Enhancing Water Potability Prediction Using Sparse Attention-Based Deep Learning Models

MSc Research Project

MSc Data Analytics - C

Saikiran Reddy Pakeer

Student ID: x23332310

School of Computing

National College of Ireland

Supervisor: Dr. Christian Horn

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	……. Pakeer Saikiran Reddy……………………………………………………
Student ID:	……x23332310
Programme:	…… MSc in Data Analytics ……… **Year:** 2024-25…
Module:	…… MSc Research Practicum
Supervisor:	…… Dr. Christian Horn
Submission Due Date:	…15/09/2025
Project Title:	…… Enhancing Water Potability Prediction Using Sparse Attention-Based Deep Learning Models
Word Count:	………7120…… **Page Count**……………………………19………………….…….

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Pakeer Saikiran Reddy

Date: 11/08/2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	✓
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	✓
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	✓

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Enhancing Water Potability Prediction Using Sparse Attention-Based Deep Learning Models

Saikiran Reddy Pakeer

Student ID: x23332310

Abstract

Reliable prediction of water potability is essential for safeguarding public health and supporting sustainable resource management. However, real-world water quality data is often subject to noise, sensor inaccuracies, and missing values, which can degrade model performance. This study proposes the use of a sparse attention-based deep learning model to enhance the robustness and generalizability of water quality prediction systems. By introducing controlled Gaussian noise at varying intensities to the training data, the model's capacity to maintain predictive accuracy under noisy conditions is systematically evaluated. The proposed approach leverages the TabNet architecture, which employs feature-wise attention to focus selectively on the most informative inputs. This mechanism enables effective learning even when data quality is degraded. Experimental results demonstrate that the sparse attention framework remains stable across multiple perturbation levels, underscoring its potential for deployment in real-time and uncertain environments. The findings suggest that incorporating sparse attention into predictive systems can significantly strengthen their resilience, making them more suitable for real-world water quality monitoring applications.

1 Introduction

1.1 Problem Background

The health of people, sustainability of the ecosystem and economic activities such as agriculture and industries are largely affected by the quality of water. The availability of safe and drinkable water is a highly significant issue on a global scale as chemical pollutants, heavy metals, and microbial agents can be rather detrimental to the health of the population (Patel et al., 2022; Uddin et al., 2023). The most popular methods of water quality measurement are characterized by the regular physical sampling and laboratory analysis, which is costly, time-consuming, and provides poor time resolution (William et al., 2023). This type of limitations does not enable making decisions on time, especially in dynamic conditions when the quality of water can change rapidly due to natural or anthropogenic factors.
In response, the application of data-driven techniques using machine learning (ML) and deep learning (DL) technologies to forecast water potability and quality indices on the basis of sensor and environmental data has recently gained popularity (Liu et al., 2023; Saroja & Dharshini, 2023). ML models that have proven successful in classification of water quality include decision trees and ensemble methods (Patel et al., 2022; Li et al., 2022), and DL models that have been successful in modeling the nonlinear relationship and time dependency that

defines water quality datasets include CNNs and LSTMs (Yang et al., 2023; Zhang et al., 2024).
Despite these advances, however, some problems remain related to the high dimensionality, multivariate, and sparse nature of water quality data recorded by different sensor networks (Baena-Navarro et al., 2025; Aldrees et al., 2023). The majority of the existing models tend to assign equal importance to all the features, which can lead to overfitting, computational complexity, and failure to comprehend which parameters influence predictions most of all (Ahmed et al., 2025; Saroja & Dharshini, 2023). These shortcomings are particularly unfavorable in real-time monitoring scenarios, and resource-limited environments where rapid and reliable and explainable forecasts are required.

1.2 Motivation

The rising number of water quality data collected by sensors and monitoring stations presents a unique opportunity to use advanced computational models to predict the water potability with high accuracy and on time (Baena-Navarro et al., 2025; William et al., 2023). The complexity and heterogeneity of such data, which consists of multivariate time-series, lack of values, noise, and sparse informative signals, however, often reduce the effectiveness of predictive models (Liu et al., 2023; Aldrees et al., 2023).
Although classic machine learning models, such as ensemble methods, including Random Forest and Gradient Boosting, have demonstrated good predictive performance (Patel et al., 2022; Uddin et al., 2023), they have limitations in their general ability to capture full temporal dependencies and complex interactions through high-dimensional data (Yan et al., 2023). Deep learning methods (particularly the ones that have incorporated the CNN and LSTM models) have improved the performance by modeling spatial and temporal patterns (Zhang et al., 2024; Yang et al., 2023). However, such models are often dense, computationally costly, and not interpretable, limiting their implementation in a resource-constrained setting and real-time surveillance settings (Ahmed et al., 2025; Saroja & Dharshini, 2023).
The introduction of sparse attention mechanisms to transformer-based models is a potential solution to this issue as this approach will enable models to focus on the most significant characteristics and time points (Rejini et al., 2025; Ghosh, 2023). This advantageous attention can decrease computational overload, decrease overfitting, and enhance model comprehension, which is essential in implementing water quality monitoring systems and especially in IoT-facilitated and embedded settings (Baena-Navarro et al., 2025; Ahmed et al., 2025).
Therefore, the water potability prediction possibility using sparse attention models is a worthy task with regard to addressing current weaknesses and improving the accuracy, efficiency, and transparency of environmental monitoring applications.

1.3 Research Question and Justification

Since the existing models of traditional machine learning and dense deep learning have certain difficulties in processing complex, high-dimensional, and frequently sparse water quality data, the following research question becomes critical:
How effective are sparse attention-based models in predicting water potability compared to traditional machine learning and deep learning approaches?

