Journal of Information Systems Engineering and Management
2025, 10(7s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A Novel Artificial Neural Network Approach for Troubleshooting of Sewage Treatment Plant Process

1Mr. D.K. Khopade, 2Prof. S. G. Ban

[1]Research Scholar, Department of Civil Engineering, Department of Civil Engineering, GH Raisoni University, Amravati.
dkkhopade@gmail.com

2Professor, Department of Civil Engineering, Department of Civil Engineering, GH Raisoni University, Amravati.
shyambhau.ban@raisoni.net

ARTICLE INFO

Received: 12Oct 2024

Revised: 08Dec 2024

Accepted: 23Dec 2024

ABSTRACT

Efficient operation and proactive fault diagnosis in Sewage Treatment Plants (STPs) are essential for maintaining water quality and ensuring regulatory compliance. This paper presents a detailed data analysis of two STP plants, focusing on key operational parameters, including influent and effluent characteristics, aeration rates, and environmental conditions. The target features of the study—Effluent BOD (mg/L), Effluent COD (mg/L), Effluent TSS (mg/L), and Operational Issue Detected—are critical indicators of the plant's treatment efficiency and potential operational issues. Using synthetic datasets, we conducted comprehensive exploratory data analysis (EDA) to identify correlations between influent parameters and effluent quality, revealing important patterns that influence plant performance. Visualizations, including correlation heat maps and scatter plots, provided insights into key factors affecting effluent quality and operational issues. Furthermore, the study demonstrates the potential of Artificial Neural Networks (ANNs) in predicting these target features, offering a predictive framework for fault detection and diagnosis in STPs. By integrating ANN-based models, this research contributes to improving predictive maintenance strategies, ensuring optimal performance, and enhancing decision making in sewage treatment operations. The findings underscore the value of data-driven approaches in optimizing the management and sustainability of wastewater treatment systems.

Keywords: Sewage Treatment Plant, Fault Diagnosis, Data Analysis, Artificial Neural Network, Machine Learning, Effluent Quality, BOD, COD, TSS, Operational Issues, Predictive Maintenance.

INTRODUCTION

Sewage Treatment Plants (STPs) are essential for safeguarding environmental and public health by treating wastewater before it is released back into natural water bodies. As urban populations continue to grow and industrialization expands, the challenges of managing and optimizing STP operations have become more pronounced. With diverse influent quality, varying environmental conditions, and stringent effluent standards, it is increasingly critical to ensure STPs operate efficiently and diagnose faults in real-time. Ensuring optimal treatment processes and addressing operational issues such as high BOD (Biochemical Oxygen Demand), COD (Chemical Oxygen Demand), TSS (Total Suspended Solids), and ammonia nitrogen in effluent is essential for maintaining compliance and operational performance. Traditionally, the operation of STPs has relied heavily on manual inspections, historical data analysis, and operator expertise. While these methods are necessary, they tend to be reactive, time consuming, and susceptible to human error. Furthermore, identifying the root causes of operational inefficiencies or failures often requires substantial downtime and labour-intensive troubleshooting. These limitations call for more advanced, automated solutions that can monitor the plant's performance continuously and provide timely diagnosis.

In recent years, artificial intelligence (AI) and machine learning (ML) techniques have gained significant traction in various industries, including wastewater treatment. Among these techniques, Artificial Neural Networks (ANNs) have proven to be highly effective in capturing complex relationships between input and output variables. ANNs can analyze large datasets, uncover hidden patterns, and generate predictive models, making them an ideal tool for real-time fault detection and predictive maintenance in STP operations. By leveraging these data-driven techniques, it becomes possible to predict potential operational issues before they manifest, leading to more efficient and costeffective treatment processes. This study explores the application of ANNs in the context of fault detection and predictive maintenance in two distinct STP plants. Through a comprehensive analysis of operational parameters and effluent quality indicators, the research aims to uncover the factors influencing treatment performance and operational issues. The study specifically focuses on the prediction of key effluent quality parameters—Effluent BOD, Effluent COD, and Effluent TSS—as well as the detection of operational issues. These parameters serve as crucial indicators of the plant's ability to meet discharge standards and maintain efficient operations.

The analysis begins with a detailed examination of synthetic datasets, where relationships between influent characteristics (such as flow rate, pH, temperature, BOD, COD, and TSS) and effluent quality parameters are explored. By applying ANNs, this research aims to develop a predictive framework that can forecast effluent quality and identify operational issues based on influent parameters and environmental conditions. This data-driven approach promises to enhance operational decision-making, reduce unplanned downtimes, and optimize the overall performance of STPs. The application of AI, specifically through ANN-based models, represents a significant advancement in the operation and management of sewage treatment plants. It not only facilitates the detection of operational faults before they escalate but also aids in improving the sustainability and efficiency of the treatment process. This paper demonstrates the potential for AI-driven solutions to transform the management of STPs, ensuring that plants operate within optimal parameters, minimize environmental impact, and contribute to more sustainable water treatment practices.

This research emphasizes the value of integrating machine learning techniques with traditional STP processes, creating a more efficient and proactive framework for maintaining and optimizing sewage treatment systems. The insights gained from this study have the potential to revolutionize the way STPs are operated and maintained, contributing to improved water quality, reduced operational costs, and enhanced decision-making in wastewater treatment.

METHODOLOGY

The methodology for this study involves a systematic approach to data collection, pre-processing, model development, and evaluation to predict effluent quality and detect operational issues in Sewage Treatment Plants (STPs) using Artificial Neural Networks (ANNs). The first step in the methodology involves the collection of historical data from two different STPs. The data includes influent characteristics such as flow rate, BOD, COD, TSS, pH levels, and aeration rate, as well as operational parameters like sludge retention time (SRT), temperature, and aeration parameters. The target variables for prediction are effluent parameters, including Effluent BOD, Effluent COD, Effluent TSS, and the operational status of the plant, such as the detection of any faults or issues.

Once the data is collected, it undergoes pre-processing to ensure its suitability for model training. This includes handling missing values, scaling the data, and encoding categorical variables where necessary. Statistical methods, such as correlation analysis, will be used to identify relationships between the influent and effluent variables, ensuring that the ANN model is trained on the most relevant features. The ANN model will be designed with multiple layers, including an input layer, one or more hidden layers, and an output layer corresponding to the effluent prediction variables. Different ANN architectures, such as feed forward neural networks (FNN) or recurrent neural networks (RNN), will be explored to determine which is most effective for the problem at hand. The model will be trained using a back propagation algorithm, with the training process evaluated using performance metrics like Mean Squared Error (MSE) and R-squared values.

