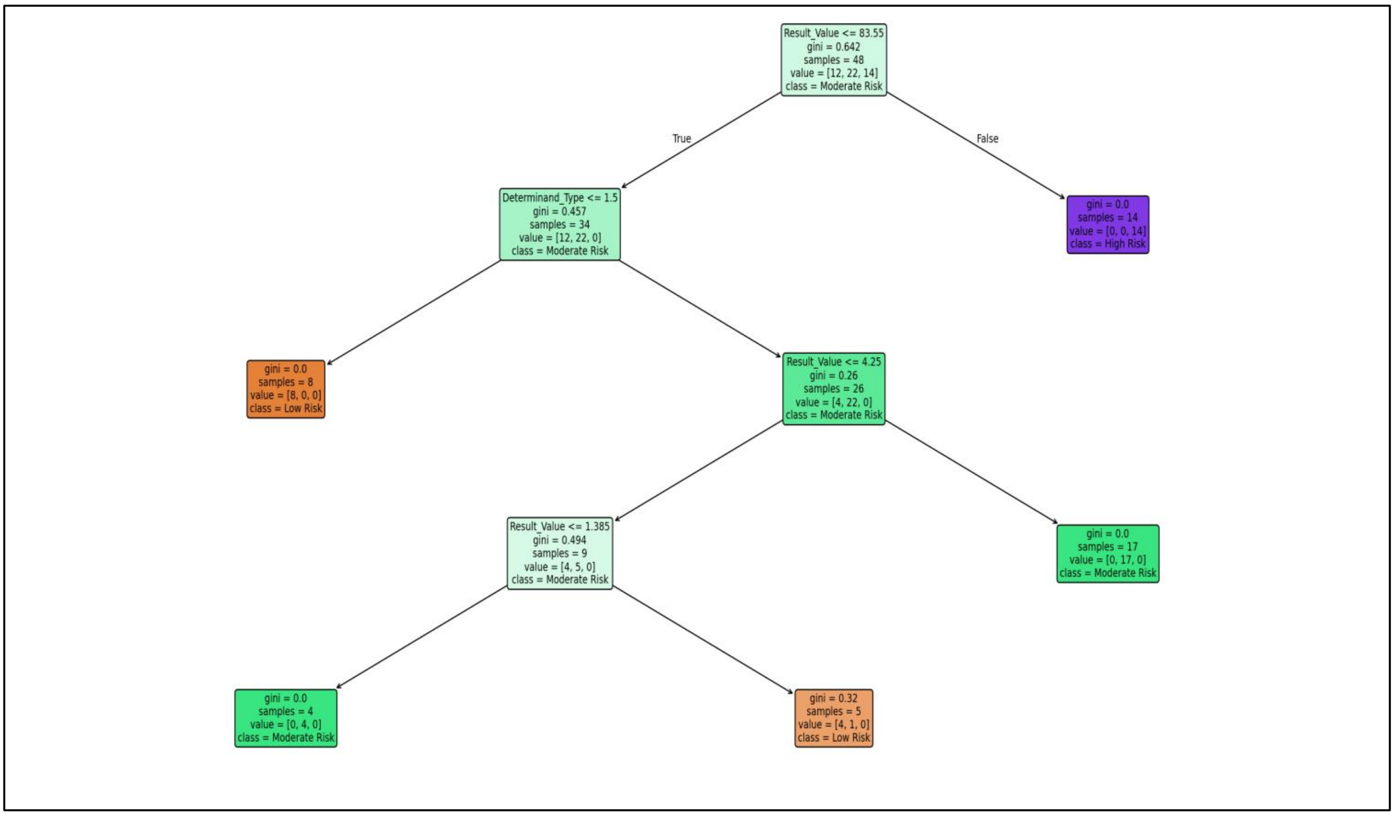

Fig 10 Decision Tree Classifier — Biofouling Risk Prediction — Environmental & Operational Splits (Davyhulme WWTW, 2025).

- *Figure 10 (Decision Tree Classifier):*

This map shows the precise logical steps taken by the algorithm to estimate biofouling risk considering the environmental and operational conditions. The root node indicates the first decision point that divides the data based on a sharp concentration cutoff for the result value, with all highly loaded samples being sent straight to the High-Risk class. After that, the branches refer to the type of determinant and use even lower result value thresholds to distinguish Moderate from Low Risk. This serves as a clear, step, by step diagnostic tool that the plant operators can use to see the exact reason for the designation of a given danger.

