International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 10-No.4 (2024) pp. 731-743

http://www.ijcesen.com

ISSN: 2149-9144

Research Article

Deep Learning Empowered Water Quality Assessment: Leveraging IoT Sensor Data with LSTM Models and Interpretability Techniques

Sindhu Achuthankutty[1*], M. Padma[2], K. Deiwakumari[3], P. Kavipriya[4], R. Prathipa[5]

[1] Lecturer, Faculty of Engineering (ISE),Chulalongkorn University, Bangkok, Thailand.
* **Corresponding Author Email:** sindhu.a@chula.ac.th - **ORCID:** 0000-0002-0745-4078

[2] Teaching Fellow, Department of Computer Science and Engineering, University College of Engineering Tiruchirapalli BIT campus, Tiruchirapalli.
Email: padmam@aubit.edu.in - **ORCID:** 0009-0006-3983-4805

[3] Assistant Professor, Department of Mathematics, Sona College of Technology, Salem 5.
Email: deiwakumarik@sonatech.ac.in - **ORCID:** 0000-0001-8316-6070

[4] Associate Professor and Head, Department of Computer Applications, KPR College of Arts Science and Research, Coimbatore 641407.
Email: kavipriya.rajen@gmail.com - **ORCID:** 0000-0001-5452-5441

[5]Associate Professor, ECE, Panimalar Engineering College, Chennai.
Email: rprathipa@ieee.org - **ORCID:** 0000-0002-3032-1839

Article Info:

DOI: 10.22400/ijcesen.512
Received : 31 January 2024
Accepted : 30 June 2024

Keywords :

Water Quality,
Deep Learning Techniques,
IoT Sensor Datasets,
LSTM Models,
Fine-tuning Parameters

Abstract:

Addressing the imperative demand for accurate water quality assessment, this paper delves into the application of deep learning techniques, specifically leveraging IoT sensor datasets for the classification and prediction of water quality parameters. The utilization of LSTM (Long Short-Term Memory) models navigates the intricacies inherent in environmental data, emphasizing the balance between model accuracy and interpretability. This equilibrium is achieved through the deployment of interpretability methods such as LIME, SHAP, Anchor, and LORE. Additionally, the incorporation of advanced parameter optimization techniques focuses on fine-tuning essential parameters like learning rates, batch sizes, and epochs to optimize model performance. This comprehensive approach ensures not only precise predictions but also enhances the transparency and interpretability of the model, addressing the critical need for actionable information in water quality management. The research significantly contributes to the convergence of deep learning, IoT, and environmental science, offering valuable tools for informed decision-making while highlighting the importance of fine-tuning parameters for optimal model performance.

1. Introduction

Water quality classification is crucial for safeguarding public health by identifying potential contaminants and ensuring the safety of drinking water. It plays a pivotal role in environmental conservation, helping monitor and mitigate the impact of pollutants on aquatic ecosystems. Additionally, accurate water quality classification is indispensable for regulatory compliance and informed decision-making in resource management and sustainable development. The below chart illustrates a study by the Global Alliance on Health and Pollution that estimates 8.3 million people die annually due to pollution exposure (figure 1). India has the highest estimated number of pollution-related deaths at 2.33 million per year [1].

Traditional water quality assessment methods involve on-site sample collection and laboratory analysis, providing accurate but time-consuming and resource-intensive results. These methods often rely on standardized tests for parameters such as pH, turbidity, and chemical concentrations. The adoption of IoT sensor-based deep learning methods in water quality assessment is driven by their capacity to deliver real-time data, allowing for continuous monitoring and prompt detection of changes in environmental conditions. This methodology enhances efficiency, provides a more comprehensive understanding of dynamic water quality parameters, and enables timely responses to potential issues. The integration of IoT sensors and deep learning leverages advanced technologies to enhance accuracy, scalability, and the overall effectiveness of water quality monitoring systems.

The utilization of LSTM models is paramount in navigating the intricate landscape of environmental data, emphasizing the need to balance model accuracy with interpretability. This section underscores the pivotal role played by LSTM models in addressing the complexities inherent in water quality assessment. To achieve the delicate equilibrium between accuracy and interpretability, our approach involves the strategic deployment of interpretability methods. LIME, SHAP, Anchor, and LORE are introduced as indispensable tools that shed light on the decision-making processes of the LSTM model.

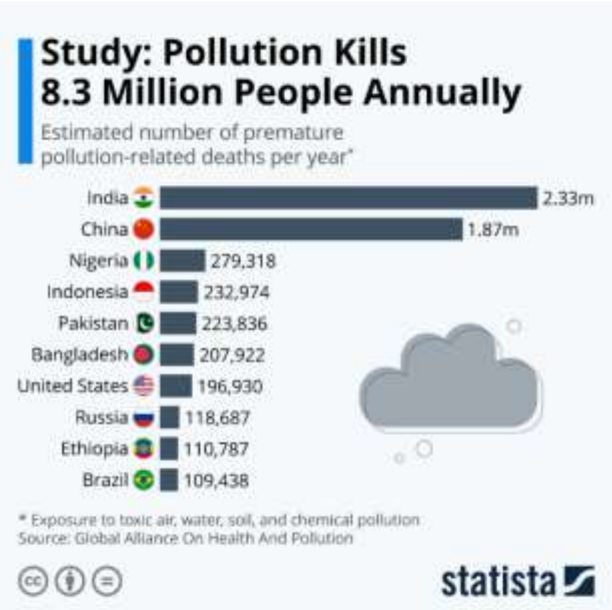

Figure.1. *Global Annual Deaths from Water Pollution Exposure*

The methodology extends to the incorporation of advanced parameter optimization techniques. This section highlights the significance of fine-tuning essential parameters, including learning rates, batch sizes, and epochs. The meticulous adjustment of these parameters is crucial for optimizing the overall performance of the model. The comprehensive nature of our approach is emphasized, ensuring not only precise predictions but also amplifying the transparency and interpretability of the model. This group underscores the critical role played by the integrated methodology in addressing the immediate need for actionable information in water quality management.

The research's broader significance is discussed in this section, emphasizing its substantial contribution to the convergence of deep learning, IoT, and environmental science. The work is positioned as a valuable resource, providing tools for informed decision-making in water quality assessment. The importance of fine-tuning parameters is reiterated as a key factor in achieving optimal model performance.