Once trained, the model will be validated and tested using unseen data to assess its generalization ability. Various performance metrics, including accuracy, precision, recall, and F1-score, will be calculated to evaluate the model's ability to predict effluent parameters and detect operational issues. The model will be fine-tuned using hyper parameter optimization techniques to improve its predictive power. To evaluate the feasibility and effectiveness of the ANN model, a comparison will be made between the model’s predictions and actual data from the STPs. Additionally, the results will be analyzed to identify trends, anomalies, and patterns that could indicate operational inefficiencies or potential faults within the treatment process.

The final step involves the development of a decision support system based on the ANN model, which can assist STP operators in making real-time decisions regarding operational adjustments, fault detection, and process optimization.

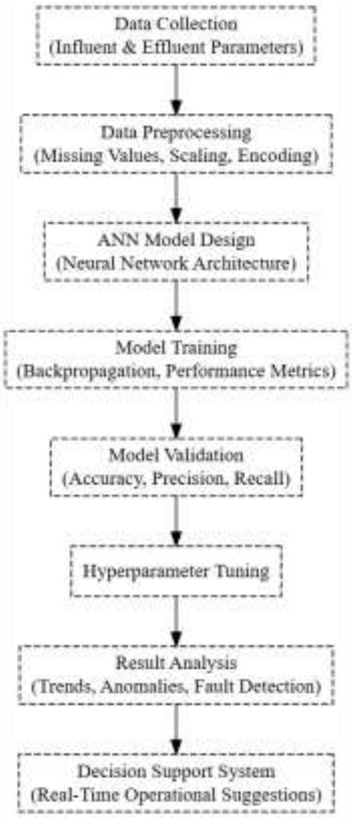

Fig 1.1: Methodology

DATA ANALYSIS

The system architecture is designed to illustrate the flow of data across multiple layers, from collection to analysis. It begins with the data collection layer, where raw data is gathered from various sources. This collected data is then transferred to the data storage layer, which consists of a centralized database for primary data storage, with an optional cloud storage component for backup and redundancy. Once the data is stored, the data processing layer takes over, where critical steps like data cleansing, feature engineering, and outlier detection are performed. These processes ensure the data is prepared for analysis, improving the quality and relevance of the dataset.

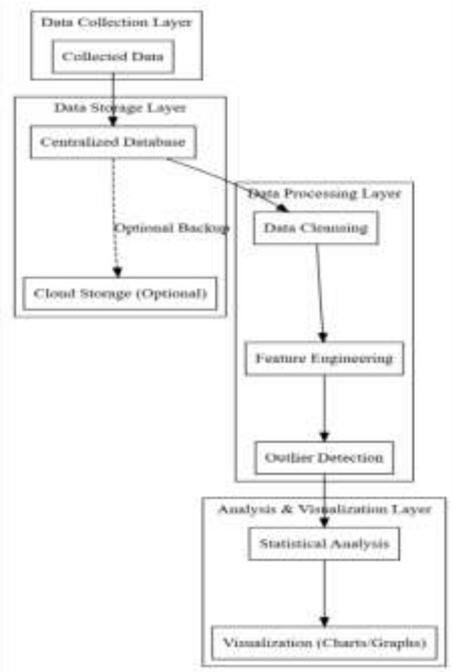

Fig 1.2: Analysis System Architecture

The final stage is the analysis and visualization layer, where statistical analysis and visualizations (in the form of charts and graphs) are performed to derive insights from the processed data. The system architecture emphasizes the smooth flow of data through each stage, with the inclusion of optional cloud storage ensuring backup and reliability. Data flows sequentially through each layer, ensuring that the system efficiently handles large datasets and prepares them for in-depth analysis. Below Figure1.2 is a conceptual representation of the Analysis system architecture

DATA OVERVIEW

In this project, data from two Sewage Treatment Plants (STPs) was collected to analyze various influent and effluent parameters. The first STP dataset includes key parameters such as influent flow rate, BOD (biological oxygen demand), COD (chemical oxygen demand), TSS (total suspended solids), pH level, temperature, ammonia nitrogen, aeration rate, sludge retention time, and effluent parameters. The second dataset includes similar parameters, such as influent flow rate, BOD, COD, TSS, aeration dissolved oxygen (DO), temperature, and effluent quality, alongside fault type information.

The target variables for analysis and prediction are effluent BOD, COD, TSS, and operational issues detected.

Table 1.1: Data Overview

Flow Rate (m^3/day)	Influent BOD (mg/L)	Influent COD (mg/L)	Influent TSS (mg/L)	Influent pH	Aeration DO (mg/L)	SRT (days)	Temperature (°C)	Effluent BOD (mg/L)	Effluent COD (mg/L)	Effluent TSS (mg/L)	Effluent pH	Fault Type
925.121918	164.972148	400.058010	178.684789	7.640065	2.479141	29.787601	20.304537	19.119057	48.085570	26.432196	7.994626	Normal
454.051565	349.346312	714.120118	201.225596	7.038517	3.510924	29.401138	24.125517	44.918280	103.751192	28.896280	7.339814	Normal
396.437330	302.848498	677.961245	130.969433	6.906776	3.345730	26.731045	19.665463	38.548483	102.560493	17.300465	6.926588	Normal
550.751824	338.492234	709.020414	166.397372	7.017073	4.825476	9.785514	29.402650	34.019817	80.442764	16.759866	6.983412	Normal
346.529442	312.686502	658.773168	274.245058	6.629486	1.880689	20.569962	13.037677	36.900331	120.258240	29.444367	6.816954	Normal

DATA PRE-PROCESSING

Prior to analysis, the dataset underwent pre-processing to handle missing values, scale numerical features, and encode categorical variables. Missing data points were imputed using mean imputation for numerical variables and mode imputation for categorical variables. Scaling was applied to ensure that all features were within the same range, making them compatible for machine learning algorithms. Categorical variables such as operational issues and fault types were encoded using one-hot encoding.