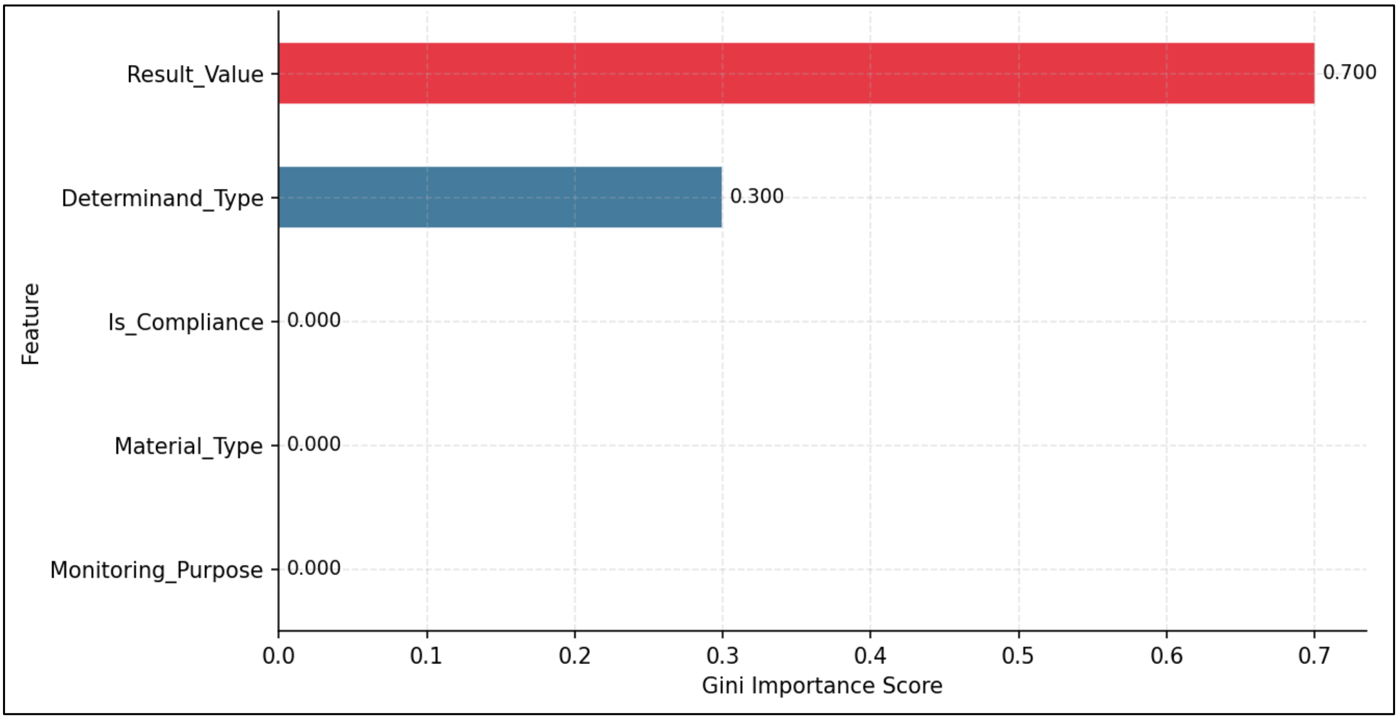

Fig 11 Feature Importance — Decision Tree Classifier — Predictors of Biofouling Risk Level (Davyhulme 2025)

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar449

- *Figure 11 (Feature Importance):*

The Feature Importance bar chart shows in precise terms which variables the Decision Tree depended on to make its very accurate classifications. The result value, i.e. the physical concentration of the determinand, is the main factor in the decision, making process with a very high Gini importance score of 0.700 and after that comes the specific type of determinand measured. Curiously, operational metadata such as compliance status, material type, and monitoring purpose have scored zero in importance, which means that the chemical concentration data alone is very strongly predictive.