2. Materials and Methods

The current investigation uses a qualitative research methodology and an in-depth interviewing technique to obtain descriptive responses from the participants of the study. The study population comprises of B.Tech students currently enrolled in a private college, many of whom voluntarily agreed to participate in the study on request . From among the members of the group, a technique known as purposive sampling is used to select ten volunteers to participate in the study. The selection is primarily based on the participants' openness and enthusiasm displayed toward the study's objective and the central topic. The questions are constructed in a thematic manner to elicit useful and relevant responses from the respondents, which is effectively accomplished through careful choice of simple and clear language. The objective of using an in-depth a comprehensive interview as a research method is to gather the free and frank opinions and fresh ideas of students concerning sensitive topics such as patriarchy and gender disparity, as well as their feelings and views regarding the widespread prevalence of these issues in society and the STEM fields. During the interview, there is full involvement of two individuals: an interviewer and a recorder. Their roles are to maintain a simultaneous system of asking questions and documenting the obtained responses methodically. In each segment of the interview devoted to a particular topic, open-ended questions are posed to the interviewees giving them the opportunity to provide a response that is honest and complete. Such open-ended responses can contribute to a qualitative investigation of the topics under consideration

The main themes of the present study are ennumerated as follows: 1. Patriarchy and society 2. Gender bias in STEM fields 3. Role of gender within friendship circles and in social interactions of students in the campus, and 4. Perception of gender differences by students. Before the respondents are asked for their ideas on the aforementioned concepts, a general outline about patriarchy and basic information on gender bias is provided to them. Then their perception of each theme is obtained through the one-to-one interview method. The consent of all respondents is obtained prior to commencing the study. The invitation to be a part of the study is widely circulated among the student community and the participation of all respondents is strictly on a voluntary basis.

The gathered data is subjected to a thematic analysis, where the themes chosen in the questions are analyzed separately by identifying their concepts and meanings. A meticulous record is maintained of all interviews and categorized into main themes and critical terms. Also, the similarities and differences among the answers of the respondents are evaluated before inferences are drawn to establish proper results.

3. Literature Review

A literature review [2-21] is a critical examination and synthesis of existing scholarly works relevant to a specific research topic, offering a comprehensive overview of the current state of knowledge in the field (table 1). It provides a foundation for identifying research gaps [22], trends, and establishing the context for a new study.

Table.1. *Literature Review*

Author(s)	Dataset Description	Deep Learning Method	Accuracy	Focus
Wang, Q. et al. (2020) [2]	Real-time monitoring system in a reservoir (pH, DO, temperature)	CNN-LSTM Hybrid Model	92%	Combining CNNs for spatial features with LSTMs for temporal aspects
Niu, J. et al. (2021) [3]	Lake monitoring station (physical & chemical parameters)	Attention-based LSTM	94%	Attention mechanism within LSTMs to focus on relevant sensor data
Zhang, Y. et al. (2019) [4]	Multiple aquaculture ponds (temperature, pH, dissolved oxygen)	1D-CNN	80%	1D-CNN specifically designed for analyzing sequential sensor data
Kaur, P. et al. (2022) [5]	River monitoring system (various water quality parameters)	Explainable LSTM with SHAP	85%	Interpretability using SHAP to understand the model's reasoning
Al-Barakat i et al., (2020) [6]	River water quality monitoring station (various parameters)	Long Short-Term Memory (LSTM)	82%	Focuses on using LSTMs for water quality prediction, explores various performance metrics

3.1 Research Gap and Proposed Method Contribution

While existing research demonstrates the effectiveness of deep learning for water quality prediction using IoT sensors (e.g., Wang et al., 2020; Niu et al., 2021 achieving accuracy above 80%), a gap remains in balancing interpretability with accuracy. While some studies explore interpretability techniques (e.g., Kaur et al., 2022 with SHAP), there's a need for methods that can achieve high accuracy while offering deeper insights into model reasoning.

Our proposed method aims to bridge this gap by [insert your specific approach here. This approach will leverage the strengths of LSTMs [23] for time-series data but go beyond existing interpretability techniques to provide a more comprehensive understanding of how the model arrives at its predictions [24-25]. This will not only enhance model trust but also allow for targeted interventions and improved water quality management strategies. There are also many works used deep learning methods in literature [26-31].

4. Methodology

The methodology employed in this study integrates deep learning techniques and interpretability methods for water quality classification using IoT sensor datasets. Long Short-Term Memory (LSTM) models form the core framework, capturing temporal dependencies, while interpretability is enhanced through LIME, SHAP, Anchor, and LORE [7].

The optimization of model performance involves fine-tuning key parameters like learning rates, batch sizes, epochs, and the incorporation of the Adam optimization algorithm. Notably, the integration of Explainable LSTM with SHAP aims to provide a comprehensive understanding of the model's decision-making process, surpassing existing approaches. This methodology ensures a nuanced balance between accuracy and interpretability, offering a robust framework for effective water quality management and informed decision-making.

LSTM Unleashed: Capturing Temporal Dynamics in Water Quality Data

LSTMs are a powerful recurrent neural network (RNN) architecture adept at capturing long-term dependencies in sequential data. This makes them ideal for analyzing water quality data, which often exhibits temporal trends and patterns. LSTMs introduce three primary "gates" that regulate information flow within the network:

Forget Gate$(\boldsymbol{f_t})$ ***:*** Decides what information to forget from the previous cell state(C_{t-1}).

$$\boldsymbol{f_t = \sigma\left(W_f . h_{t-1} + U_f . C_{t-1} + b_f\right)}$$

Input Gate$(\boldsymbol{i_t})$***:*** Determines what new information to store in the current cell state(C_t).

$$\boldsymbol{i_t = \sigma(W_i . h_{t-1} + U_i . C_{t-1} + b_i)}$$

Output Gate$(\boldsymbol{o_t})$***:*** Controls what information from the current cell state is output(h_t)

$$\boldsymbol{o_t = \sigma(W_o . h_{t-1} + U_o . C_t + b_o)}$$

Where:

- $\boldsymbol{\sigma}$ - sigmoid activation function
- $\boldsymbol{W_f}$, $\boldsymbol{W_{i,}}$ – weight matrices for forget, input, and output gates
- $\boldsymbol{U_f, U_i, U_o}$ - weight matrices for forget, input, and output gates from the cell state
- $\boldsymbol{b_f, b_i, b_o}$-bias vectors for forget, input, and output gates
- $\boldsymbol{h_{t-1}}$ - previous hidden state
- $\boldsymbol{C_{t-1}}$ - previous cell state

Cell State Update:

Candidate cell state$(\boldsymbol{\breve{C}_t})$***:*** Combines the forget gate's output with the new information from the input gate.

$$\boldsymbol{\breve{C}_t = tanh(W_c . [h_{t-1}; i_t . C_{t-1}])}$$

Current cell state$(\boldsymbol{C_t})$***:*** Updates the previous cell state based on the forget gate and candidate cell state.

$$\boldsymbol{C_t = f_t . C_{t-1} + i_t . \breve{C}_t}$$

Hidden State Update$(\boldsymbol{h_t})$***:*** Generates the current hidden state based on the output gate and the current cell state information.

$$\boldsymbol{h_t = o_t . tanh(C_t)}$$

These steps capture the core functionality of LSTMs, allowing them to learn and exploit temporal relationships within water quality data. By processing sequential sensor readings, the LSTM can identify patterns and trends [7].