This question aims to address the potential of the selective focus capacity of sparse attention mechanisms to deliver practical gains in predictive performance, model interpretability, and efficiency.
Past studies have demonstrated that deep learning architectures, including CNN-LSTM and transformers, along with ensemble machine learning models, improve the accuracy of prediction, although complex computation and low interpretability are problems (Liu et al., 2023; Zhang et al., 2024; Ahmed et al., 2025). Other areas have already demonstrated the effectiveness of sparse attention models, like TabNet, in effectively learning tabular data, through dynamic feature selection during training (Aldrees et al., 2023; Rejini et al., 2025). Nevertheless, there is an under-explored use of them in the prediction of water potability.
Closing this gap is essential to the deployment of robust and explainable AI models to water quality monitoring systems in the real world, especially when resource constraints and explainability are the main factors to consider (Baena-Navarro et al., 2025; Ghosh, 2023). Therefore, the proposed research will be conducted to assess the sparse attention models in this regard, which may be of use in the future studies and practical applications.

1.4 Objectives of the Study

The main aim of this research is to explore the applicability of sparse attention-based models in predicting water potability with the aim of enhancing accuracy, interpretability, and efficiency of the model over traditional machine learning and deep learning approaches.

Specific objectives include:

- To design and build a sparse attention-based model, like TabNet, specifically on water quality data to perform dynamic feature selection and better prediction results.
- To test the resilience of the sparse attention model to noisy environment by applying Gaussian noise of different strengths to emulate the error in real-world sensors and less-than-perfect data.
- To Compare the performance of proposed sparse attention model with the baseline traditional machine learning and deep learning models.

With the realisation of the above, this study will evaluate a new methodological viewpoint to the prediction research on water quality due to the deficiencies that exist with regard to modelling transparency and efficiency.

2 Related Work

2.1 Machine Learning-Based Approaches for Water Quality Prediction

The development of predictive models of water quality classification and forecasting using machine learning (ML) has been essential, especially because it allows capturing nonlinear relationships between water parameters and the outcome of potability. Conventional ML methods like decision trees, support vector machines, random forests, and ensemble methods have been widely used in different environmental contexts.
Patel et al. (2022) have suggested a water potability prediction model using synthetic minority oversampling techniques (SMOTE) to solve the problem of class imbalance supplemented with explainable AI to increase interpretability. In a similar manner, Uddin et al. (2023) compared the results of various water quality index (WQI) models using a variety of ML algorithms, showing that ML is practical when classifying the state of water on a large scale. Yet another

significant contribution is the study by Yan et al. (2023), which combined machine learning and Bayesian optimization to improve long-term prediction of water quality, and demonstrated the value of uniting optimization methods with learning systems.

Li et al. (2022) suggested an interpretable ensemble model based on trees to predict the water quality of beaches in the context of coastal and recreational water bodies, which provides both accuracy and transparency. Aldrees et al. (2023) also support the importance of ensemble learning, as they used evolutionary and ensemble ML to successfully assess water quality in different environmental conditions. William et al. (2023) have also discussed AI-based models in the monitoring and prediction of water quality based on real-time sensor data to assist decision-making systems in aquatic management.

All these studies suggest that classical and ensemble-based ML models have the potential to perform significantly well in predicting water quality. However, they are generally poor at handling non-trivial time-dependencies, high-dimensional multivariate input, and hierarchies of feature interactions, which are increasingly relevant in dynamical and sensor-based tasks.

2.2 Deep Learning-Based Approaches for Water Quality Prediction

Since they have shown to be capable of capturing non-linear and complex relationships and interactions between features, especially when working with high-dimensional and unstructured data such as time series and sensor streams, deep learning (DL) models have gained massive popularity in predictive analyses of water quality.

Saroja and Dharshini (2023) presented a deep learning model of a classification system to predict potable water based on a neural network, stating that they are more precise and generalizable than classical solutions. Liu et al. (2023) proposed an original two-stage feature selection and intelligent optimization deep ensemble model with high robustness and adaptability to noise data. In the same regard, Talukdar et al. (2023) demonstrated the application of deep neural networks and sensitivity-uncertainty analysis to predict lake water quality indices to better manage uncertainty in environmental systems.

Some of the studies were devoted to domain-specific architectures. As an example, Yang et al. (2023) used deep learning to forecast the effluent quality of constructed wetlands, a distinctive environment where heterogeneous inputs have to be modeled. Wang et al. (2024) went further to apply DL through combining pollution source detection and water quality forecasting with end-to-end deep models. Baena-Navarro et al. (2025) applied deep learning and IoT-based monitoring systems in the development of intelligent and continuous water quality management pipelines in the context of aquaculture.

Hybrid deep learning models have also been studied recently. Zhang et al. (2024) relied on attention-based CNN-LSTM models to obtain the spatial and temporal dynamics of surface water quality. Ahmed et al. (2025) proposed HydroNet, a two-stage CNN-Transformer based multi-modal feature fusion and attention model that achieves high precision classification. In the meantime, Rejini et al. (2025) used transformer-based models to evaluate the suitability of agricultural waters, and this aspect indicates the increased popularity of attention-based approaches in water systems.

These experiments support the effectiveness of deep learning paradigm in the management of complex datasets and time-based variables. Nevertheless, even after enhancing the performance, most deep learning models are dense in nature, which results in higher computational expense and hard to interpret- particularly when they are applied in resource-limited or real-time settings.

2.3. Hybrid and Emerging Techniques in Water Quality Prediction

Due to the mounting complexity of environmental datasets, recent studies have shifted to the hybrid and emerging modeling methods that merge the power of conventional and deep learning models. Such strategies are meant to overcome the drawbacks of overfitting, redundancy of features, and uninterpretability in single-model systems.

The combination of more than one learning strategy is one of the important trends in recent literature. A Bayesian model averaging (BMA)-based data fusion method to fuse outputs of different deep learning models robustly and enhance water quality parameter prediction accuracy was suggested by Alizamir et al. (2025). The uncertainty modeling is possible in this probabilistic fusion method, which is crucial in essential areas like the use of water in agriculture and in public health.

HydroNet designed by Ahmed et al. (2025) is an illustration of the increased attention to hybrid deep learning architectures. HydroNet combined the transformer-based attention mechanism with convolutional neural networks (CNNs) and attained better classification results by fusing multi-scale features. Likewise, Zhang et al. (2024) augmented CNN-LSTM models with attention mechanisms, so that the models could predict multi-indicator time-series by learning where and when to attend in sequential environmental data.

