RESEARCH ARTICLE

Secure Aware Outlier Detection in Underwater Wireless Sensor Networks using Deep Learning

Sunbal Rani[1,*] **and Saqib Shahid Rahim**[1]

[1] Department of Computing, Abasyn University, Peshawar, Pakistan

Abstract

The emerging field of underwater sensor networks (UWSN) has a vital potential shaping the modern and future landscape that assists in measuring water quality, pollution tracking, and identification of underwater habitats. The challenging conditions in the UWSN environments raise data and security concerns in terms of outliers related to the complicated communication system, poor visibility, and limited resources. The data quality and network efficiency may be affected due to these unwanted conditions, giving rise to certain malicious activities in the network. This study aims to enhance the outlier identification process in terms of security and quality perspectives using the Long Short-Term Memory (LSTM) framework. The focus is to identify the temporal patterns and differentiate between various outliers in critical UWSN conditions. Results reveal that the proposed framework achieved high accuracy up to 95% and surpassed the other traditional machine learning models. It is worth mentioning that underwater sensor data have a complicated pattern that can be more appropriately handled using deep learning models, including LSTM, in comparison to traditional machine learning models.

Keywords: underwater sensor networks (UWSN), anomaly detection (AD), outlier detection (OD), deep learning (DL), long short-term memory (LSTM).

1 Introduction

Marine environments and disaster detection are designed using underwater acoustic sensor networks. Wireless Sensor Networks (WSNs) are also applied in forestry to identify fires, track forest growth, and evaluate ecological conditions. Wireless acoustic pipeline systems that monitor air quality help authorities detect significant changes in pollution, further demonstrating the adaptability of Wireless Sensor Networks (WSNs) [1, 2].

The transfer of information by wireless means over the oceans is vital for systems that monitor oceanic states and support sensor networks [3]. Underwater Wireless Sensor Networks (UWSN) are employed in collecting the information and transmitting it underwater. Some of the tracking applications of UWSNs are monitoring the oil industry, controlling fish farms, pollution control, weather observing, natural disasters prediction, search and rescue missions, and impact study of marine life [4].

UWSNs are the state-of-the-art technology that is

Submitted: 19 June 2025
Accepted: 16 July 2025
Published: 18 August 2025

Vol. 1, **No.** 1, 2025.
10.62762/TISC.2025.610386

***Corresponding author:**
Sunbal Rani
sunbal.rani@abasyn.edu.pk

Citation
Rani, S., & Rahim, S. S. (2025). Secure Aware Outlier Detection in Underwater Wireless Sensor Networks using Deep Learning. *ICCK Transactions on Information Security and Cryptography*, 1(1), 25–37.