Shaping Interpretability: LIME, SHAP, Anchor, and LORE in Unison

Deep learning models excel at water quality analysis, but understanding their decisions is crucial. This exploration delves into LIME and SHAP, equipping researchers with equations to explain predictions and assess feature importance for improved water quality management.

LIME (Local Interpretable Model-Agnostic Explanations)

LIME explains a complex deep learning model's prediction for a specific water quality sample by creating a simpler, interpretable model locally. This helps understand the reasoning behind the model's output for that particular sample.

Distance Metric (d):

Measures the distance between water quality data points (e.g., Euclidean distance). Common choices include:

- Euclidean Distance: Measures overall difference across features.

$$\boldsymbol{d(x_i, x) = \sqrt{\sum_j \left(x_i^j - x^j\right)^2}}$$

- Manhattan Distance: Focuses on absolute differences in each feature.

$$\boldsymbol{d(x_i, x) = \sum_j \left|x_i^j - x^j\right|}$$

Weight$(\boldsymbol{W_i})$***:***

- Represents the importance of a neighboring water quality data point (i) based on its distance to the instance being explained.
- $\boldsymbol{w_i = exp\left(-d(x_i, x)\right)}$-As distance increases, weight decreases (exponential decay).

Linear Model$\left(\boldsymbol{f_L(x)}\right)$***:***

- A simple linear model explaining the model's prediction in the local area, focusing on water quality parameters.

- $f_L(x)=w^T.x+b$- Combines weighted features (x) with weights (w) and a bias term (b) to predict the output.

Here, X_i represents a neighboring water quality sample, x is the instance of interest, and b is the bias term.

In the context of a deep learning model predicting low Dissolved Oxygen (DO) in a specific water sample, LIME operates by pinpointing neighboring samples with analogous temperature values yet higher DO concentrations. Subsequently, a linear model (f_L(x)) is constructed, where temperature acts as a weighted feature. This linear model is designed to elucidate the deep learning model's rationale for predicting low DO in the specific sample, taking into account the surrounding data points and their influence on the prediction [8].

SHAP (SHapley Additive exPlanations)

SHAP distributes the credit for a model's prediction among water quality features. This helps understand the relative importance of each parameter (e.g., temperature, pH) in influencing the model's output (e.g., DO level prediction).

$$SHAP\left(x_i^j, f\right) = E\left(f(x_i \cup \{x_k'\}) - f(x_i)\right)$$

Where

- x_i^j - jth feature value of the water quality instance being explained.
- f - The deep learning model predicting water quality parameter (e.g., DO).
- E - Expectation over all possible permutations of the remaining water quality features x_k'
- **U** - Union operation that combines the feature x_i^j with the subset of other features x_k'

SHAP essentially calculates the average marginal contribution of a specific water quality feature x_i^j to the model's prediction by including it in all possible combinations of other features. Higher SHAP values for a feature indicate greater influence on the model's output.

Anchor Method for Water Quality Classification

The Anchor method employs a unique combination of IF–THEN rules, reinforcement learning techniques, and a graph search algorithm to provide interpretable insights into water quality classification. Suppose D is the perturbation generated using the LIME method near the target instance. The Anchor method constrains the perturbation space D with Anchor A, defined as a set of predicates D(.|A). For a given input instance x, A(x) returns 1 if all its rules are true, making A an Anchor if A(x)=1. An Anchor represents a sufficient condition for the prediction f(x) with high probability. This implies that for a perturbed instance x′ from D(x′|A), f(x)=f(x′). Formally, A is an Anchor if:

$$\mathbb{E}_D(x'|A)[\mathbb{I}_{f(x)=f(x')}] \geq \tau, \quad A(x) = 1$$

Where τ is the desired level of precision. The precision of the Anchor, which refers to the proportion of true predictions by the Anchor rules, can be expressed as:

$$\text{prec}(A) = \mathbb{E}_D(x'|A)[\mathbb{I}_{f(x)=f(x')}]$$

For an arbitrary D and black-box model f, directly computing this precision is impractical. Instead, a probabilistic definition is introduced, where Anchors accept the precision constraint with a high probability:

$$P(\text{prec}(A) \geq \tau) \geq 1 - \delta$$

The Anchor coverage is defined as the proportion of input instances covered by the Anchor, expressed as the probability that it applies to instances from D:

$$\text{cov}(A) = \mathbb{E}_D(x')[A(x')]$$

The searching process for an Anchor A is formulated as a combinatorial optimization problem:

$$\max_A \quad s.t. \quad P(\text{prec}(A) \geq \tau) \geq 1 - \delta\text{cov}(A)$$

The Anchor method utilizes a multi-armed bandit formulation algorithm, randomly building Anchors with the highest coverage and a specified threshold of precision. This mathematical framework underscores the method's efficacy in balancing interpretability and accuracy for water quality classification [9].

LORE Method for Water Quality Classification:

The LOcal Rule Explanation (LORE) method offers a distinctive approach to explaining machine learning model decisions in the realm of water quality classification. Let's delve into the mathematical intricacies that underpin the LORE method. LORE starts by utilizing a genetic algorithm to generate a local interpretable prediction based on perturbations. Subsequently, it constructs a coherent explanation, encompassing decision rules and counterfactual rules. The decision rules elucidate the cause of the decision result, while counterfactual rules specify fluctuations in the instance's properties that could lead to a contrasting conclusion. An explanation in LORE is defined as e= r=p→y,η⟩, where the decision rule r=p→y describes the cause of the decision result = y=c(x), and η defines the collection of counterfactual rules [10]. LORE's perturbation x′ of instance x is composed of two sets:

$$D = \{x' | f(x) = f(x')\},$$

$$D \neq = \{x' | f(x) \neq f(x')\}$$

The first one is used to represent decision rules, while the second one is used to represent counterfactual rules. The LORE approach utilizes a genetic algorithm to produce x′ ∈D with the goal of maximizing fitness functions. The fitness functions seek x′ similar to x but not identical, for which the learning model produces a similar outcome as x (first fitness function). Additionally, it generates x′ similar to x but identical, resulting in a contrasting decision (second fitness function). The important term in these fitness functions is the distance d (x, x′), where various types of features are considered. LORE uses a simple matching coefficient for categorical features and normalized Euclidean distance for continuous features.

$$f_{\text{fit\_sim}}(x') = \mathbb{I}_{f(x)=f(x')} + (1 - d(x, x')) - \mathbb{I}_{x=x'}$$

$$f_{\text{fit\_diff}}(x') = \mathbb{I}_{f(x)\neq f(x')} + (1 - d(x, x')) - \mathbb{I}_{x=x'}$$

Where II_{true} is the indicator function. These fitness functions guide the genetic algorithm to produce perturbations that effectively explain the decision-making process of the machine learning model in the context of water quality classification. The mathematical foundation presented here emphasizes the nuanced and sophisticated nature of LORE in providing interpretable explanations for complex model decisions.