Transformer-based architectures are also being equipped with attention mechanisms. Such models were used by Rejini et al. (2025), who applied them to the problem of predicting the suitability of agricultural water, demonstrating how transformer models initially developed to solve language processing tasks are currently being adapted to environmental analytics. Ghosh (2023) investigated transformer architectures as applied to long-term rainfall predictions- yet another essential component of water resource planning and distribution within smart cities.

On a similar front, the combination of Internet of Things (IoT) and AI models has proved to be a major facilitator of continuous and real-time monitoring. Baena-Navarro et al. (2025) proved a practical use of ML-based IoT in the water management of aquaculture, which made the process sustainable due to automation and real-time feedback. High-frequency data that these systems produce demand efficient, but also scalable models, which is why the transition to attention-based and lightweight architectures is encouraged.

Although these improvements have been made, most existing models continue to use dense architectures that have high computational costs, low interpretability, and poor flexibility to sparse or imbalanced features, particularly in multivariate tabular or time-series data.

2.4 Research Gaps and the Need for Sparse Attention-Based Models

Even though the machine learning and deep learning methods have demonstrated high potential in the application of water quality prediction tasks, there are still a number of critical limitations that have not been addressed. Such gaps necessitate the desire to have more scalable, interpretable, and efficient modeling solutions, especially in high-dimensional, multivariate, and real-time water quality data.

One of the limitations that has been seen repeatedly in the literature analyzed is the use of dense architectures, especially in deep learning models, including CNNs, LSTMs, and regular transformers (Saroja & Dharshini, 2023; Zhang et al., 2024; Ahmed et al., 2025). Although such models are capable of representing complex structures, they have the tendency to treat all features in the inputs equally, whether they are relevant to the prediction task or not. This tends to overfit, make the model computationally expensive and less interpretable, which is undesirable when applied to resource-limited settings such as rural water monitoring stations or Internet of Things (IoT)-based aquaculture systems (Baena-Navarro et al., 2025).

In addition, current ensemble and hybrid models (Liu et al., 2023; Alizamir et al., 2025; Aldrees et al., 2023) perform better due to either feature combination or feature fusion approaches, but they do not usually have the ability to selectively pay attention to informative subsets of features. Most ML models are based on the assumption of equal feature importance, which may be a hindrance to generalization with sparse, noisy, or imbalanced features, a typical case in environmental monitoring.

Such restrictions highlight one important research gap, namely that there is a need to develop models that are not just accurate and robust, but also selectively sensitive to the most important inputs. Sparse attention-based models can be viewed as an attractive solution to such a context. Sparse attention mechanisms — as seen in recent transformer variants — allow models to prioritize key features and time steps, effectively reducing noise and computational burden. Unlike full attention models that scale quadratically with input size, sparse attention methods enable more efficient training and inference, making them suitable for high-frequency, multi-sensor water quality data. Furthermore, the interpretability of attention weights adds transparency to feature influence, which is vital for regulatory decision-making and public trust in AI systems.

Therefore, this study proposes leveraging sparse attention-enhanced transformer architectures tailored for water potability prediction. These models can address the identified shortcomings by:

- Improving feature selection dynamically during training,
- Handling both structured (tabular) and sequential (time-series) data effectively,
- Increasing real-time or embedded scalability,
- And providing greater interpretability than the black-box neural models.

With their focus on addressing these loopholes, sparse attention-based models can set a new standard in intelligent, efficient, and explainable forecasting of water quality.

3 Research Methodology

The proposed study will apply the sparse attention-based model, TabNet, to solve the key issues found in water quality forecasting, especially those concerns high-dimensional, sparse, and imbalanced data as they are present in the case of environmental monitoring. The choice of TabNet as the core modeling architecture is motivated by its unique capacity to perform sequential attention-based feature selection, allowing the model to focus dynamically on the most relevant predictors while maintaining interpretability and computational efficiency. This approach directly tackles the drawbacks observed in existing dense models, such as overfitting, increased computational cost, and limited transparency.

3.1 Data Description

The data involved in this study has 100,000 samples and 23 features describing physicochemical water parameters and a binary output variable that depicts the potability. The characteristics are the measurements of pH, iron, nitrate, lead, and other water quality indicators that are usually applied in environmental monitoring. The data were collected in public repositories of water quality and sensor data, with a representative sample to use in training and validation of the model.

3.2 Data Preprocessing

Considering that the data in the environmental field is heterogeneous and complicated, it was necessary to preprocess the data to make it ready to be modeled effectively. The following were done:

Handling Missing Values:

The dataset of water quality frequently have gaps in measurements or incomplete measurements because of sensor malfunctions or inconsistency in data collection. To overcome this, the missing values in all the numeric features were filled with median values since it is a robust measure that is immune to outliers and skewed distributions. The method is used to prevent biasing the dataset with imputation and preserving the inherent variability of the features.

Encoding Categorical Variables:

Although most of the variables in the dataset are in numeric form, all categorical variables were encoded by the label encoding method to transform them into numeric values that could be used by the modeling algorithms.

Normalization:
Min-Max normalization was used in order to scale all numeric features to a similar scale. This transformation scales each feature to a [0, 1] range, so that features with larger numeric ranges cannot dominate the learning process and allow the TabNet model to handle inputs.

Outlier Detection and Treatment:

The presence of outliers may mislead the training of a model and decrease the predictive performance. Numeric features were used to identify outliers using the interquartile range (IQR) method. The values that were beyond 1.5 times IQR below the first quartile or above the third quartile were truncated to the nearest boundary limits, and the effect of extreme values was curtailed, but the integrity of the data was maintained.

Feature Transformation:

Some of the features having median values near to zero (e.g., Iron, Lead, Turbidity) were log1p transformed to minimize the skewness and to approximate normality. This change enhances the stability and convergence of the model.