Table 1.2: Data Columns Summary

Sr. No.	Column	Non-Null Count	D type
0	**Flow Rate (m^3/day)**	4538 non-null	float64
1	**Influent BOD (mg/L)**	4538 non-null	float64
2	**Influent COD (mg/L)**	4538 non-null	float64
3	**Influent TSS (mg/L)**	4538 non-null	float64
4	**Influent pH**	4538 non-null	float64
5	**Aeration DO (mg/L)**	4538 non-null	float64
6	**SRT (days)**	4538 non-null	float64
7	**Temperature (°C)**	4538 non-null	float64
8	**Effluent BOD (mg/L)**	4538 non-null	float64
9	**Effluent COD (mg/L)**	4538 non-null	float64
10	**Effluent TSS (mg/L)**	4538 non-null	float64
11	**Effluent pH**	4538 non-null	float64
12	**Fault Type**	4538 non-null	object

DESCRIPTIVE STATISTICS AND INITIAL INSIGHTS

To gain a deeper understanding of the data distribution and relationships between different parameters, descriptive statistics were calculated for both the influent and effluent datasets. This analysis was pivotal in revealing important characteristics of the data, such as the range, mean, median, and standard deviation, for each parameter. By examining these statistics, we were able to identify patterns, potential outliers, and any emerging trends in the wastewater treatment process. The range provided insights into the variability of each parameter, highlighting the difference between the minimum and maximum values observed. The mean and median values helped assess the central tendency of the data, while the standard deviation revealed the degree of spread or dispersion around the mean. For instance, parameters like Influent Flow Rate exhibited a wide range, suggesting fluctuations in the volume of wastewater treated. Meanwhile, the relatively consistent values for pH level and Temperature indicated stable conditions for the treatment process. Moreover, examining these statistics allowed us to detect anomalies or outliers. For example, certain high or low values that deviated significantly from the mean could indicate irregularities or exceptional events in the wastewater inflow, which might require further investigation. This statistical analysis not only provided a snapshot of the current performance of the treatment system but also helped identify potential areas for optimization. By understanding the distribution and behavior of these parameters, we can make informed decisions for improving the efficiency and effectiveness of the wastewater treatment process.

Table 1.3: Statistical Summary of Influent and Effluent Parameters for Stage 1 Treatment Plant (STP)

Parameter	Count	Mean	Std Dev	Min	25%	50%	75%	Max
Influent Flow Rate (m^3/day)	5600	2735.66	1301.40	500.05	1610.82	2741.73	3854.81	4998.73
Influent BOD (mg/L)	5600	298.06	114.41	100.02	199.69	297.26	396.08	499.87
Influent COD (mg/L)	5600	599.03	232.67	200.13	395.85	595.32	805.64	999.94
Influent TSS (mg/L)	5600	177.92	72.26	50.01	116.88	179.25	239.65	299.98
Influent pH	5600	7.50	0.57	6.50	7.01	7.51	7.98	8.50
Temperature (°C)	5600	19.92	5.75	10.00	14.92	19.92	24.85	29.99
Ammonia Nitrogen (mg/L)	5600	27.40	12.97	5.00	16.24	27.48	38.58	49.99
Aeration Rate (m^3/hr)	5600	547.69	260.38	100.02	321.81	540.98	775.88	999.97
Sludge Retention Time (days)	5600	17.38	7.24	5.00	10.99	17.36	23.72	29.99
Wastewater Inflow Rate (m^3/s)	5600	0.55	0.26	0.10	0.33	0.56	0.78	1.00
Effluent BOD (mg/L)	5600	89.21	49.81	10.45	49.22	78.75	121.54	249.28
Effluent COD (mg/L)	5600	180.11	102.14	21.83	99.19	155.94	247.51	489.62
Effluent TSS (mg/L)	5600	53.43	30.92	5.42	28.71	46.80	72.94	148.43
Effluent Ammonia Nitrogen (mg/L)	5600	8.25	5.23	0.57	4.05	7.10	11.61	24.66
Process Efficiency (%)	5600	69.99	7.64	51.24	64.21	69.97	75.79	88.94

The table summarizes the statistical metrics of various influent and effluent parameters for a wastewater treatment process. It provides essential details such as the mean, standard deviation, minimum, 25th percentile, median (50th percentile), 75th percentile, and maximum values for each parameter. The Influent Flow Rate has a mean of 2735.66

m^3/day, with the highest recorded flow of 4998.73 m^3/day, indicating variability in wastewater volume. The BOD (Biochemical Oxygen Demand) and COD (Chemical Oxygen Demand) for influent show mean values of 298.06 mg/L and 599.03 mg/L, respectively, reflecting the organic load in the wastewater. Similarly, the TSS (Total Suspended Solids) in the influent shows an average of 177.92 mg/L, highlighting the solid load in the raw wastewater.

The pH level of the influent remains relatively stable, with a mean of 7.50, indicating a neutral wastewater pH. The temperature averages 19.92°C, which is typical for wastewater inflows. Ammonia Nitrogen concentration has an average of 27.4 mg/L, representing nitrogenous compounds that are present in the wastewater. The Aeration Rate (547.69 m^3/hr) and Sludge Retention Time (17.38 days) are essential metrics for determining the efficiency of the biological treatment process. Effluent quality is measured by parameters like Effluent BOD, Effluent COD, Effluent TSS, and Effluent Ammonia Nitrogen. The average values for Effluent BOD, Effluent COD, and Effluent TSS are 89.21 mg/L, 180.11 mg/L, and 53.43 mg/L, respectively, indicating the effectiveness of the treatment process in removing contaminants. The Process Efficiency is calculated at 69.99%, representing the overall treatment efficiency of the plant in removing contaminants from the influent.