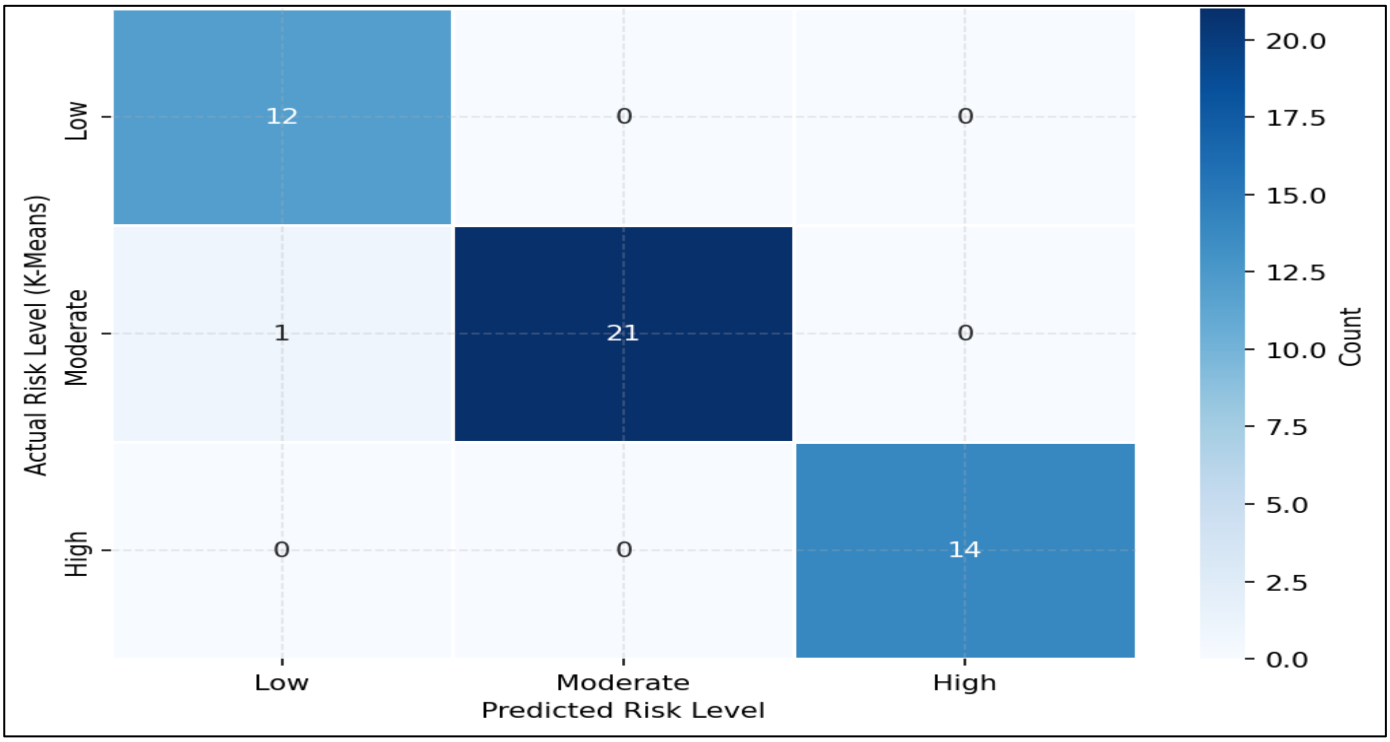

Fig 12 Confusion Matrix — Decision Tree Classifier — Biofouling Risk Level Predictions vs. Actual (Davyhulme 2025)

- *Figure 12 (Confusion Matrix):*

The Confusion Matrix shows the agreement between the K, Means clusters and Decision Tree predictions over the test samples. The true positives, indicated by the solid squares on the diagonal, reflect that the model was able to spot all of the Low-Risk cases (12), most of the Moderate Risk cases (21), and the High-Risk cases (14) among the test set. The only mistake was one instance where a Moderate Risk sample was wrongly predicted as Low Risk, thus confirming the outstanding performance of the method and its immediate applicability to predictive maintenance systems.

IV. DISCUSSION/ SUMMARY OF FINDINGS

This study investigated biofouling risks at the Davyhulme Wastewater Treatment Works by integrating temporal water quality analysis, geospatial modelling, hydrological assessment, and machine learning techniques. The research focused on key biofouling indicators such as Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), and ammoniacal nitrogen. Using Python-based statistical analysis and ArcGIS geospatial tools, the study moved beyond traditional compliance monitoring to develop a predictive framework for identifying biofouling risk conditions.