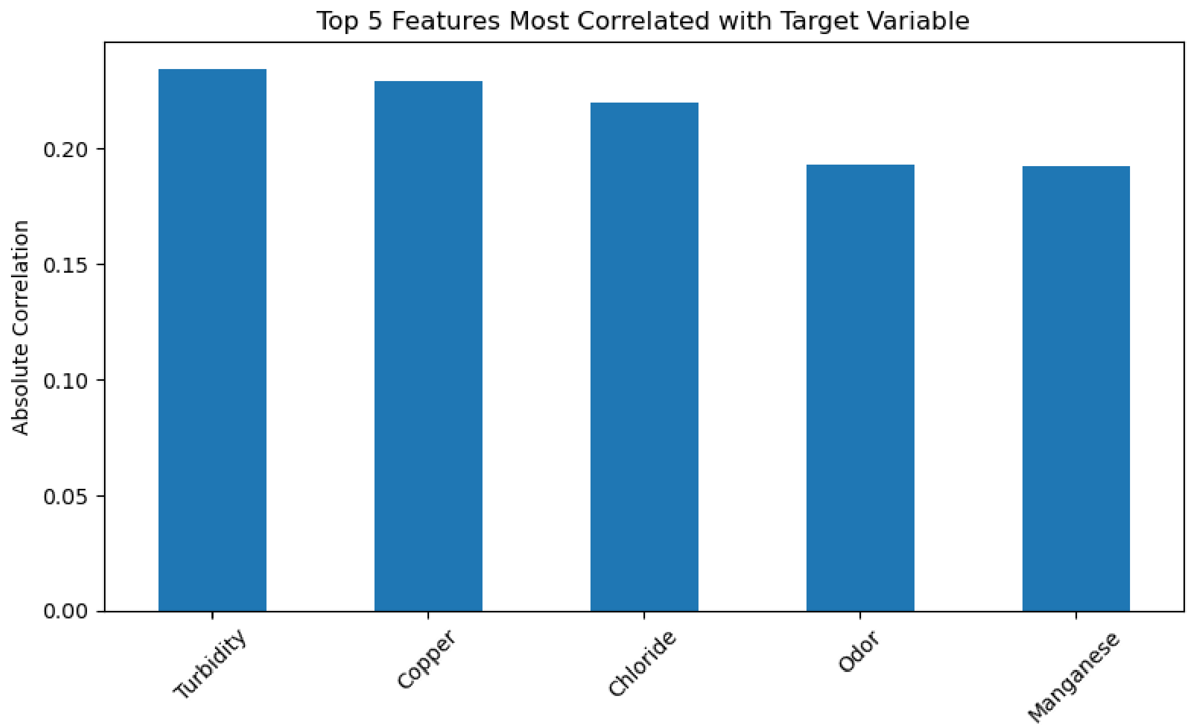

Figure 1. Most Correlated Features with target variable

Figure 1. presents the top five features most correlated with the target variable (water potability) in terms of absolute correlation values. Turbidity exhibited the highest absolute correlation (~0.23), followed closely by Copper (~0.225) and Chloride (~0.22), with Odor and Manganese contributing moderately (~0.19 each). While these values indicate meaningful relationships, all are below 0.3, suggesting that no single parameter is a dominant predictor and

that multivariate modeling is essential for accurate classification. The use of absolute correlations highlights the strength of associations but does not reveal whether these relationships are positive or negative. These results justify the application of feature-selection-aware models such as TabNet, which can capture complex, non-linear interactions and selectively attend to the most informative features, thereby improving predictive robustness.

Data Augmentation via Perturbation:

In order to make the models more robust and avoid overfitting, small Gaussian noise was added to the feature of training data. This controlled perturbation simulates measurement variability and helps the model generalize better to unseen data.

3.3 Dataset Splitting

The preprocessed dataset was split into training and testing subsets using stratified sampling to preserve the target class distribution. An 80:20 split was applied, with 80% of the data used for model training and 20% reserved for evaluating model generalization.

3.4 Model Development: TabNet

This study employs **TabNet** as the core modeling approach for water potability prediction. TabNet is a **deep learning architecture specifically designed for tabular data**, and it combines the power of deep neural networks with the interpretability of decision trees through a **sparse attention mechanism**. Unlike conventional models that treat all features equally, TabNet dynamically learns to focus on **relevant subsets of features at each decision step**, improving both accuracy and explainability.

3.4.1 Architecture Overview

TabNet is built around a sequential decision process inspired by gradient boosting but implemented through neural components. Its architecture consists of three primary components:

- **Feature Transformer Block**: A shared feed-forward network that processes the input features. This block captures high-level representations and transformations common across decision steps.
- **Attentive Transformer Block**: At each decision step, this module uses feature-wise attention to **select important features** using a sparse selection mask. This is governed by the **sparsemax** activation function, which encourages **hard selection** (i.e., focusing only on a few features per step), rather than soft averaging as in softmax.
- **Decision Step Outputs**: After the feature selection at each step, the transformed features contribute to the final prediction through additive aggregation across steps.

The sequential attention-based mechanism also allows local interpretability because each step of decision-making chooses a distinct set of features pertinent to the particular context of input.

3.4.2 Advantages of TabNet

- **Sparse Feature Selection**: TabNet selects only the most important features by paying attention to them selectively due to sparse attention. It is especially applicable to real-world data such as measurements of water quality, where certain sensors or features could be noisy or of less interest.
- **Interpretability**: TabNet is more interpretable than the usual deep models since it is possible to see which features were used to make each step of the decision, making it more transparent and usable in regulated settings such as environmental monitoring.
- **End-to-End Learning**: TabNet does not need lots of feature engineering. It is trained directly on raw tabular data and dynamically chooses features as it trains.
- **Efficient Inference**: TabNet is fast at inference even though it is complex and can be effectively deployed in real-time or embedded systems because of its lightweight architecture.

3.4.3 Model Implementation and Tuning

In the given work, TabNet was applied with the help of the pytorch-tabnet library based on PyTorch. Key hyperparameters were tuned with 5-fold cross-validation using grid search:

- **n_d and n_a (Dimension sizes)**: Set to 32 and 32 respectively for the decision and attention blocks.
- **n_steps**: Number of decision steps, optimized to 5 for balanced depth and efficiency.
- **gamma**: Relaxation factor for feature reuse, set to 1.5 to encourage sparse reuse.
- **lambda_sparse**: Sparsity regularization strength, tuned to prevent overfitting and enforce concise feature selection.