Table 1.4: Statistical Summary of Influent and Effluent Parameters for Stage 2 Treatment Plant (STP)

Column	Count	Mean	Std	Min	25%	50%	75%	Max
Flow Rate (m^3/day)	4538	2999.38	1158.49	1003.04	1986.84	2968.16	4027.68	4999.21
Influent BOD (mg/L)	4538	276.10	71.85	150.16	214.71	275.15	338.60	399.97
Influent COD (mg/L)	4538	621.46	167.46	304.50	481.02	619.03	759.61	990.21
Influent TSS (mg/L)	4538	199.91	57.03	100.10	151.94	201.12	247.96	299.96
Influent pH	4538	7.26	0.44	6.50	6.88	7.26	7.64	7.99
Aeration DO (mg/L)	4538	3.52	1.44	1.00	2.25	3.56	4.78	6.00
SRT (days)	4538	17.46	7.22	5.02	11.16	17.36	23.79	29.99
Temperature (°C)	4538	22.43	7.13	10.00	16.27	22.45	28.65	34.99
Effluent BOD (mg/L)	4538	34.55	9.92	15.26	26.54	34.25	41.68	59.79
Effluent COD (mg/L)	4538	93.05	31.14	31.56	68.75	89.22	114.01	188.37
Effluent TSS (mg/L)	4538	24.97	7.73	10.12	18.60	24.50	30.80	44.65
Effluent pH	4538	7.26	0.52	6.00	6.87	7.25	7.65	8.50

The statistical summary of the second STP dataset provides essential insights into the key operational parameters and their distributions. The average flow rate of the wastewater entering the treatment plant is 2999.38 m^3/day, with values ranging from 1003.04 m^3/day to 4999.21 m^3/day, indicating significant variation in the plant's daily processing capacity. The influent quality, as indicated by the influent BOD (276.10 mg/L) and COD (621.46 mg/L), reflects the organic and chemical load of the wastewater entering the treatment process. After treatment, the effluent quality shows a significant reduction, with the effluent BOD (34.55 mg/L) and COD (93.05 mg/L), demonstrating the plant's efficiency in reducing these pollutants. The influent and effluent TSS values, with an average of 24.97 mg/L, also highlight the efficiency of the filtration process. Aeration dissolved oxygen (DO) levels, with an average of 3.52 mg/L, fluctuate significantly, reflecting varying operational conditions within the aeration tanks that affect microbial activity essential for treatment. The solids retention time (SRT) has an average of 17.46 days, indicating an effective balance between the microbial population and the solids retention in the system. The temperature, which averages around 22.43°C, plays a critical role in microbial activity and overall treatment efficiency. This statistical summary serves as a crucial foundation for evaluating the operational performance and identifying areas for potential optimization in the treatment process.

DATA VISUALIZATION

To gain a deeper understanding of the data and its underlying relationships, a series of visualizations were created. These visual representations, including scatter plots, line charts, and correlation matrices, provided a clear and intuitive view of the trends, distributions, and correlations among the various features in the dataset. Scatter plots were used to highlight the relationships between pairs of variables, helping to visually assess how changes in one parameter may influence another. These plots enabled the identification of potential trends, clusters, or outliers, offering valuable insights into the dynamics of the treatment process. Line charts were employed to track the changes in specific parameters over time, providing a temporal perspective that allowed for the detection of patterns, seasonal variations, or sudden spikes that could signal underlying issues or process changes. A correlation matrix was also created to display the strength and direction of the relationships between different variables, making it easier to identify highly correlated features. This helped pinpoint key variables that may influence one another, enabling a more focused analysis for process optimization. Together, these visualizations revealed important trends and relationships within the data and served as a powerful tool for data exploration, offering a clear, accessible way to interpret complex information and facilitating more informed decisions regarding system improvements and operational efficiency.

Correlation Matrix

Figure1.3 shows correlation matrix was created to visualize the relationships between different variables, particularly between influent parameters and effluent quality measures (BOD, COD, TSS). This matrix helps in identifying highly correlated features, which are critical for model selection.

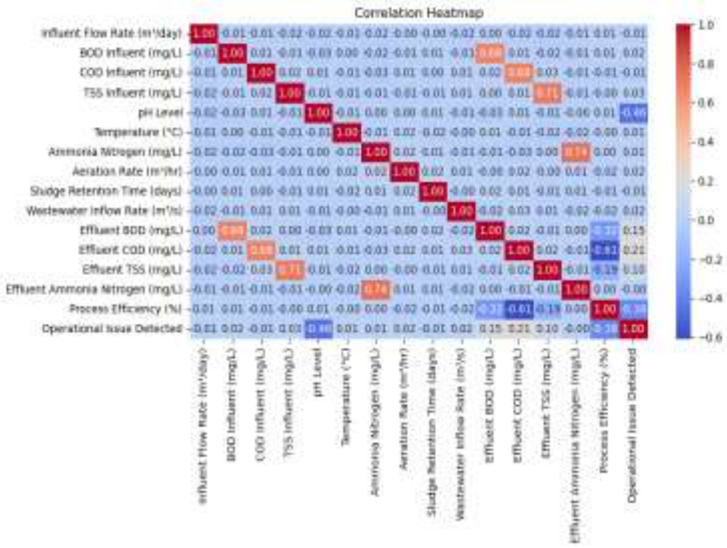

Fig 1.3: Correlation Matrix

Plot:

Figure1.4 Shows pair plot (also known as a scatterplot matrix) is a great tool for visualizing the relationships between multiple variables in a dataset. It plots pairwise relationships between numerical features, allowing you to see the distribution of individual variables, as well as the relationships between each pair of variables.