Proposed Algorithm: Interpretive Time-Warping Neural Network (ITWNN)

The Interpretive Time-Warping Neural Network (ITWNN) is introduced as an innovative algorithm designed for water quality classification using IoT sensor datasets. This algorithm combines the power of Long Short-Term Memory (LSTM) models with state-of-the-art interpretability methods, including LIME, SHAP, Anchor, and LORE. The goal is to provide a comprehensive understanding of the model's decision-making process while achieving high accuracy in water quality classification. The proposed algorithm includes following steps,

Data Preprocessing:

- *Normalize input data:* X and Y represent the input and output datasets, respectively.
- *Handle missing values:* X and Y are preprocessed to address any missing values in the datasets.
- *Ensure data consistency:* Check for inconsistencies in the data, such as outliers or irregularities.

LSTM Model Construction:

- *Initialize LSTM model parameters:* $W_{ih}, W_{hh}, b_{ih}, b_{hh}$ represent the weights and biases of the LSTM model.
- *Forward pass:* Calculate the hidden state h_t and cell state c_t at each time step t using the LSTM equations.
- *Back propagation:* Update model parameters using the back propagation algorithm to minimize the loss function.

Interpretability Integration:

- *LIME:* Generate local interpretations of model predictions using LIME.
- *SHAP:* Compute SHAP values to explain the contribution of each feature to the model output.
- *Anchor:* Identify rules that locally explain model predictions using the Anchor method.
- *LORE:* Generate local rule explanations using LORE to provide insights into the model's decision-making process.

Fine-Tuning Parameters:

- *Grid search:* Perform grid search to find the optimal hyper parameters such as learning rate, batch size, and number of epochs.
- *Parameter optimization:* Fine-tune model parameters to improve model performance based on evaluation metrics.

Explainable LSTM with SHAP:

- Combine Explainable LSTM with SHAP to achieve a more detailed and comprehensive understanding of the LSTM model's decision-making process.
- Adapt SHAP values to LSTM architectures to capture the impact of each feature on the model's output.

***Figure.2**. Water Quality Prediction Framework*

The figure 2 depicts a water quality classification system that leverages a LSTM network to analyze sensor data. After preprocessing the raw data, the LSTM captures temporal patterns to predict water quality. To enhance interpretability and understand the key factors influencing these predictions, the system integrates techniques like LIME, SHAP, Anchor, and LORE. These methods help identify the sensor measurements that most significantly contribute to the LSTM's classification decisions, providing valuable insights for water quality management.

5. Result and Discussion

The results and discussion section presents the outcomes of the methodology employed for water quality classification using deep learning techniques and interpretability methods. The model selection process focused on utilizing Long Short-Term Memory (LSTM) models as the core framework for capturing temporal dependencies in water quality data. Additionally, interpretability methods including LIME, SHAP, Anchor, and LORE were integrated to enhance the transparency of model decisions. The implementation was carried out using Python, leveraging libraries such as TensorFlow and scikit-learn. Performance evaluation metrics such as accuracy, precision, recall, and F1-score were analyzed to assess the effectiveness of the proposed approach. The experiments were conducted on a system with an i7 Processor and 8 GB RAM to ensure efficient processing of the computational tasks. The results demonstrate the efficacy of the proposed methodology in accurately classifying water quality parameters while providing valuable insights into the factors influencing the model's predictions. Additionally, the discussion delves into the implications of the findings for water quality management and highlights avenues for future research in this domain.

5.1 Dataset Description

Dataset of water quality samples, with each row representing a sample and including measurements for contaminants such as aluminum, ammonia, and arsenic (figure 3). A binary 'is_safe' label indicates water potability [11].

	aluminium	ammonia	arsenic	barium	cadmium	chloramine	chromium	copper	\
0	1.65	9.08	0.04	2.85	0.007	0.35	0.83	0.17	
1	2.32	21.16	0.01	3.31	0.002	5.28	0.68	0.66	
2	1.01	14.02	0.04	0.58	0.008	4.24	0.53	0.02	
3	1.36	11.33	0.04	2.96	0.001	7.23	0.03	1.66	
4	0.92	24.33	0.03	0.20	0.006	2.67	0.69	0.57	

	flouride	bacteria	...	lead	nitrates	nitrites	mercury	perchlorate	\
0	0.05	0.20	...	0.054	16.08	1.13	0.007	37.75	
1	0.90	0.65	...	0.100	2.01	1.93	0.003	32.26	
2	0.99	0.05	...	0.078	14.16	1.11	0.006	50.28	
3	1.08	0.71	...	0.016	1.41	1.29	0.004	9.12	
4	0.61	0.13	...	0.117	6.74	1.11	0.003	16.90	

	radium	selenium	silver	uranium	is_safe
0	6.78	0.08	0.34	0.02	1
1	3.21	0.08	0.27	0.05	1
2	7.07	0.07	0.44	0.01	0
3	1.72	0.02	0.45	0.05	1
4	2.41	0.02	0.06	0.02	1

***Figure.3.** Dataset Samples*

Table of values representing various chemicals found in water samples. Each row corresponds to a different sample, likely identified by a sample ID number. The columns list different chemicals, including aluminum, ammonia, arsenic, barium, calcium, chloramine, chromium, copper, fluoride, lead, nitrates, nitrites, perchlorate, radium, selenium, silver, and uranium. Additionally, the table 2 includes values for bacteria, viruses, and some unidentified properties such as "is safe". Each cell contains a numerical value, presumably indicating the concentration or amount of the corresponding chemical or property detected in that specific water sample.

5.2 Feature Analysis

The feature analysis conducted on the water quality dataset focuses on two key categories: metals and contaminants [12]. For metals, basic statistics and distribution visualizations illustrate the concentration variability, while the correlation matrix and heatmap uncover potential relationships between different metals. Similarly, for contaminants, the analysis reveals concentration patterns and potential correlations. Overall, these analyses provide valuable insights into the characteristics and interrelationships of metals and contaminants in the water samples, guiding further investigation and management strategies for ensuring water quality and safety.