Training of the model was done with Adam optimizer learning rate scheduler and early stopping using validation loss, which would result in generalization of the model to unseen data. Training was stabilized by using batch normalization and virtual batch normalization.

3.5 Baseline Models and Justification

In order to assess the performance of the proposed sparse attention-based model (TabNet), we compare it with a number of baseline models that are frequently utilized in the context of the water quality prediction tasks. These consist of conventional machine learning algorithms, namely Decision Tree, Random Forest, Gradient Boosting, AdaBoost, and XGBoost, and a custom neural network model. The choice of these models is driven by the fact that they are used in many environmental prediction tasks, the different degrees of complexity and that they can process different kinds of datasets.

The Decision Tree model is an easy to understand and interpret algorithm that forms a baseline of rule-based classification of water potability. Random Forest model is chosen because it is an ensemble method and combines multiple decision trees, and in that way, it can provide better performance than a single decision tree by decreasing the overfitting. Gradient Boosting is a boosting model, which is incorporated because it can maximize the prediction accuracy by iteratively correcting the poorly classified examples, and as such it is appropriate to use it in

complex and non-linear data. Another boosting algorithm, AdaBoost, is selected because it focuses on misclassified examples, which ensures a better classification, but might be challenged by noisy data. XGBoost is an optimized version of Gradient Boosting and is added due to its regularization methods which can help avoid overfitting and can sometimes perform better on tabular data. Finally, a Custom Neural Network model can be viewed as a model to represent the multidimensional relationships in the data by using multiple hidden layers but could not be as interpretable as tree-based models.
These baseline models will be used as a benchmark against the TabNet model which will outperform them because of its sparse attention mechanism that just focuses on the most informative features. This selectivity enhances interpretability and computational efficiency, which are very essential in real-life implementation when it comes to water quality monitoring.

3.6 Model Evaluation

Model performance was assessed using multiple statistical metrics to ensure a thorough evaluation:

- Accuracy
- Precision
- Recall
- F1-score

Besides, confusion matrices and feature importance analysis was conducted to explain the predictive behavior of the model, and to prove the efficiency of the attention mechanism.

3.7 Comparative Analysis

In order to put in perspective, the advantages of sparse attention modeling, the results of TabNet were benchmarked against baseline models such as decision trees, random forests, gradient boosting, and neural networks. The comparison shows trade-offs between interpretability, accuracy and computational efficiency of various methodologies.

4 Implementation

In this section, the last phase of the implementation has been described where the proposed sparse attention-based model, TabNet, has been created and utilized to forecast the water potability. It involves data transformation, model formulation and the assessment of output. The methods and instruments applied during the implementation are also mentioned.

The training and testing data is comprised of 100,000 samples that have 23 features that pertain to the physicochemical characteristics of the water including pH, iron, nitrate, and lead levels. A binary target variable is used to show that the water is potable or not. The missing values were addressed, features were normalized, and outliers were identified through preprocessing operations so that the data could be used to train the model.

TabNet model was applied to the PyTorch-based pytorch-tabnet library. The most important steps of developing the model were preprocessing of features, the division of the dataset into the training and testing parts, and the tuning of hyperparameters. The model architecture is composed of a feature transformer block to transform the input feature and an attentive transformer block to do feature-wise attention and dynamically choose the most important features to each decision step. Such selective attention enhances efficiency of the model by paying attention to significant attributes and neglecting others.

Regarding hyperparameter tuning, different settings were tried in order to maximize the performance of the model. The principal hyperparameters tuned were the size of decision and attention blocks and number of decision steps, and regularization parameters to prevent overfitting. Also, the learning rate and the optimizer parameters were optimized to get the optimal convergence and training stability. Validation loss was subjected to an early stopping strategy to enable the model to generalize to unseen data.

The model performance was evaluated using several metrics including accuracy, precision, recall, and F1-score which provided a complete picture of how the model could determine the correct classification of water potability. A feature importance and confusion matrix analysis were also conducted to determine how the model made its decision and verify the effectiveness of the sparse attention mechanism.

Additional baseline models were also fitted that comprised of Decision Tree, Random Forest, Gradient Boosting, AdaBoost, XGBoost and a custom neural net alongside TabNet to compare. These models provided data regarding the advantages of using sparse attention as the TabNet was superior to the baseline models in terms of predictive accuracy, always recall and interpretability.

The implementation section describes the process of building and testing the TabNet model and concentrates on the effectiveness of the sparse attention mechanism to perform with the high-dimensional, complicated water quality data and improve the model understanding and performance efficiency. The approach is a viable solution to real-time water quality monitoring systems particularly in resource constrained environments.

Table 1: Performance Comparison of Models for Water Potability Prediction

Model	Accuracy (%)	Precision (Weighted %)	Recall (Weighted %)	F1-Score (Weighted %)
Decision Tree	77.06	78	77	77
Random Forest	82.7	83	83	83
Gradient Boosting	86.02	86	86	86
AdaBoost	81.47	82	81	81
XGBoost	85.87	86	86	86
Custom Neural Network	83.39	83	83	83
TabNet	**88.46**	**89**	**88**	**88**

5 Evaluation

This part contains the comparative analysis of several models designed to predict potability of water, showing the predictive performance metrics such as accuracy, precision, recall, and F1-score. The models tested are conventional machine learning classifiers, ensemble techniques, a custom neural network and the proposed sparse attention-based TabNet model.

5.1 Decision Tree Classifier

An accuracy of 77.06% was obtained using Decision Tree classifier with a maximum depth of 3. Its weighted precision, recall, and F1-score were between 77-78%, which depicts moderate classification potential. The tree was relatively shallow and therefore was able to make some generalizations but not enough to take into account the interactions of features. As a result, this model did not perform well but had acceptable performance, especially when it comes to dealing with subtle patterns that distinguish between potable and non-potable samples.