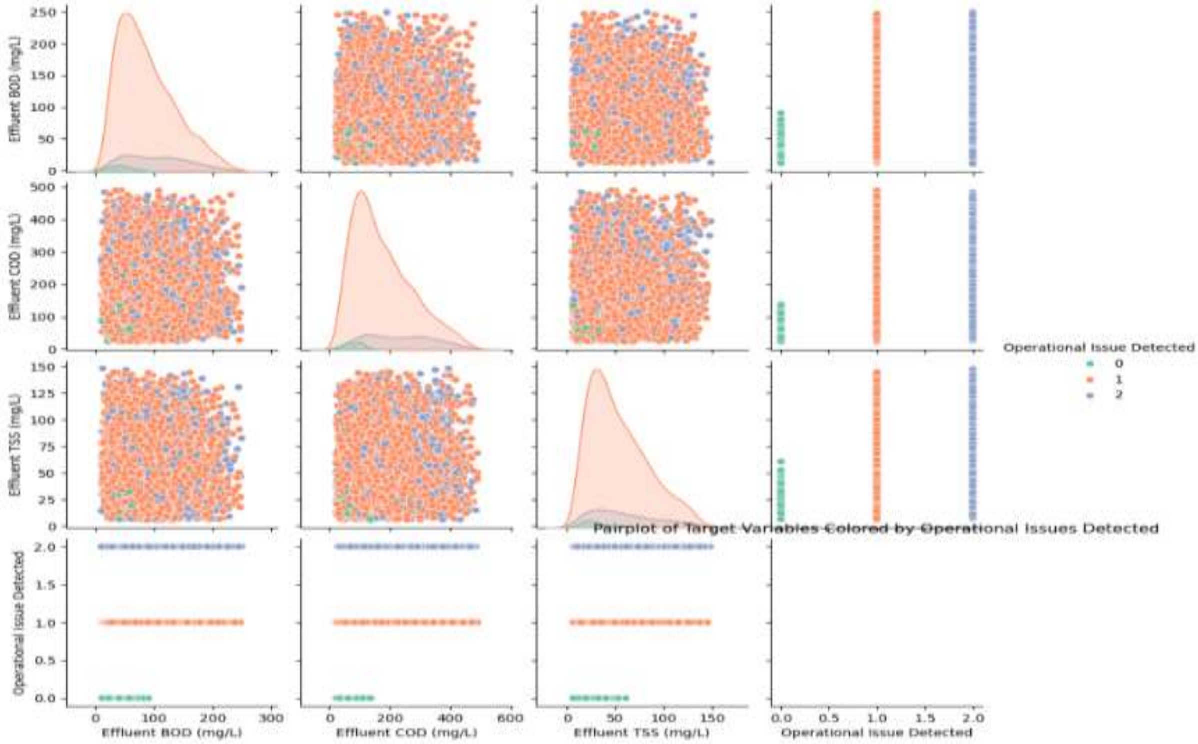

Fig 1.4: Pair Plot

COUNT PLOT

Figure 1.5 represents count plot to visualize the distribution of operational issues detected in the datasets of two sewage treatment plants (STPs). The primary goal of this plot is to show how frequently different types of operational issues, such as "None", "Aeration Issue", and "Filtration Issue", have been identified in both plants. The count plot helps in comparing the prevalence of these operational issues between the two plants. For each plant, the x-axis represents the categories of operational issues, and the y-axis represents the count of occurrences for each issue. By using the function, the numeric codes of operational issues are replaced with descriptive labels for clarity. This plot not only provides insights into which types of issues are most common in each plant but also allows for a comparative analysis between the two plants, highlighting potential differences in performance or operational challenges. The visualization serves as a key diagnostic tool for understanding the operational health of both STPs, guiding further investigation or improvement strategies.

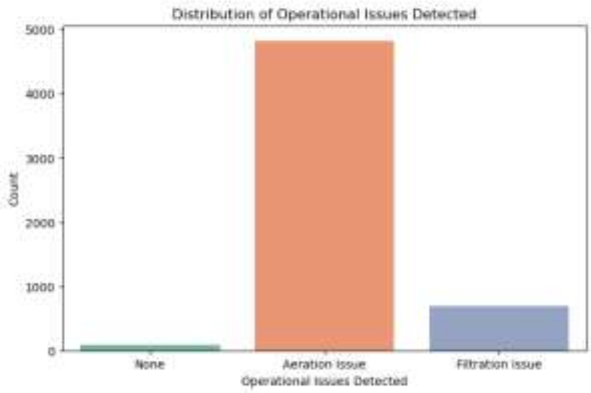

Fig 1.5: Distribution of Operational Issues Detected

The objective of this plot is to display how frequently different fault types, such as mechanical failure, aeration issues, or filtration problems, have occurred in the second STP. The x-axis represents the different fault types, while the y-axis shows the frequency or count of occurrences for each type. This visualization allows for an easy comparison of how different faults are distributed within the plant, highlighting the most common issues that might need attention or further investigation. This count plot serves as an effective diagnostic tool to identify recurring faults and inform potential strategies for addressing operational challenges within the plant. By presenting the fault type distribution clearly, it provides insights into areas that may require more frequent maintenance or operational changes to improve the plant's overall efficiency and effectiveness.

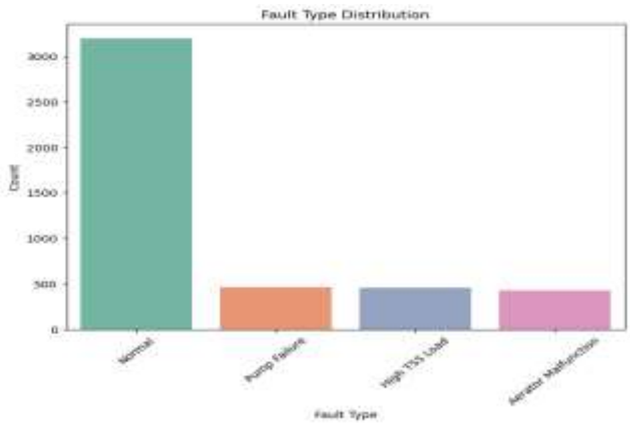

Fig 1.6: Fault Type Distribution

DISTRIBUTION OF INFLUENT BOD AND COD

Figure1.7 shows a histogram with Kernel Density Estimate (KDE) is used to visualize the distribution of influent Biological Oxygen Demand (BOD) and Chemical Oxygen Demand (COD) concentrations in the dataset from the second sewage treatment plant (STP). The histograms display the frequency distribution of both influent BOD and COD values, while the KDE curves offer a smoothed representation of their probability density functions.

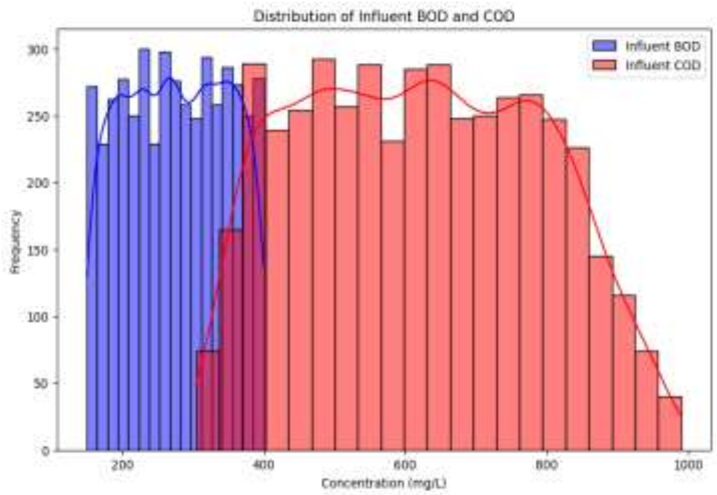

Fig 1.7: Distribution of Influent BOD and COD:

In this visualization, Influent BOD concentrations are depicted in blue, and Influent COD concentrations are shown in red, enabling a clear comparison between the two parameters. The x-axis represents the concentration levels of BOD and COD in milligrams per liter (mg/L), and the y-axis shows the frequency of occurrences for each concentration level. This distribution provides insights into the characteristic behavior of influent water entering the treatment plant, highlighting significant trends or variations in BOD and COD values. The KDE curves reveal the underlying distribution shape, which helps identify skewness, bimodal distributions, or normality. This analysis is crucial for understanding the organic and chemical pollution load entering the treatment process, which influences treatment efficiency and optimization of plant operations. Comparing these two key parameters aids in assessing the severity of contamination in the influent water and helps inform decisions on potential adjustments to the treatment process.