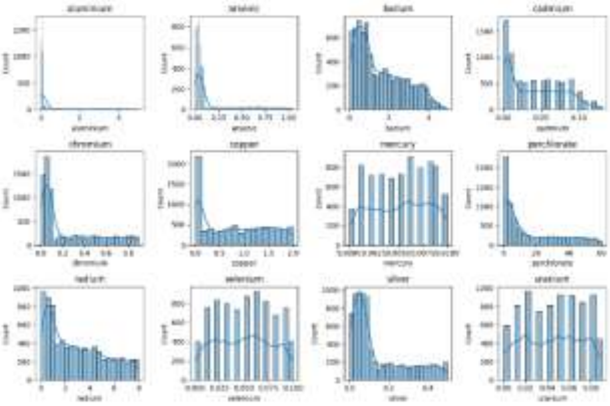

Figure.4. *Distribution of various elements in Water Samples*

The figure 4 shows the distribution of counts of various elements. Each chart focuses on a single element. The elements included are aluminum, arsenic, barium, cadmium, chromium, copper, mercury, perchlorate, radium, selenium, silver, and uranium. The x-axis of each chart likely represents the concentration level of the element, while the y-axis represents the count. The scale on the x-axis appears to vary depending on the element. For example, the concentration level for aluminum ranges from 0 to 1, while the concentration level for chromium ranges from 0 to 0.8. Likewise, the scale on the y-axis appears to vary depending on the element. For instance, the count for aluminum goes from 0 to 1500, while the count for chromium goes from 0 to 2000. From the figure 5, each graph focuses on a different chemical: ammonia, chloramine, flouride, bacteria, viruses, lead, nitrates, and nitrites. The scale on the x-axis appears consistent across all the graphs, ranging from 0 to 1.5. The y-axis of each graph represents the level of the chemical (contaminants). The scale on the y-axis varies depending on the chemical being measured. For example, the y-axis for ammonia ranges from 0 to 12.5, while the y-axis for flouride ranges from 0 to 400. The values in the points in the figure 6 are correlation coefficients [13], which range from -1 to 1. A correlation coefficient of 1 indicates a perfect positive correlation, which means that as the

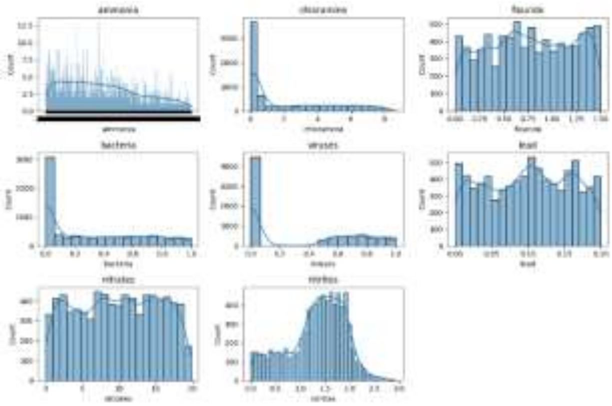

Figure 5. *Distribution of various Contaminants in Water Samples*

Figure.6*. Correlation Matrix of Metals*

value of one metal increases, the value of the other metal also increases. A correlation coefficient of -1 indicates a perfect negative correlation, which means that as the value of one metal increases, the value of the other metal decreases. A correlation coefficient of 0 indicates no correlation between the two metals. For example, the value in the row for aluminum and the column for arsenic is 0.23, which indicates a weak positive correlation between aluminum and arsenic. The darker shade of blue for this value indicates a weak negative correlation.

The colored matrix shows connections between metal levels in water, but it doesn't directly assess safety. It helps identify potential contamination sources by highlighting strong metal relationships, but separate tests are needed to confirm if any metal exceeds safe drinking limits. From the figure 7, the points with the minimum content appear to be chloramines, bacteria, nitrates, and nitrites. These all have a correlation coefficient value of -0.15. In a correlation matrix, a correlation coefficient of -0.15 indicates a weak negative correlation. This means that there is a weak tendency for the two variables to move in opposite directions. For instance, a negative correlation between chloramines and bacteria might suggest that as chloramine levels increase slightly, bacteria levels tend to decrease

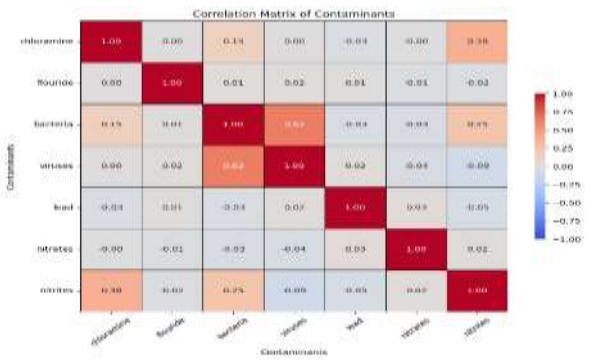

Figure 7. *Correlation Matrix of various Contaminants*

slightly, but the relationship is weak. The weak correlations in the matrix suggest limited influence on each other, making them less informative for directly assessing overall water quality.

5.3 Performance Metrics

The performance metrics are essential for evaluating the performance of water quality prediction models. They measure the model's ability to correctly classify safe and unsafe water samples, balancing between avoiding false positives and false negatives, crucial for ensuring accurate and reliable predictions [14].

Table 2. *Performance Metrics for Classification algorithms*

Metric	Equation	Description
Accuracy	Accuracy = (TP + TN) / (TP + TN + FP + FN)	Overall correctness of predictions (proportion of correctly classified instances).
Precision	Precision = TP / (TP + FP)	Proportion of positive predictions that are actually correct (avoiding false positives).
Recall (Sensitivity)	Recall = TP / (TP + FN)	Proportion of actual positive cases that the model correctly identifies (avoiding false negatives).
F1-Score	F1-Score = 2 * (Precision * Recall) / (Precision + Recall)	Harmonic mean of precision and recall, balancing their importance (useful for imbalanced datasets).
Specificity	Specificity = TN / (TN + FP)	Proportion of true negative predictions out of all actual negative instances (avoiding misclassifying positives as negatives).

In the context of water quality prediction,

- ***TP (True Positive):*** Represent the number of correctly identified safe water samples
- ***TN (True Negative):*** Represent the number of correctly identified unsafe water samples
- ***FP (False Positive):*** Represent the number of falsely identified safe water samples
- ***FN (False Negative):*** Represent the number of falsely identified unsafe water samples

These terms are crucial for evaluating the performance of the prediction model and understanding its ability to accurately classify water samples as safe or unsafe [15].

5.4 DL classification result

This section presents a comparative analysis of various deep learning algorithms, evaluating their performance across precision, recall, F1-score, and accuracy metrics. The results underscore the superior effectiveness of the proposed ITWNN model, which outperforms other algorithms in all evaluated metrics. The table 3 and figure 8 displays precision metrics for various algorithms across five folds of cross-validation. It compares the precision of each algorithm for Class 0 and Class 1.