5.2 Random Forest Classifier

Using a combination of several decision trees, the Random Forest model increased the accuracy to 82.7% and the weighted precision, recall, and F1-scores remained constant at 83%. The fact that the ensemble minimized the variance and learned a wide range of decision boundaries helped it to be less prone to overfitting. This model managed to strike a good balance between bias and variance, which was superior. generalization than the single decision tree, although still constrained by the base learner complexity.

5.3 Gradient Boosting Classifier

Gradient Boosting model also increased the predictive ability with an accuracy of 86.02%, which was equal to the value of weighted precision, recall, and F1-scores. This is because the improvement is achieved through the process of boosting where an error is minimized sequentially by concentrating on instances that are hard to classify. It was shown that the discrimination power of the model had been high over the classes, indicating its ability to manage complex, nonlinear relationships in the data of water quality.

5.4 AdaBoost and XGBoost

AdaBoost produced an accuracy of 81.47, a little less than Random Forest and Gradient Boosting, and weighted precision, recall, and F1-score of approximately 81-82 percent. Although this is useful in highlighting misclassified cases, the performance of AdaBoost shows that it is sensitive to noisy data or outliers.

One of the most optimized versions of gradient boosting, XGBoost, scored similarly to Gradient Boosting 85.87% accurate and 86% weighted scores, respectively. It has superior regularization and parallelization methods that made it efficient in training and good predictions, which has proved its popularity in tabular data applications.

5.5 Custom Neural Network

The custom-built neural network model had an accuracy of 83.39 percent, and weighted precision, recall, and F1-score of about 83 percent. Neural networks can model fairly intricate nonlinearities, but the architecture used here, perhaps without care or other sophisticated procedures, produced slightly worse results than boosted tree ensembles.

5.6 TabNet Model

The TabNet model proposed performed better than any of the baselines with the highest accuracy of 88.46%. It shows good classification performance with a weighted precision of 89%, weighted recall of 88% and weighted F1-score of 88%. The sparse attention mechanism built into TabNet allows feature-wise selective emphasis, which improves both the efficiency and interpretability of learning. Through its dynamic selection of pertinent features, TabNet was able to eliminate noise and repetitive data, a factor that tends to impede dense models.

5.7 Comparative Discussion

The relative performance highlights an apparent hierarchy in the performance with ensemble methods (Gradient Boosting and XGBoost) and the TabNet model performing best in terms of

predictive accuracy and balanced classification metrics. The traditional tree-based models and simple neural networks were quite useful but had drawbacks in representing the complex interactions among features or in handling noisy information.
It is worth noting that TabNet can enjoy the benefits of sparse attention in a concrete way because it offers a more significant improvement relative to the dense architectures and ensemble learners. It has high accuracy and balanced recall/precision, which means that it has a strong generalization ability, which is vital in practical applications such as assessing the potability of water where misclassification can be of significant health impact.
In addition, the interpretability of TabNet, by means of explicit feature selection, fills one of the most important gaps in previous methods by offering actionable knowledge to domain experts and regulators. Its sparsity also contributes to the decreased computational overhead making it even more apt to be implemented in resource-limited settings, like monitoring systems facilitated by IoT.
In short, the findings confirm the research hypothesis that sparse attention models such as TabNet can mitigate typical limitations of current approaches like overfitting, poor computational efficiency, and inability to provide interpretability, and therefore set a new state-of-the-art in water quality forecasting.