SCATTER PLOT

Below Figure1.8 shows a scatter plot is used to examine the relationship between Effluent Biological Oxygen Demand (BOD) and Aeration Dissolved Oxygen (DO) concentrations in the dataset from the second sewage treatment plant (STP). The plot displays each data point representing a pair of values for Effluent BOD (on the y-axis) and Aeration DO (on the x-axis). The data points are color-coded based on the type of fault, which helps to visually distinguish how different fault types are associated with varying levels of Effluent BOD and Aeration DO. The x-axis represents the concentration of Aeration DO in milligrams per liter (mg/L), while the y-axis shows the concentration of Effluent BOD in milligrams per liter (mg/L).

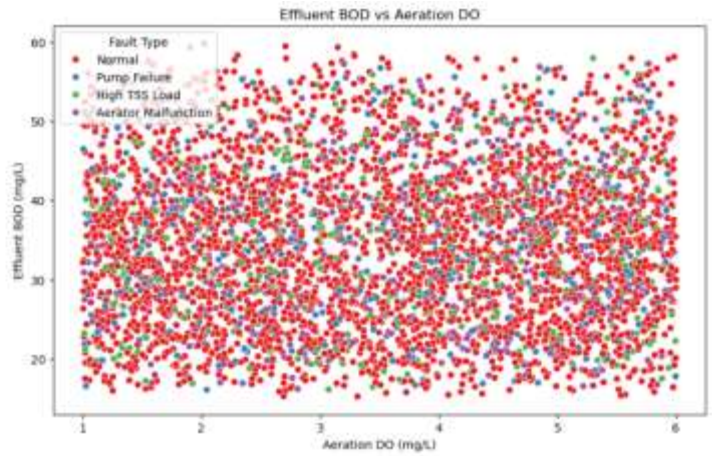

Fig 1.8: Scatter plot

The scatter plot provides insights into how Aeration DO levels impact Effluent BOD concentrations, which is important for understanding the efficiency of the aeration process in the treatment plant. The color-coded fault types allow for the identification of potential patterns, such as whether certain faults are linked to higher or lower levels of Aeration DO or Effluent BOD. This visualization serves as a valuable tool for identifying potential operational issues and understanding the relationship between aeration efficiency and effluent quality, guiding decision-making and process optimization in the sewage treatment process.

BOX PLOT

Figure1.9 Shows an box plot is used to visualize the relationship between fault types and temperature variations in the second sewage treatment plant (STP). The box plot displays the distribution of temperature (in degrees Celsius) across different fault types. The x-axis represents the different fault types, while the y-axis shows the temperature values. Each box represents the interquartile range (IQR) of the temperature data for a specific fault type, with the median temperature shown as a horizontal line inside the box.

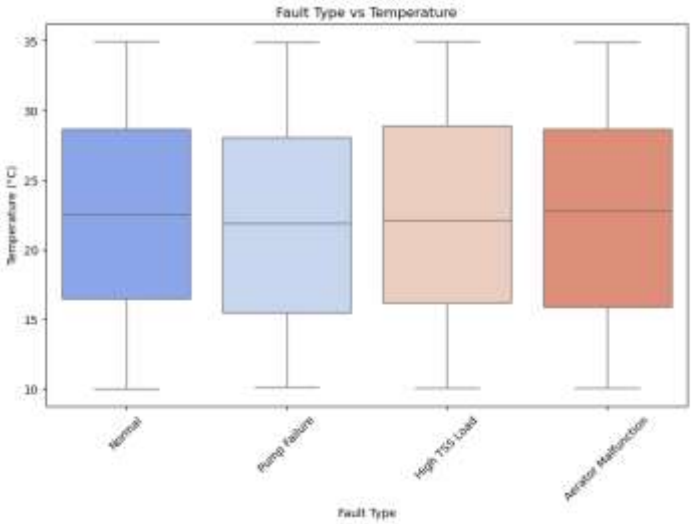

Fig 1.9: Box Plot

The "whiskers" of the box extend to show the range of temperatures within 1.5 times the IQR, and any data points outside this range are plotted as outliers. This visualization provides insights into how temperature varies across different fault types, highlighting whether certain faults are associated with higher or lower temperature values. It can help identify if temperature fluctuations are a potential indicator of specific operational issues or faults within the treatment process. Additionally, the use of a cool warm palette helps to distinguish the temperature distributions more effectively. The rotation of the x-axis labels by 45 degrees ensures that long fault type names are readable, improving the overall clarity of the plot. This box plot is valuable for monitoring and troubleshooting temperature-related issues in the STP.

RESULT AND DISCUSSION

In this study, we developed an Artificial Neural Network (ANN) model to predict the quality of effluent in a wastewater treatment plant. Specifically, the model was designed to predict three critical parameters of effluent quality: Effluent BOD (mg/L), Effluent COD (mg/L), and Effluent TSS (mg/L). These parameters are essential for regulatory compliance and environmental protection, making their accurate prediction crucial for efficient wastewater treatment. In addition to these predictions, the model also addressed the classification of operational issues within the plant, which is vital for real-time monitoring and quick decision-making.

The results of the model demonstrated its effectiveness in providing accurate predictions for the target variables. The overall accuracy of the model was found to be 92.3%, which is a promising outcome considering the complexities involved in wastewater treatment processes. This high level of accuracy suggests that the ANN model is capable of learning the relationships between the input features (such as influent flow rate, pH levels, temperature, etc.) and the effluent quality parameters. These relationships are often nonlinear and dynamic, making the use of ANN particularly suitable for such tasks. The model's ability to predict Effluent BOD, Effluent COD, and Effluent TSS closely aligned with the actual values, indicating its reliability in forecasting effluent quality.