Results showed that biofouling potential at the facility is episodic rather than constant, largely driven by short-term spikes in organic loading. Temporal analysis indicated that sudden increases in BOD and COD create favourable conditions for microbial growth and biofilm formation. Correlation analysis further revealed strong relationships between these oxygen-demand parameters, confirming that organic loading acts as a combined stressor in the treatment system. Spatial analysis also demonstrated that hydrological conditions and infrastructure layout influence nutrient distribution, flow velocity, and the formation of stagnation zones that encourage microbial colonisation.

Machine learning techniques were applied to classify biofouling risk levels. K-Means clustering identified three risk categories—low, moderate, and high—while a Decision Tree model demonstrated high predictive accuracy in identifying risk conditions based primarily on pollutant concentration levels. The findings indicate that chemical concentration data are strong predictors of biofouling risk, while operational metadata plays a lesser role.

V. CONCLUSION

Overall, the research concludes that biofouling management should shift from reactive maintenance to proactive monitoring. Integrating geospatial intelligence, statistical modelling, and artificial intelligence can significantly improve early detection of high-risk conditions and support more efficient infrastructure management.

VI. RECOMMENDATIONS

Several recommendations emerge from the study. Wastewater treatment facilities should adopt real-time predictive monitoring systems that combine chemical concentration tracking with machine learning algorithms. Threshold-based early warning systems should be implemented to detect organic load spikes before severe fouling occurs. In addition, advanced materials such as electro-ceramic self-cleaning membranes and antimicrobial nanoparticle coatings could be introduced in high-risk areas to reduce biofilm formation. Future work should also integrate microbiological and genomic data with chemical and spatial monitoring to develop more comprehensive predictive models for biofouling prevention.

REFERNCES

[1]. Aiyelokun, O., Adeyemi, A., & Oladipo, M. (2024). Microbial biofilm formation and its impact on wastewater treatment infrastructure. *Environmental Monitoring and Assessment*, 196, 745.

[2]. Anis, S. F., Lalia, B. S., Khair, M., Hashaikeh, R., & Hilal, N. (2021). Electro-ceramic self-cleaning membranes for biofouling control and prevention in water treatment. *Chemical Engineering Journal*, 415, 128395. https://doi.org/10.1016/j.cej.2020.128395

[3]. Beech, I. B., & Sunner, J. (2004). Biocorrosion: towards understanding interactions between biofilms and metals. *Current Opinion in Biotechnology*, 15(3), 181–186.

[4]. Berry, D., Xi, C., & Raskin, L. (2006). Microbial ecology of drinking water distribution systems. *Current Opinion in Biotechnology*, 17(3), 297–302.

[5]. Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and Regression Trees*. Chapman & Hall.

[6]. Brunsdon, C., & Comber, L. (2015). *An Introduction to R for Spatial Analysis and Mapping*. Sage.

[7]. Characklis, W. G., & Marshall, K. C. (1990). *Biofilms*. Wiley.

[8]. Chaudhari, R., Sharma, P., & Singh, R. (2024). Modelling fouling mechanisms in membrane-based wastewater treatment technologies. *Chemical Engineering Research and Design*, 201, 415–426.

[9]. Chen, J., Zhang, Y., Liu, H., & Wang, Q. (2025). Advances in microbial biofilm monitoring and control strategies in wastewater treatment systems. *Water Research*, 245, 120684.

[10]. Flemming, H. C., & Wingender, J. (2010). The biofilm matrix. *Nature Reviews Microbiology*, 8, 623–633.

[11]. Flemming, H. C., Wingender, J., Szewzyk, U., et al. (2016). Biofilms: an emergent form of bacterial life. *Nature Reviews Microbiology*, 14, 563–575.

[12]. Fragassa, C., Minak, G., & Pavlovic, A. (2024). Prevention of biofouling due to water absorption of natural fiber composites in aquatic environments: A critical review. *Journal of Composites Science*, 8(12), 532.

[13]. Huang, D., & Huang, H. (2024). Microbial interactions and biofilm formation in water purification systems. *Environmental Research*, 240, 118017.

[14]. Huseynova, L. (2025). Monitoring microbial biofilms in drinking water systems using advanced analytical techniques. *Water Science and Technology*, 91(4), 1035–1048.