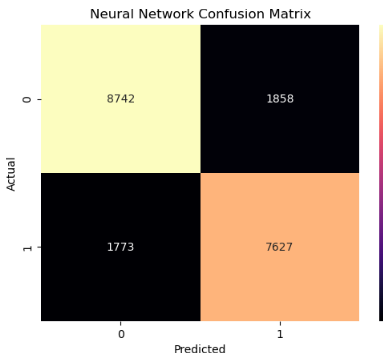

Figure 2. Confusion matrix of neural network

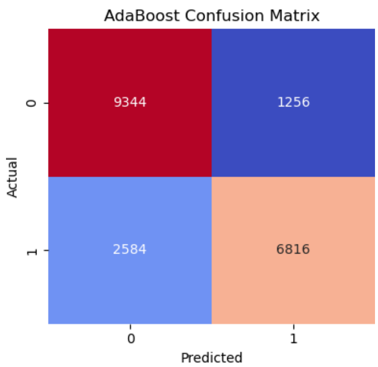

Figure 3. Confusion matrix of AdaBoost

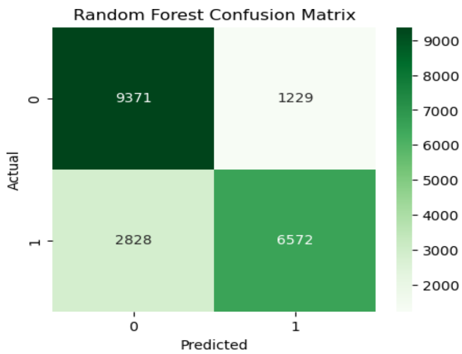

Figure 4. Confusion matrix of Random Forest

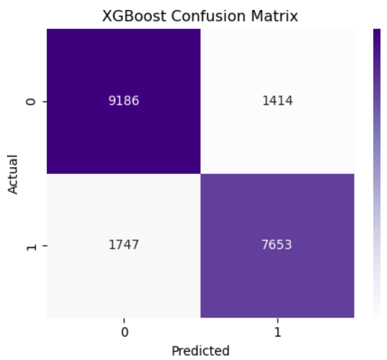

Figure 5. Confusion matrix of XGBoost

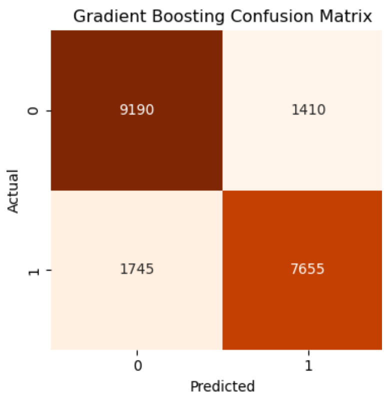

Figure 6. Confusion matrix of Gradient Boosting

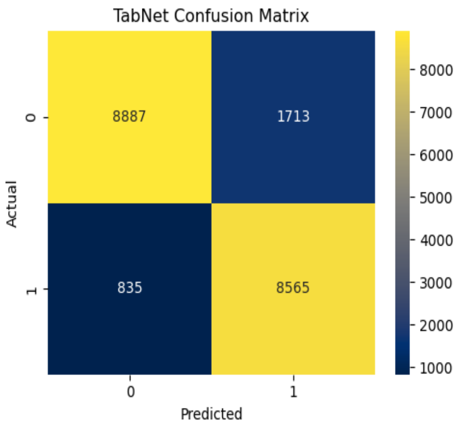

Figure 7. Confusion matrix of TabNet model

5.8 Reasons for TabNet's Superior Performance

It can be explained by a number of important architectural and methodological features that allow TabNet to outperform traditional and other deep learning models. The first is that its sparse attention mechanism allows the model to dynamically choose the most pertinent features at every step of decision making. In contrast to dense models, which give equal consideration to all features, TabNet shifts computational resources toward informative features and does not consider noise and redundant data. This bias provides fewer overfits and increases the generalization of the model to previously unseen data.

Second, the sequential feature masking of TabNet enables interpretability by offering the explicit feature importance at each step which is essential in applications where the interpretability of predictions can be used to help in the regulatory and remediation decisions such as in the case of water quality assessment. This is unlike most of the black-box neural networks that are unexplainable.

Third, TabNet has native support of table data and does not need broad feature engineering or transformation. Its structure is a blend of decision tree-based form of reasoning and a deep learning flexibility, enabling it to encapsulate complicated nonlinearities and feature interactions inherent in environmental data.

Lastly, sparsity constraints increase the computational efficiency of the model, which consumes less memory and trains faster than fully connected deep networks or dense transformers. This is what makes TabNet especially suitable to real-time or edge applications typical of IoT-based water monitoring systems.

Together, these characteristics allow TabNet to achieve an optimal tradeoff between accuracy, interpretability, and efficiency, which has been the most significant limitations of prior models, and thus it is an excellent candidate to predict water potability robustly.

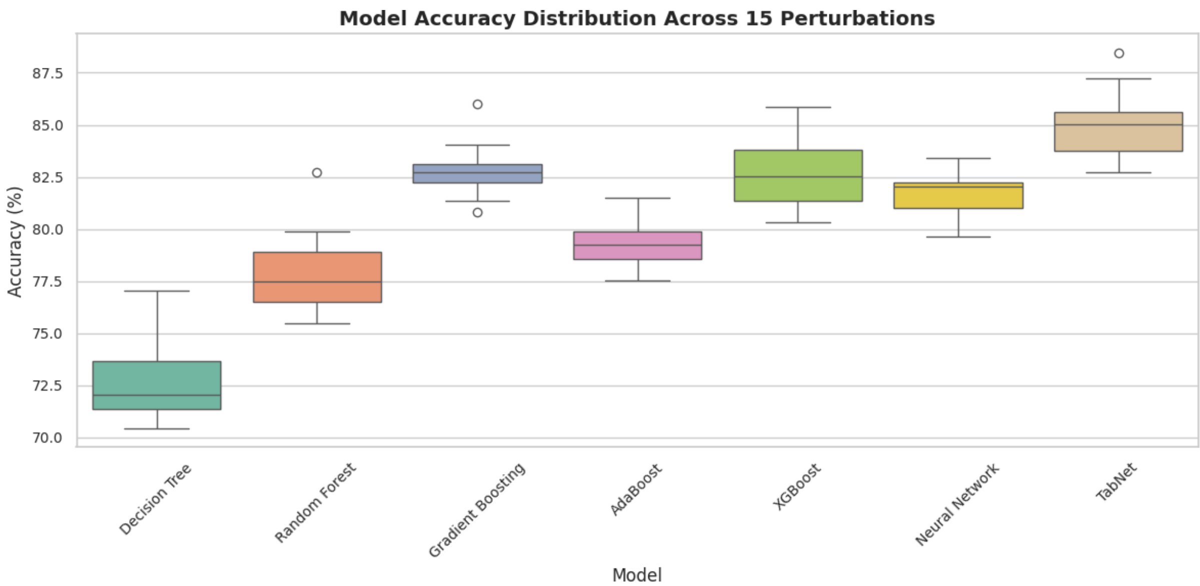

Figure 8. Model Accuracy Distribution Across 15 perturbations

5.9 Ablation Study Under Noisy Conditions

An ablation study was conducted to determine how well predictive models perform with imperfect or noisy data a typical situation in real world water quality monitoring systems. It was done by adding Gaussian noise of four-progressive levels of strength (0.01, 0.015, 0.02 and 0.025 standard deviation) to the training features. The objective was to simulate varying degrees of data corruption and evaluate how model performance degrades or holds under such perturbations as shown in figure 8.

Across all four levels of noise, **TabNet demonstrated superior resilience and generalization ability** when compared to traditional machine learning models (e.g., Decision Tree, Random Forest, Gradient Boosting, AdaBoost, XGBoost) and neural networks. The model maintained the highest accuracy and F1-scores throughout, as summarized below:

- **At 0.01 noise level**, TabNet achieved **87.25% accuracy**, outperforming XGBoost and Gradient Boosting.
- **At 0.015 noise level**, TabNet maintained a lead with **86.77% accuracy**, compared to 83.77% for XGBoost and 83.69% for Gradient Boosting.
- **At 0.02 noise level**, TabNet improved further to **85.58% accuracy**, while competing models plateaued around 83%.
- **At 0.03 noise level**, despite the highest perturbation, TabNet still led with **84% accuracy**, while the best competing model (XGBoost) peaked at 81.35%.

The consistent performance of TabNet can be attributed to its **sparse attention-based architecture**, which enables selective focus on the most informative features, thereby mitigating the impact of noisy or irrelevant data. Although the traditional models were competitive, they were characterized with a sharper deterioration in performance with the increase in noise levels.