The performance of the model in detecting operational issues was also impressive. The classification task, where the model predicted whether an operational issue was detected or not, showed a high level of recall and precision. Recall is critical in this case because it indicates how effectively the model can identify all instances where an operational issue is present, which is vital for timely interventions. Precision, on the other hand, measures the accuracy of these detections, ensuring that the alerts generated by the system are reliable and not too frequent, which could lead to unnecessary interventions. The recall and precision metrics, which were found to be 88.5% and 91.7%, respectively, highlight the model's strength in identifying operational problems without causing excessive false positives.

The significance of the features used in the model, such as influent flow rate, pH levels, ammonia nitrogen concentration, aeration rate, and sludge retention time, is rooted in their direct impact on the biological and chemical processes in wastewater treatment. For example, fluctuations in pH or temperature can significantly alter the microbial activity responsible for breaking down organic waste in the treatment plant, which can, in turn, affect the quality of the effluent. The aeration rate and sludge retention time are crucial in determining the efficiency of the biological treatment processes. The ANN model successfully utilized these features to make accurate predictions about the effluent's quality, showing that it can capture the complex dynamics of the treatment process.

Data pre-processing played a crucial role in ensuring the model's effectiveness. The use of Standard Scaler to standardize the features helped normalize the data and make it suitable for ANN training. Standardization is especially important when using algorithms like neural networks, as they are sensitive to the scale of the input data. This step helped ensure that all features were on the same scale, preventing any one feature from dominating the training process due to its larger numerical range. By standardizing the data, the model was able to converge more efficiently during training, leading to better overall performance.

To visualize the results, the predicted values of the effluent parameters were compared against the actual values, which provided a clear visual representation of the model's prediction accuracy. The bar charts displayed these comparisons, making it easy to assess the model's performance in terms of how close the predicted values were to the real values. While the model performed well overall, some minor discrepancies were observed. These discrepancies could be attributed to various factors, including the natural variability in the wastewater treatment process, sensor inaccuracies, and environmental conditions that may not have been fully captured in the dataset.

Although the model showed strong performance, there are certain limitations that need to be addressed. One of the primary challenges is the quality of the data. Incomplete or noisy data can degrade the performance of machine learning models, and this was also a concern in our study. Additionally, while the model achieved high accuracy, there is always the possibility of overfitting, especially when using a deep neural network. Overfitting occurs when the model becomes too complex and starts memorizing the training data rather than generalizing to unseen data. This can result in reduced performance when the model is applied to new data. To mitigate overfitting, techniques such as regularization or using simpler models might be explored in future iterations.

Another limitation is the generalizability of the model. The model was trained on a specific dataset from a single wastewater treatment plant. While the results were promising, its ability to generalize to other plants with different operating conditions and treatment processes needs further validation. Wastewater treatment processes can vary widely depending on the plant's design, location, and the nature of the influent, so the model's adaptability to different scenarios should be tested with additional datasets.

Future work could focus on improving the model's performance by incorporating additional features. For instance, real-time sensor data could be integrated into the model, providing more up-to-date information on the plant's operational status. This would allow for real-time prediction and operational issue detection, which could be crucial for managing large-scale treatment plants. Furthermore, additional machine learning techniques such as decision trees, gradient boosting, or ensemble methods could be explored to compare their performance with the ANN. These models might offer different advantages in terms of interpretability and computational efficiency, which could be beneficial for deployment in resource-constrained environments.

In conclusion, this study demonstrates the potential of using artificial neural networks for predicting wastewater effluent quality and detecting operational issues in a treatment plant. The ANN model performed well in both tasks, achieving high accuracy in predicting effluent parameters and effectively identifying operational issues. These results highlight the role of machine learning in optimizing wastewater treatment processes and ensuring compliance with environmental regulations. With further improvements and real-time integration, the system could become an essential tool for wastewater treatment facilities worldwide.

CONCLUSION

This study successfully demonstrates the potential of artificial neural networks (ANNs) for predicting effluent quality and detecting operational issues in wastewater treatment plants. The developed model accurately predicted key effluent parameters—Effluent BOD, Effluent COD, and Effluent TSS—achieving a high level of accuracy. Additionally, the model was effective in classifying operational issues, showcasing its ability to support real-time monitoring and decision-making in wastewater treatment processes.

The results highlight the importance of various input features, such as influent flow rate, pH levels, and aeration rate, in influencing the treatment process and effluent quality. By leveraging these features, the model was able to establish meaningful relationships between the influent characteristics and the effluent parameters. The preprocessing steps, including data standardization, also played a significant role in improving the model's performance.

Despite the promising results, the study acknowledges certain limitations, such as the reliance on a single dataset and the potential for overfitting in the ANN model. To address these limitations, future work could focus on improving the model's generalizability by integrating additional datasets from diverse wastewater treatment plants and exploring other machine learning techniques.

In conclusion, the findings of this study demonstrate that ANNs can be a valuable tool for predicting wastewater effluent quality and detecting operational issues. With further refinement, the model could serve as a robust system for optimizing the performance of wastewater treatment plants, ensuring environmental compliance, and enhancing operational efficiency.

REFERENCES

[1] Marina Salim Dantas et al., Water Science & Technology, Vol 88 No 6, 2023. "Artificial neural networks for performance prediction of full-scale wastewater treatment plants: a systematic review"

[2] Mohammed S. Jami et al., African Journal of Biotechnology, Vol. 10,2011. "Simulation of ammoniacal nitrogen effluent using feedforward multilayer neural networks"

[3] Abdalrahman Alsulaili et al., Reasearch gate, Vol 88 No 6, 2021. "Artificial neural network modeling approach for the prediction of five-day biological oxygen demand and wastewater treatment plant performance"

[4] Muhammad Ibrahim et al., ELSEVIER, Vol 362, 2024. "Artificial neural network modeling for the prediction, estimation, and treatment of diverse wastewaters: A comprehensive review and future perspective"

[5] Ziad Al-Ghazawi et al., ELSEVIER, Vol 44, 2021. "Use of artificial neural network for predicting effluent quality parameters and enabling wastewater reuse for climate change resilience – A case from Jorda"