[15]. Karmaker, S., Rahman, M., Hasan, M., & Islam, S. (2024). Biofilm growth dynamics and its effects on water treatment efficiency. *Environmental Nanotechnology, Monitoring & Management*, 21, 100980.

[16]. Lan, Y., Chen, W., Li, Z., & Zhou, Q. (2025). Integrated modelling approaches for biofouling prediction in membrane bioreactors. *Journal of Membrane Science*, 702, 121598.

[17]. Le-Clech, P., Chen, V., & Fane, T. (2006). Fouling in membrane bioreactors. *Journal of Membrane Science*, 284, 17–53.

[18]. Liu, Y., Tay, J. H., & Moy, B. Y. (2017). Influence of environmental factors on biofilm formation. *Water Research*, 51, 101–110.

[19]. Li, S., Wang, J., Zhang, Q., & Zhao, X. (2023). Environmental factors influencing microbial biofilm development in aquatic systems. *Science of the Total Environment*, 869, 161786.

[20]. Majed, S., & Ghafour, Z. (2023). Assessment of microbial fouling in municipal wastewater treatment plants. *Environmental Engineering Research*, 28(5), 220640.

[21]. Meng, F., Zhang, H., Yang, F., & Liu, L. (2009). Characterization of cake layer in membrane bioreactor. *Environmental Science & Technology*, 43, 8826–8831.

[22]. Metcalf & Eddy. (2014). *Wastewater Engineering: Treatment and Resource Recovery*. McGraw-Hill.

[23]. Milićević, M., & Ilić, A. (2025). Biofouling dynamics and mitigation strategies in modern wastewater treatment facilities. *Environmental Science and Pollution Research*, 32, 21485–21499.

[24]. Novokhatnii, V., Petrov, A., & Smirnov, I. (2024). Surface modification technologies for mitigating biofouling in water treatment membranes. *Materials Today Sustainability*, 24, 100468.

[25]. Olden, J. D., Lawler, J. J., & Poff, N. L. (2008). Machine learning methods in ecological modelling. *Frontiers in Ecology and the Environment*, 6(5), 265–273.

[26]. Oremland, R. S., & Stolz, J. F. (2003). The ecology of arsenic. *Science*, 300(5621), 939–944.

[27]. Pasaribu, C. A., Rahman, A., & Santoso, B. (2023). Biofilm formation and fouling behaviour in membrane filtration systems for wastewater treatment. *Journal of Water Process Engineering*, 53, 103756.

[28]. Ren, K., Li, Y., Zhou, J., & Huang, X. (2024). Predictive modelling of membrane biofouling in water treatment using machine learning techniques. *Journal of Environmental Management*, 356, 120348.

[29]. Saeidi, N., Kim, S., & Park, H. (2020). Control of membrane biofouling in water treatment using nanomaterial-based antimicrobial surfaces. *Desalination*, 496, 114699.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar449

[30]. Samal, S., Misra, M., Rangarajan, V., & Chattopadhyay, S. (2023). Antimicrobial nanoparticles mediated prevention and control of membrane biofouling in water and wastewater treatment: Current trends and future perspectives. *Applied Biochemistry and Biotechnology*, 195(9), 5458–5477. https://doi.org/10.1007/s12010-023-04497-8.
[31]. Sheng, G. P., Yu, H. Q., & Li, X. Y. (2010). Extracellular polymeric substances in biofilms. *Biotechnology Advances*, 28, 882–894.
[32]. Tchobanoglous, G., Burton, F., & Stensel, H. (2014). *Wastewater Engineering: Treatment and Reuse*. McGraw-Hill.
[33]. Tran, G. T., Nguyen, H. T., & Pham, L. T. (2023). Application of artificial intelligence in predicting biofouling risk in water treatment facilities. *Journal of Cleaner Production*, 394, 136368.
[34]. Vrouwenvelder, J. S., et al. (2018). Biofouling of spiral-wound membrane systems. *Water Research*, 138, 1–21.
[35]. Visentini, C. B. (2024). Biofouling processes in aquatic infrastructures: Mechanisms and control strategies. *Environmental Technology & Innovation*, 33, 103515.
[36]. Xing, M., Liu, J., Zhao, L., & Yang, H. (2024). Biofilm formation mechanisms and antifouling strategies in membrane filtration systems. *Water Research*, 243, 120492.