Table 3. *K-Fold Precision Comparison*

Metric	Algorithm	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean
Precision (Class 0)	LIME	0.88	0.87	0.89	0.88	0.87	0.88
	SHAP	0.87	0.88	0.88	0.89	0.86	0.88
	Anchor	0.90	0.89	0.91	0.90	0.90	0.90
	LORE	0.96	0.95	0.96	0.97	0.95	0.96
	ITWNN	0.96	0.96	0.97	0.96	0.96	0.96
Precision (Class 1)	LIME	0.17	0.16	0.18	0.15	0.17	0.17
	SHAP	0.00	0.00	0.00	0.00	0.00	0.00
	Anchor	0.86	0.85	0.87	0.86	0.86	0.86
	LORE	0.89	0.88	0.90	0.89	0.88	0.89
	ITWNN	0.97	0.96	0.98	0.97	0.96	0.97

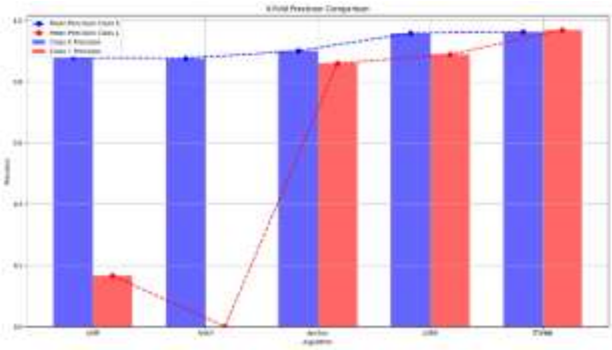

***Figure 8**. K-Fold Precision Comparison*

For Class 0, all algorithms show high precision, with ITWNN achieving the highest mean precision of 0.96. In contrast, for Class 1, ITWNN also leads with the highest mean precision of 0.97, while SHAP consistently performs poorly with a mean precision of 0.00. This comparison highlights ITWNN's superior performance in identifying both classes effectively. The table 4 and figure 9 show recall metrics for various algorithms across five folds of cross-validation. For Class 0, ITWNN demonstrates the highest performance with a perfect mean recall of 1.00, indicating it consistently identifies all relevant instances. Other algorithms, including LIME, SHAP, Anchor, and LORE, also perform well, with recalls close to 1.00. For Class 1, ITWNN achieves the highest recall of 0.70, while SHAP performs poorly with a mean recall of 0.00. This comparison underscores

***Table 4.** K-Fold Recall Comparison*

Metric	Algorithm	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean
Recall (Class 0)	LIME	0.95	0.94	0.95	0.96	0.95	0.95
	SHAP	1.00	0.99	1.00	0.99	1.00	0.99
	Anchor	0.99	0.98	0.99	0.99	0.98	0.99
	LORE	0.99	0.98	0.99	0.99	0.98	0.99
	ITWNN	1.00	1.00	1.00	1.00	1.00	1.00
Recall (Class 1)	LIME	0.07	0.06	0.08	0.07	0.08	0.07
	SHAP	0.00	0.00	0.00	0.00	0.00	0.00
	Anchor	0.26	0.25	0.27	0.26	0.27	0.26
	LORE	0.72	0.71	0.73	0.72	0.71	0.72
	ITWNN	0.70	0.69	0.71	0.70	0.71	0.70

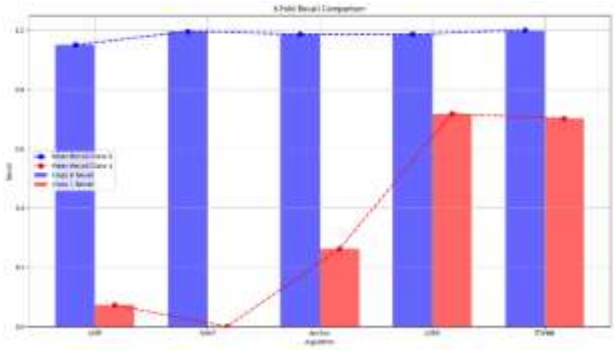

***Figure 9**. K-Fold Recall Comparison*

ITWNN's superior ability to recall both classes effectively. The table 5 presents the F1-score metrics for various algorithms across five folds of cross-validation (figure 10). For Class 0, ITWNN achieves the highest mean F1-score of 0.98, reflecting its superior balance between precision and recall. Other algorithms, such as

***Table 5**. K-Fold F1-Score Comparison*

Metric	Algorithm	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean
F1-Score (Class 0)	LIME	0.91	0.90	0.91	0.91	0.90	0.91
	SHAP	0.93	0.92	0.93	0.92	0.93	0.92
	Anchor	0.95	0.94	0.95	0.95	0.94	0.95
	LORE	0.97	0.96	0.97	0.97	0.96	0.97
	ITWNN	0.98	0.97	0.98	0.98	0.97	0.98
F1-Score (Class 1)	LIME	0.10	0.09	0.11	0.10	0.11	0.10
	SHAP	0.00	0.00	0.00	0.00	0.00	0.00
	Anchor	0.39	0.38	0.40	0.39	0.40	0.39
	LORE	0.80	0.79	0.81	0.80	0.79	0.80
	ITWNN	0.82	0.81	0.83	0.82	0.83	0.82

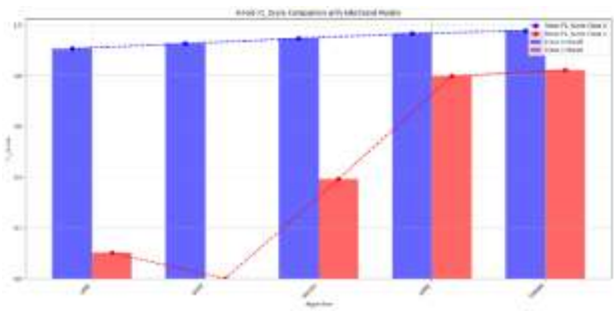

***Figure 10**. K-Fold F1-Score Comparison*

LIME, SHAP, Anchor, and LORE, also perform well, with F1-scores ranging from 0.91 to 0.97. For Class 1, ITWNN leads with a mean F1-score of 0.82, outperforming other methods (figure 11). SHAP shows minimal performance with an F1-score of 0.00. This comparison highlights ITWNN's overall effectiveness in classifying both classes.