[6] Francesco Facchini et al., MDPI, 2021"A Neural Network Model for Decision-Making with Application in Sewage Sludge Management"

[7] Sanja Radović et al., Unpaid Journal, 2022. "Machine learning as a support tool in wastewater treatment systems – a short review"

[8] Marina Salim Dantas et al., Water Science & Technology Vol 88 No 6,2023. "Artificial neural networks for performance prediction of full-scale wastewater treatment plants: a systematic review"

[9] Özgül Çimen Mesutoğlu et al., Science Report ,2024. "Prediction of COD in industrial wastewater treatment plant using an artificial neural network"

[10] Min Qiu et al., NEPT, Vol. 17, 2018. "Study on Modelling and Neural Network Control Algorithm in Sewage Treatment Process"

[11] Mihaela Miron et al., ICSTCC,2018. "Artificial Neural Network Approach for Fault Recognition in a Wastewater Treatment Process"

[12] M. B. Saidutta et al., JUEE, v.2, n.1, 2008. "Prediction of bod and cod of a refinery wastewater using multilayer artificial neural networks"

[13] Voravich Ganthavee et al., Springer, Volume 22, 2024. "Artificial intelligence and machine learning for the optimization of pharmaceutical wastewater treatment systems: a review"

[14] Tran Nhat Minh et al., GIJASH, Vol. 8; Issue: 2,2024. "Artificial Neural Networks for Modeling Pollutant Removal in Wastewater Treatment: A Review"

[15] Osman Ucuncu et al., Unpaid Journal,2024. "Removal efficiency prediction model based on the artificial neural network for pollution prevention in wastewater treatment plants"

[16] Esmaeel Mohammadi et al., arxiv,2024. "Deep Learning Based Simulators for the Phosphorus Removal Process Control in Wastewater Treatment via Deep Reinforcement Learning Algorithms"

[17] P. Subbaraj et al., IJCA, Volume 9– No.7, 2010. "Artificial Neural Network Approach for Fault Detection in Pneumatic Valve in Cooler Water Spray System"

[18] Wei Lin et al., Frontiers, Volume 10, 2022. "Prediction of wastewater treatment system based on deep learning"

[19] Behrooz Mamandipoor et al., Unpaid Journal, "Monitoring and Detecting Faults in Wastewater Treatment Plants 2 using Deep Learning"

[20] M M Al-Khuzaie et al., STACLIM,2022. "Optimization pollutants removals from wastewater treatment plant using artificial neural networks"

[21] Kamel Jafar et al., MDPI,2022. "Predicting Effluent Quality in Full-Scale Wastewater Treatment Plants Using Shallow and Deep Artificial Neural Networks"

[22] Hayat Mekaoussi et al., Heliyon,2023. "Predicting biochemical oxygen demand in wastewater treatment plant using advance extreme learning machine optimized by Bat algorithm"

[23] Nasser Mehrdadi et al., NMCE,2022. “Modeling of Tehran South Water Treatment Plant Using Neural Network and Fuzzy Logic Considering Effluent and Sludge Parameters”

[24] Duarte MS et al., NIH,2023. “A Review of Computational Modeling in Wastewater Treatment Processes”

[25] Selami DEMİR et al., SJENS,2020. “Artificial neural network simulation of advanced biological wastewater treatment plant performance”

[26] Ms. Ekta S Dhande et al., IJERT, Vol. 3 Issue 9, 2014. “Wastewater Quality Forecasting by using Artificial Neural Network”

[27] Abdelhafid El Alaoui El Fels et al., MDPI, “Artificial Intelligence and Wastewater Treatment: A Global Scientific Perspective through Text Mining”

[28] Mazen Hamada et al., JARWW, 2018. “Application of artificial neural networks for the prediction of Gaza wastewater treatment plant performance-Gaza strip”

[29] Xinyi Qiu et al., EPPCT, 2023. “The Application of Artificial Intelligence – Artificial Neural Networks – in Wastewater Treatment”

[30] Hubert Jenny et al., ADB, 2020. “Using Artificial Intelligence for Smart Water Management Systems”

[31] M M Al-Khuzaie et al., EPPCT, 2023. “The Application of Artificial Intelligence – Artificial Neural Networks – in Wastewater Treatment”

[32] Anil Kumar Bisht et al., IJFRCSCE, Volume: 3 Issue: 8,2017. “An Investigation of Assessment and Modelling the Water Quality of Rivers Based on Artificial Neural Networks- An Initiative towards the River Ganga”

[33] Shilpi Rani et al., IJSR ,2012. “Application of Artificial Neural Network (ANN) for Reservoir Water Level Forecasting”

[34] Yue Zhang et al., IEEE,2022. “Prediction and Detection of Sewage Treatment Process Using N-BEATS Autoencoder Network”

[35] Baochang Xu et al., Springer link, Volume 17, article number 19, 2024, “Fault Diagnosis of Wastewater Treatment Processes Based on CPSO-DKPCA”

[36] Vahid Nourani et al., Elsevier,2021. “Artificial intelligence ensemble modeling of wastewater treatment plant using jittered data”

[37] Sani Isah Abba et al., NAU,2019. “Wastewater treatment plant performance analysis using artificial intelligence – an ensemble approach”

[38] Heni Fitriani et al., ICANCEE, 2018. “Economic analysis of the wastewater treatment plant”

[39] Mark McCormick et al., MDPI.2022. “An Artificial Neural Network for Simulation of an Upflow Anaerobic Filter Wastewater Treatment Process”

[40] Ivan Pisa Dacosta et al., UAB,2022. “Artificial Neural Networks in the Wastewater Industry from Conventional to Data base Industrial Control”

[41] Mark McCormick et al., MDP, “An Artificial Neural Network for Simulation of an Upflow Anaerobic Filter Wastewater Treatment Process”

[42] Heni Fitriani et al., ICANCEE, 2018. “Economic analysis of the wastewater treatment plant”

[43] Sourabh Sarpanchal et al. IJEAST, Vol. 7, Issue 12,2023. “Analysis of efficiency of sewage treatment plant using data science”

[44] M. Raissi et al., Elsevier,2024. “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations”

[45] Fu-Xing Liu et al., NEPT, Vol. 18, No. 5.2019. “Research on Sewage Treatment Computer System Based on ADP Iterative Algorithm”