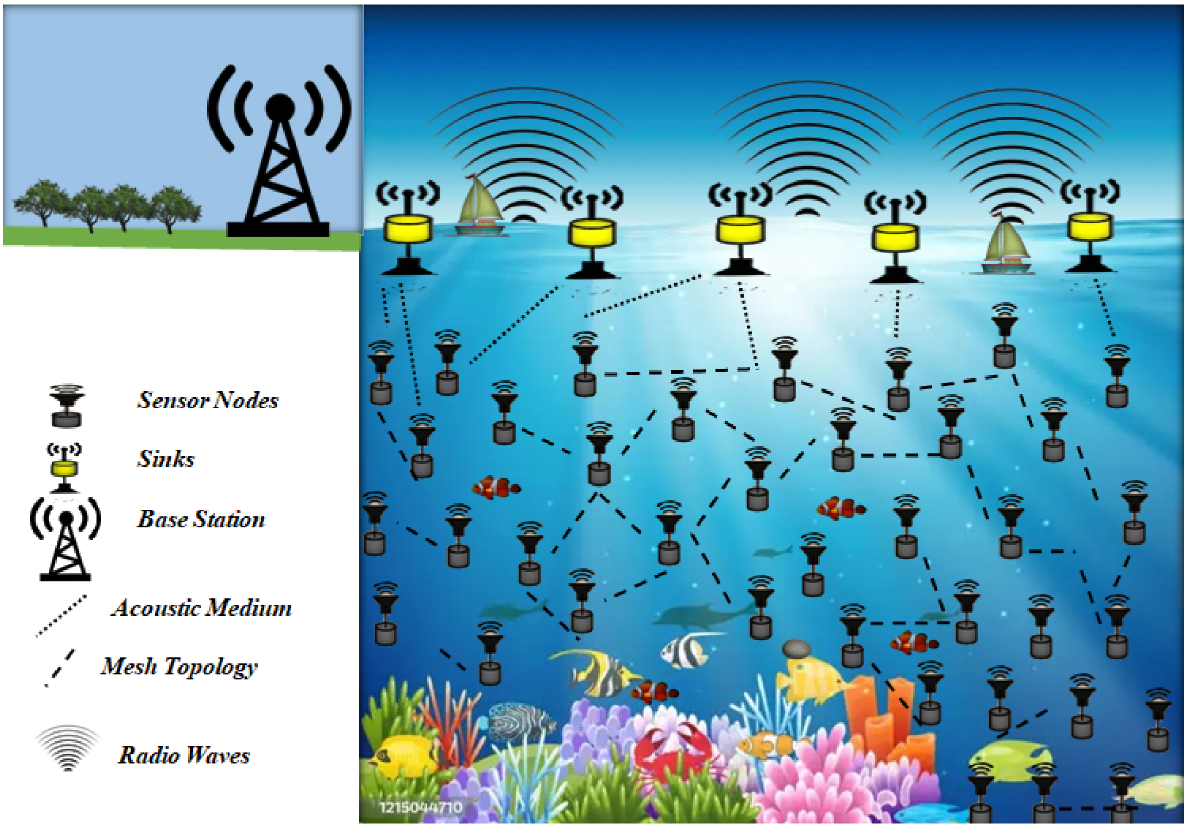

Figure 1. Underwater sensor network.

now changing the world of marine exploration and environmental monitoring. Through the positioning of nodes, both underwater and on the surface, these networks have a secure infrastructure to help in the precise monitoring of some areas of the ocean bed, as shown in Figure 1 [5]. UWSNs facilitate the seamless transfer of data among nodes and base stations, continuously modifying communication channels to guarantee optimal communication channels, given the changing environmental conditions. Equipped with smart sensors and autonomous vehicles, these networks operate on wireless networks, coordinating various activities to ensure continuous transfer of information on scientific research and environmental management without data loss. In a complicated framework of UWSNs, the communication systems are running on diverse carriers, such as acoustic, electromagnetic, and optical waves [1]. Among them, acoustic communication can be termed as the best strategy of underwater communication. Sound wave transmission through water enables transferring information efficiently and, therefore, real-time monitoring and responsive environment changes. Based on the specifics of acoustic waves, UWSNs have a high reach and reliability, which stimulates a breakthrough in oceanic research and resource management [6]. UWSNs can be used beyond data transportation; they can also be essential in the monitoring and analysis of the environment and predictive tracking. Such networks are critical for predicting weather patterns, predicting natural disasters, as well as tracking vital environmental conditions, such as pressure, temperature, and pollution [7].

Acoustic underwater communication is a means of communication that uses sound waves in the transmission of information underwater, and is used in tasks including navigation, sensing [8]. It works great even on long ranges as well as through other strategic obstacles, due to which it not only works well on clear water but murky ones as well. The method is two-way and provides flexibility in the application underwater. However, it is limited in some way [9]. Background noise may restrict the clarity, and sound travels less as it moves towards a long distance, thus slowing its impact. Moreover, the sounds employed in communication may be able to upset marine animals and therefore change their way of operation, thereby affecting their natural habits. Despite these challenges, underwater acoustic communication can be a reliable technology in a broad spectrum of underwater applications because of its versatility and range [10–12].

UWSNs play an essential role in surveillance and the interrogation of different marine environment properties. These networks adopt different architectures to fulfill specific operational needs and address unique challenges.

Sensor nodes in one-dimensional UWSNs are defined in a line, either in a row or at particular depths. The nodes act individually and translate into a small network by themselves to gather data within the immediate environment. During the exploration, the nodes will be fixed to make sure that the received information is accurate [13]. Upon collecting the data, nodes emerge to relay the information to distant stations. The acoustic signals convey communication among the nodes during the submersion, thus successfully allowing the nodes to coordinate and exchange data effectively. This system favors surveillance in a fixed direction and relaying information acquired to the surface to be analyzed further [14]. In a two-dimensional underwater wireless sensor network (UWSN), sensor nodes are grouped and deployed underwater, with each cluster headed by a leader node, commonly known as an anchor node. These nodes do not move; they are attached to the surface of the ocean. All the sensor nodes in a cluster gather the data and send it to the anchor node in the cluster [15]. The linked nodes would then collect data from all the connected nodes and feed it into surface buoy hubs. In this architecture, communication between the two dimensions is possible. Additionally, horizontal communication between sensor nodes and their anchor nodes, as well as vertical communication between anchor nodes and surface buoy hubs, can be facilitated. Communications between the anchor nodes and the surface buoy hubs are also carried out through acoustic communication determination because the distance between them is usually significant.