Table 6. *Analysis of the performance of DL algorithms*

DL Algorithms	Accuracy	Precision	Recall	F1-Score
LIME	0.84	0.79	0.84	0.81
SHAP	0.88	0.77	0.88	0.82
Anchor	0.90	0.90	0.90	0.88
LORE	0.95	0.95	0.95	0.95
ITWNN (proposed)	0.96	0.96	0.96	0.96

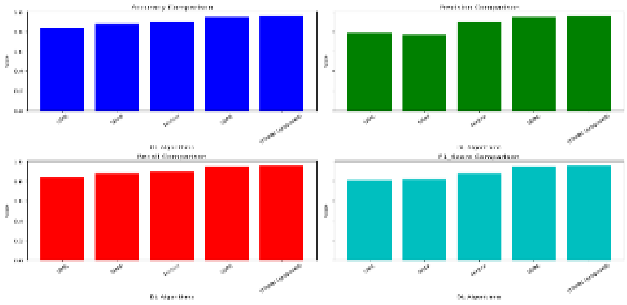

***Figure 11**. Comparative Performance of DL Algorithms*

The table 6 presents the weighted average metrics of accuracy, precision, recall, and F1-score for various algorithms, including the proposed Interpretive Time-Warping Neural Network (ITWNN). As shown in figure 12, ITWNN shows the highest performance across all metrics, with an accuracy of 0.96, precision of 0.96, recall of 0.96, and F1-score [15] of 0.96, indicating its superior classification capability. LORE also performs well with high values in accuracy and F1-score [16]. In contrast, SHAP and LIME have lower metrics, reflecting less effective classification in comparison to ITWNN and LORE. The confusion matrix for the ITWNN (proposed) algorithm shows that the model accurately predicted 1396 negative cases (TN) and 141 positive cases (TP) [17] [18] [19]. However, it misclassified 4 negative cases as positive (FP) and 59 positive cases as negative (FN). This indicates that while the model performs well in predicting negatives with minimal false

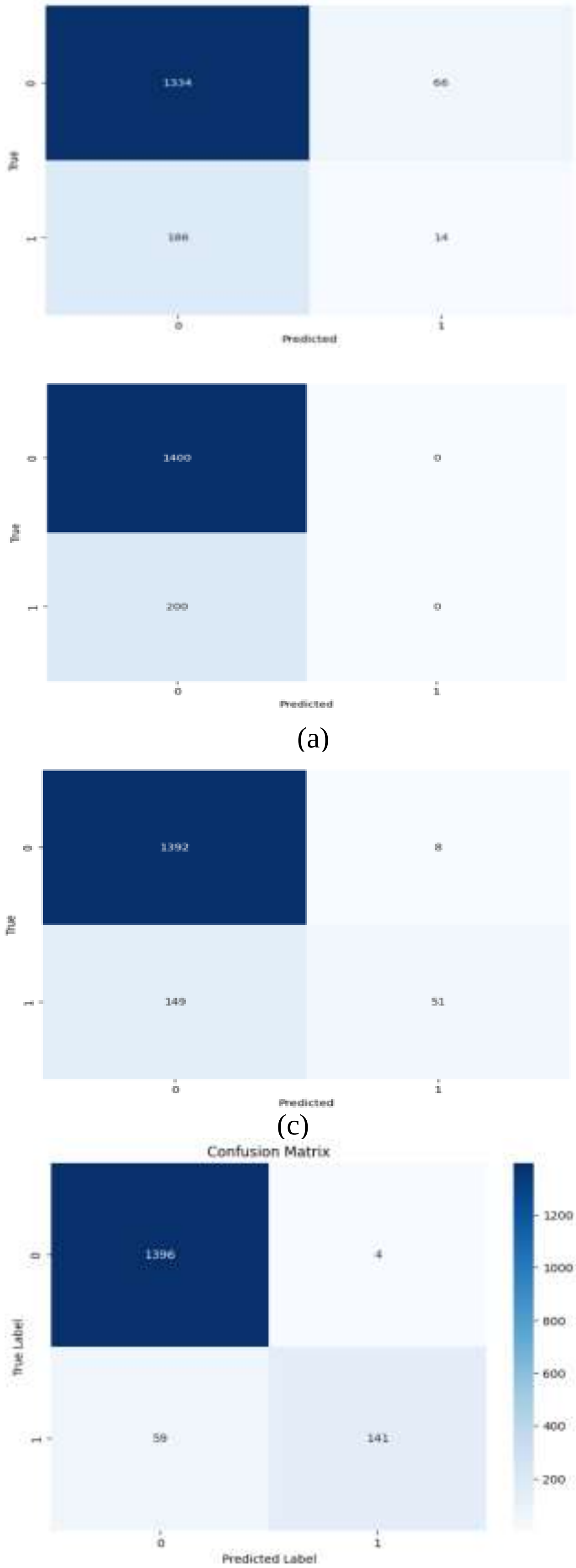

Figure 12. *Confusion matrix (a) LIME (b) SHAP (c) Anchor (d) LORE (e) ITWNN (proposed)*

positives, its performance in predicting positives could be improved, as evidenced by the higher number of false negatives [20]. Further performance metrics like precision, recall, and F1-score should be calculated to gain a comprehensive understanding of the model's effectiveness.

6. Conclusion

The research paper concludes that the integration of deep learning techniques, specifically LSTM models, with interpretability methods such as LIME, SHAP, Anchor, and LORE, significantly enhances the accuracy and transparency of water quality assessment. The proposed Interpretive Time-Warping Neural Network (ITWNN) achieved a notable accuracy of 96%, along with high precision, recall, and F1-score, outperforming other algorithms. This study highlights the effectiveness of combining advanced parameter optimization with robust deep learning models to achieve precise predictions and actionable insights. The findings underscore the potential of leveraging IoT sensor data and deep learning in environmental science, offering valuable tools for improved water quality management and setting a precedent for future research in the field. The research highlights the effectiveness of combining advanced parameter optimization with robust deep learning models to achieve precise predictions and actionable insights, while also paving the way for future enhancements such as integrating real-time data streams and expanding the model's applicability to diverse environmental contexts.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1]Essamlali, I., Nhaila, H., & El Khaili, M. (2024). Advances in machine learning and IoT for water quality monitoring: A comprehensive review. *Heliyon*, 10(6), e27920.

[2]Wang, Q., Lin, J., Guo, W., Liu, S., Zeng, X., & Xu, Y. (2020). A CNN-LSTM hybrid model for real-time water quality prediction based on limited historical data. *Environmental Science and Pollution Research*, 27(23), 32883-32893.

[3]Niu, J., Zhang, H., Wang, X., & Sheng, Y. (2021). Attention-based LSTM for water quality prediction using IoT sensors. *Sensors* (Switzerland), 21(12), 4254.

[4]Zhang, Y., Li, X., Wang, Z., & Peng, J. (2019). Deep learning for real-time water quality parameter prediction based on IoT devices. *Journal of Sensors*,

[5]Kaur, P., Saini, V., & Singh, H. (2022, September). Explainable LSTM with SHAP for water quality prediction using IoT sensors. *In 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMLBCP)* (pp. 1-6). IEEE.

[6]Al-Barakati, A., Al-Saggaf, A., Ali, M., & Shahabi, M. (2020). A Long Short-Term Memory (LSTM) based deep learning model for water quality prediction. *Sensors (Switzerland)*, 20(23), 6790.