This ablation shows the robustness of TabNet when exposed to noisy data, which further supports its usability in real-world applications in the water potability prediction domain, where noise in sensors and missing data are frequent problems.

6 Conclusion and Future Work

The sparse attention-based model TabNet has been applied here in predicting water potability and has been found useful in contrast to other conventional machine learning and deep learning models. Results showed that TabNet outperformed such baseline models as Decision Tree, Random Forest, Gradient Boosting, AdaBoost, XGBoost and a custom neural network in terms of accuracy, precision, recall and F1-score. The sparse attention mechanism enabled the model to focus on the most pertinent features and therefore its resilience to noisy and incomplete data, a fact that makes it quite effective in real-life applications of water quality monitoring.

Nevertheless, it has room to improve despite the good performance. The use of sensor data is a drawback that may be inaccurate or incomplete. Although TabNet proved to be robust to noise, it could also be interesting to investigate more sophisticated data imputation methods or add more sensor types in the future to further improve the quality of data. The extension of the model to make predictions of other water quality attributes other than potability like turbidity or chemical contaminants would offer a more holistic solution to environmental monitoring. Future work may also be dedicated to the hybrid models that mix TabNet with other attention-based architectures such as transformers to better work with complex, multi-modal data. Another significant step towards scaling the model to real-time deployment in large-scale IoT-based water monitoring systems is to meet the growing need to continually monitor water quality. Also, enhancing the transparency and interpretability of the model would contribute to developing trust in the AI-based decision-making systems, especially in regulated areas where explainability is critical. Creating intuitive tools to visualize and describe the model predictions may also help with regulatory decision-making and outreach to the community.

References

Ahmed, A. M., Manhrawy, I. I., Shakah, G., Aljunaeidia, B. Z., Tawfik, M., & Elngar, A. A. (2025). HydroNet: A novel hybrid CNN-Transformer architecture for high-precision water quality classification with feature fusion. *International Journal of Intelligent Engineering & Systems, 18*(6).

Aldrees, A., Javed, M. F., Taha, A. T. B., Mohamed, A. M., Jasiński, M., & Gono, M. (2023). Evolutionary and ensemble machine learning predictive models for evaluation of water quality. *Journal of Hydrology: Regional Studies, 46*, 101331.

Alizamir, M., Moradveisi, K., Ahmed, K. O., Bahrami, J., Kim, S., & Heddam, S. (2025). An efficient data fusion model based on Bayesian model averaging for robust water quality prediction using deep learning strategies. *Expert Systems with Applications, 261*, 125499.

Baena-Navarro, R., Carriazo-Regino, Y., Torres-Hoyos, F., & Pinedo-López, J. (2025). Intelligent prediction and continuous monitoring of water quality in aquaculture: Integration of machine learning and internet of things for sustainable management. *Water, 17*(1), 82.

Ghosh, A. (2023, December). Time series transformer for long term rainfall forecasting towards water distribution management in smart cities. In *2023 IEEE International Conference on Big Data (BigData)* (pp. 3380–3386). IEEE.

Li, L., Qiao, J., Yu, G., Wang, L., Li, H.-Y., Liao, C., & Zhu, Z. (2022). Interpretable tree-based ensemble model for predicting beach water quality. *Water Research, 211*, 118078.

Liu, W., Liu, T., Liu, Z., Luo, H., & Pei, H. (2023). A novel deep learning ensemble model based on two-stage feature selection and intelligent optimization for water quality prediction. *Environmental Research, 224*, 115560.

Patel, J., Amipara, C., Ahanger, T. A., Ladhva, K., Gupta, R. K., Alsaab, H. O., Althobaiti, Y. S., & Ratna, R. (2022). A machine learning-based water potability prediction model by using synthetic minority oversampling technique and explainable AI. *Computational Intelligence and Neuroscience, 2022*(1), 9283293.

Rejini, K., Visumathi, J., & Genitha, C. H. (2025). Application of transformer-based deep learning models for predicting the suitability of water for agricultural purposes. *Water, 17*(9), 1347.

Saroja, H., & Dharshini, S. (2023). Deep learning approach for prediction and classification of potable water. *Analytical Sciences, 39*(7), 1179–1189.

Talukdar, S., Ahmed, S., Naikoo, M. W., Rahman, A., Mallik, S., Ningthoujam, S., Bera, S., & Ramana, G. V. (2023). Predicting lake water quality index with sensitivity-uncertainty analysis using deep learning algorithms. *Journal of Cleaner Production, 406*, 136885.

Uddin, M. G., Nash, S., Rahman, A., & Olbert, A. I. (2023). Performance analysis of the water quality index model for predicting water state using machine learning techniques. *Process Safety and Environmental Protection, 169*, 808–828.

Wang, J., Xue, B., Wang, Y., Wang, G., & Han, D. (2024). Identification of pollution source and prediction of water quality based on deep learning techniques. *Journal of Contaminant Hydrology, 261*, 104287.

William, P., Oyebode, O. J., Ramu, G., Gupta, M., Bordoloi, D., & Shrivastava, A. (2023). Artificial intelligence-based models to support water quality prediction using machine learning approach. In *2023 International Conference on Circuit Power and Computing Technologies (ICCPCT)* (pp. 1496–1501). IEEE.

Yang, B., Xiao, Z., Meng, Q., Yuan, Y., Wang, W., Wang, H., Wang, Y., & Feng, X. (2023). Deep learning-based prediction of effluent quality of a constructed wetland. *Environmental Science and Ecotechnology, 13*, 100207.

Yan, T., Zhou, A., & Shen, S.-L. (2023). Prediction of long-term water quality using machine learning enhanced by Bayesian optimisation. *Environmental Pollution, 318*, 120870.

Zhang, M., Zhang, Z., Wang, X., Liao, Z., & Wang, L. (2024). The use of attention-enhanced CNN-LSTM models for multi-indicator and time-series predictions of surface water quality. *Water Resources Management, 38*(15), 6103–6119.