The use of an underwater wireless sensor network can be made three-dimensional, where the sensor nodes are set at various depth levels to sense some event in the water. One of the common ways of interfacing the underwater sensor nodes and the surface buoys is by connecting using adjustable cables [15]. Each sensor is secured to the seabed and connected to a floating buoy that can be inflated through a pump to regulate its position as needed. These sensors should be placed below the surface of water and will require a reputable connection method to make them work reliably. It is possible to collect data from sensors through two primary methods. The former approach involves connecting all underwater sensors to the floating buoy via variable wires, allowing for adaptive depth changes [12]. The architecture of UWSN is four-dimensional in 3D UWSN architecture with the combination of mobile underwater sensor networks [14]. The mobile UWSNs exhibit remotely operated vehicles that gather information from the anchor nodes. This is made available in far stations where it is analyzed. Depending on the proximity of the node to ROVs, information is transferred to the ROVs through each sensor node that can transfer information directly to ROVs. Depending on the nature of the data to be transmitted between the underwater sensor nodes and the ROV and the distance of the nodes to the ROV, the communication is affected [16].

In short, the study helps in building an LSTM-based model for anomaly detection to enhance data accuracy and reliability in UWSNs. It addresses the overall issues of noise, signal disturbance, and environmental variance to improve the detection of outliers. The study covers data preprocessing, model optimization, and performance assessment, ensuring the proposed method is practical and scalable for real-time underwater use.

The structure of the research paper is as follows: Section 2 gives a literature review. Section 3 contains the proposed methodology. Section 4 will discuss the experiment and analysis of the findings. Future directions end the study in section 5.

2 Related Work

WSNs have attracted much attention in recent years for solving various problems, including energy efficiency, resource and power management, and traffic control [2, 17, 18]. A new field in research has been introduced, the prediction of behavior of underwater sensor nodes using DL, hence giving great understanding into the underwater environment [3]. Several methods have been suggested over the years to apply to UWSNs in the direction of identifying suspicious happenings on the bottom of the ocean using diverse algorithms. The literature review provides an overview of various approaches that incorporate varied methods and techniques in the context of UWSNs.

AD in UWSNs adopts identifying sensor anomalies in the water environment. With the help of modern technologies such as machine learning and deep learning, it will be possible to identify problems generated by environmental changes or sensor errors, and provide proper monitoring and valid results. In [16], the authors presented an approach in which an unsupervised AD system on an LSTM model was considered to interpret the interplay of nodes to identify anomaly detection. The study highlights the challenges related to anomaly detection algorithms in terrestrial sensor networks, as opposed to underwater networks, where anomalous behaviors differ. The suggested model has both a prediction stage and a detection stage, which are based on an LSTM network to detect anomalies. Results show an accuracy of 80% and 84% in the first and second experiments, respectively. This method will have good levels of accuracy as well as minimize the false alarms in UWSNs with minimum computational complexity.

In [19], the authors have developed a new outline to identify anomalies in the underwater acoustic network. In this research, the authors implement a self-regressing prediction-based system in which each sensor node finds anomalous data by itself. The fuzzy inference system is then used in detecting the sensor that is malfunctioning. The approach achieved 90% accuracy in the real underwater acoustic setting and minimal false alarms in the underwater acoustics environment. Results show that the technique is efficient in identifying problems and does not bring numerous false alarms. The method is very efficient in identifying anomalies without delivering countless false positives.

In [20], the authors developed an algorithm for anomaly detection in water supply systems using sensors and machine learning algorithms. This technique employs KNN and NB models to analyze data collected by the sensors. The method proved effective, with a precision rate in anomaly detection ranging from 83% to 87%.

The authors proposed a novel approach for detecting anomalies in UWSNs based on temporal and spatial correlation [21]. Their algorithm uses GRU to comprehend the time series feature. It examines the distribution of variances between the observed and actual values using a sliding window. Thus, the abnormal data may be identified at an early stage by the use of probability density. Moreover, the algorithm handles the spatial correlation through data features incorporation and the use of Euclidean distance to make further evaluation of abnormal data. The performance of this method in detecting anomalies in data from underwater sensors is explained by experimental results.

In [22], the authors proposed a new method considering the outliers and optimum positioning of the anchor nodes to bring about a better path optimization in an underwater sensor network. The drawback in this research paper is the implementation of sensors across the dimensions. The intelligibility approach gives significance to the location of anchor nodes in ensuring that accurate localization is realized. It demonstrates that outlier data can be addressed and that the shortest path can be efficiently placed using half-quadratic minimization. The experiments show that optimizing the positioning of the anchors based on the Fisher Information Matrices, but respecting D-optimality requirements. Quantitative evidence shows that the method has major strengths compared to the current literature, especially when it comes to processing outliers. Given the study's focus on 3D localization to optimize partially connected underwater optical wireless sensor networks and its resistance to outliers, it holds high prospects for enhancing the accuracy of UWSN.