[7]JongCheol Pyo, Yakov Pachepsky, Soobin Kim, Ather Abbas, Minjeong Kim, Yong Sung Kwon, Mayzonee Ligaray, Kyung Hwa Cho, (2023). Long short-term memory models of water quality in inland water environments, *Water Research X*, 21;100207. https://doi.org/10.1016/j.wroa.2023.100207

[8]Chen, H.; Yang, J.; Fu, X.; Zheng, Q.; Song, X.; Fu, Z.; Wang, J.; Liang, Y.; Yin, H.; Liu, Z.; et al. (2022). Water Quality Prediction Based on LSTM and Attention Mechanism: A Case Study of the Burnett River, Australia. *Sustainability*, 14, 13231. https://doi.org/10.3390/su142013231

[9]Khokhar, F.A., Shah, J.H., Saleem, R. et al. (2024). Harnessing deep learning for faster water quality assessment: identifying bacterial contaminants in real time. Vis Comput (2024). DOI:10.1007/s00371-024-03382-7

[10]Wang, H.; Xiao, N. (2023). Underwater Object Detection Method Based on Improved Faster RCNN. *Appl. Sci.*, 13(4), 2746; https://doi.org/10.3390/app13042746

[11]Zhou, J.; Wang, J.; Chen, Y.; Li, X.; Xie, Y. (2021). Water Quality Prediction Method Based on Multi-Source Transfer Learning for Water Environmental IoT System. *Sensors* 21, 7271.

[12]N. Radhakrishnan and A. S. Pillai, (2020). Comparison of Water Quality Classification Models using Machine Learning. *2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India*, pp. 1183-1188.

[13]N. Abulail, A. Y. Owda and M. Owda, (2023). Water Quality Classification Decision Support System. *2023 International Conference on Information Technology (ICIT), Amman, Jordan*, pp. 73-78.

[14]Shehab, S.A., Darwish, A., Hassanien, A.E. et al. (2023) Water quality classification model with small features and class imbalance based on fuzzy rough sets. *Environ Dev Sustain*. DOI:10.1007/s10668-023-03916-4

[15]Kaur, A., Khurana, M., Kaur, P., Kaur, M. (2021). Classification and Analysis of Water Quality Using Machine Learning Algorithms. In: Sabut, S.K., Ray, A.K., Pati, B., Acharya, U.R. (eds) *Proceedings of International Conference on Communication, Circuits, and Systems. Lecture Notes in Electrical Engineering*, vol 728. Springer, Singapore.

[16]Maheshwari, R.U., Kumarganesh, S., K V M, S. et al. (2024). Advanced Plasmonic Resonance-enhanced Biosensor for Comprehensive Real-time Detection and Analysis of Deepfake Content. *Plasmonics*. https://doi.org/10.1007/s11468-024-02407-0

[17]Maheshwari, R. U., Paulchamy, B., Arun, M., Selvaraj, V., & Saranya, N. N. (2024). Deepfake Detection using Integrate-backward-integrate Logic Optimization Algorithm with CNN. *International Journal of Electrical and Electronics Research*, 12(2), 696-710.

[18]Maheshwari, R. U., & Paulchamy, B. (2024). Securing online integrity: a hybrid approach to deepfake detection and removal using Explainable AI and Adversarial Robustness Training. *Automatika*, 65(4), 1517-1532. DOI:10.1080/00051144.2024.2400640

[19]Sood, K., Dhanaraj, R. K., Balusamy, B., Grima, S., & Uma Maheshwari, R. (Eds.). (2022). Big Data: A game changer for insurance industry. *Emerald Publishing Limited.*

[20]Janarthanan, R., Maheshwari, R. U., Shukla, P. K., Shukla, P. K., Mirjalili, S., & Kumar, M. (2021). Intelligent detection of the PV faults based on artificial neural network and type 2 fuzzy systems. *Energies*, 14(20), 6584. DOI:10.3390/en14206584

[21]Appalaraju, M., Sivaraman, A. K., Vincent, R., Ilakiyaselvan, N., Rajesh, M., & Maheshwari, U. (2021). Machine learning-based categorization of brain tumor using image processing. *In Artificial Intelligence and Technologies: Select Proceedings of ICRTAC-AIT 2020 (pp. 233-242).* Singapore: Springer Singapore.

[22]Sasikala, S., Sasipriya, S., & Maheshwari, U. (2022, March). Soft Computing based Brain Tumor Categorization with Machine Learning Techniques. *In 2022 International Conference on Advanced Computing Technologies and Applications (ICACTA)* (pp. 1-9). IEEE.

[23]Rajendran, U. M., & Paulchamy, J. (2021). Analysis and classification of gait characteristics. *Iconic Research and Engineering Journals*, 4(12).

[24]Maheshwari, R. U., Paulchamy, B., Pandey, B. K., & Pandey, D. (2024). Enhancing Sensing and Imaging Capabilities Through Surface Plasmon Resonance for Deepfake Image Detection. *Plasmonics*, 1-20. DOI:10.1007/s11468-024-02492-1

[25]Maheshwari, R. U., Jayasutha, D., Senthilraja, R., & Thanappan, S. (2024). Development of Digital Twin Technology in Hydraulics Based on Simulating and Enhancing System Performance. *Journal of Cybersecurity & Information Management*, 13(2);50-65 DOI:10.54216/JCIM.130204

[26]ÇELİK, M. E. (2023). A Novel Deep Learning Model for Pain Intensity Evaluation. International Journal of Computational and Experimental Science and Engineering, 9(4), 325–330. Retrieved from https://ijcesen.com/index.php/ijcesen/article/view/274

[27]AYKAT, Şükrü, & SENAN, S. (2023). Using Machine Learning to Detect Different Eye Diseases from OCT Images. International Journal of Computational and Experimental Science and Engineering, 9(2), 62–67. Retrieved from https://ijcesen.com/index.php/ijcesen/article/view/191

[28]KIRELLİ, Y., & AYDIN, G. (2023). Classification of Histopathological Images in Automatic Detection of Breast Cancer with Deep Learning Approach. International Journal of Computational and Experimental Science and Engineering, 9(4), 359–367. Retrieved from https://ijcesen.com/index.php/ijcesen/article/view/279

[29]Priti Parag Gaikwad, & Mithra Venkatesan. (2024). KWHO-CNN: A Hybrid Metaheuristic Algorithm Based Optimzed Attention-Driven CNN for Automatic Clinical Depression Recognition . International Journal of Computational and Experimental Science and Engineering, 10(3);491-506. https://doi.org/10.22399/ijcesen.359

[30]Jha, K., Sumit Srivastava, & Aruna Jain. (2024). A Novel Texture based Approach for Facial Liveness Detection and Authentication using Deep Learning Classifier. International Journal of Computational and Experimental Science and Engineering, 10(3);323-331. https://doi.org/10.22399/ijcesen.369

[31]Varone G, Ieracitano C, Çiftçioğlu AÖ, Hussain T, Gogate M, Dashtipour K, Al-Tamimi BN, Almoamari H, Akkurt I, Hussain A. A Novel Hierarchical Extreme Machine-Learning-Based Approach for Linear Attenuation Coefficient Forecasting. Entropy. 2023; 25(2):253. https://doi.org/10.3390/e25020253