3 Methodology

This methodology section provides a detailed and systematic methodology for detecting outliers in UWSNs using an LSTM deep learning model, with the aid of comparisons with traditional machine learning methods. The methodology starts with the thorough analysis of anomalies and outliers, which will be characterized, and the significance of their role in the study of sensor data will be explained. The proposed methodology aims to enhance better prediction and trustworthy outlier detection ability in complex environments under water.

With a thorough and dynamic environment of UWSNs, precise anomaly detection plays a fundamental role in the successful gathering of consistent information as well as its transmission. This methodology includes several succeeding phases, such as the selection of the relevant data in the dataset, preprocessing, and feature extraction, as a vital process in achieving the optimal performance of the model. It will be followed by the method of model selection and training, during which a variety of configurations and hyperparameters will be tested to determine the most effective model. Lastly, it has thorough assessment and testing processes to guarantee the reliability and robustness of the outlier detection system. In this systematic process, this study will contribute significantly to the research of UWSN by improving the accuracy and reliability of the outlier detection (OD) mechanism. Figure 2 shows a detailed structure of the LSTM model.

3.1 Selection of Anomaly-Based Dataset

The dataset covers a broad scope of sensor measurements vital in identifying outliers in UWSNs. The data include variables like water temperature, depth, pressure, salinity, pH, dissolved oxygen, turbidity, signal strength, timestamp, and disturbance level, among others, the variety and dynamism of variables that underwater sensors are very likely to encounter. This dataset has been selected because it represents the complete set of variables and conditions typically observed in UWSNs. As such, it is a unique resource to develop and test strong models of anomaly detection.

3.2 Pre Processing

Preprocessing data is the initial phase to analyze the data. In this phase, the data must be clean, reliable, and consistent. The following are some of the critical stages in this phase: First, there is the data cleaning, where entries that are incomplete or do not match are removed to ensure the integrity of the dataset. Afterwards, normalization scales with digits to a fixed scale usually between 0 and 1, so that the training of models becomes more efficient and robust. Organization of temporal sequences is undertaken through formatting the data based on the feature of timestamp, which facilitates harvesting and making use of the temporal dependencies that are essential in the dataset. Lastly, the process of outlier labeling is performed, in which the data points are classified as usual or outliers and are predetermined either based on domain knowledge or using any criterion. This will make the models learn to identify the pertinent instances of anomalies. All these preprocessing processes increase the quality and relevancy of this dataset, which paves the way for more accurate and reliable outlier identification.

3.3 Features Extraction

Outlier analysis is a significant process in feature extraction in the UWSN environment. This comes in the form of choosing and extracting features of the data to complement the process of detecting anomalies. The most important features that can be adequately extracted are temporal features, they include the timing of patterns and trends, which are used to determine likely time dependence of the LSTM model, statistical features, which include the mean and standard deviation which can be used to provide a sense of distribution and variability in the data, and domain-specific features, which are based on expert knowledge use [23]. The dataset incorporates vital parameters like temperature in water, depth, pressure, salinity, pH, dissolved oxygen, turbidity, signal strength, time, and level of disturbance, all of which will help in the comprehensive outlook of the UWSN environment and, above all, help in the recognition of outliers effectively and adequately.

3.4 Model Selection and Architecture

The LSTM model is chosen due to its ability to process sequential data, which is essential for identifying outliers in the UWSN environment. Important hyperparameters in selecting a model and forming the architecture are setting up LSTM layers, i.e., trying different settings of the number of hidden layers (1, 2, 3), and units per layer (64, 128, 256). The different activation functions (Sigmoid, ReLU, among others) are tested to determine how they influence the modeling of nonlinear relationships in the data. Hyperparameter tuning is the process of optimizing model performance by altering parameters, also known as hyperparameters, e.g., learning rate, batch size, and dropout rate, with the help of algorithms such as grid search or random search. To determine the relative performance of the model, it is compared in its operational efficiency with some customary machine learning models, such as Random Forest (RF), Decision Tree (DT) and K-Nearest Neighbor (KNN), Logistic Regression (LR), Support Vector Machine (SVM), and Naive Bayes (NB). This is an overall fine-tuning of the chosen LSTM model configuration that is very suited and will be used to detect anomalies in complex underwater environments correctly.

3.5 Model Training

To enhance outlier identification in UWSNs, the LSTM model training steps are involved. First, the data is split into 80/20 datasets to train and test, respectively, covering a healthy check of the quality of the model produced. The LSTM model is optimized during training to reduce the loss function. This optimization is based on the choice of web page parameters, e.g., the number of hidden layers, the number of units per layer, the type of activation functions used, and other training features, such as the number of epochs and the validation split. The model performance is monitored continuously on a validation set to guard against overfitting. They can include early stopping and regularization techniques to stop training when the performance becomes flat, to allow the model to learn to generalize to new and unseen data. Such a systematic approach is very essential in the development of a reliable LSTM model that could be

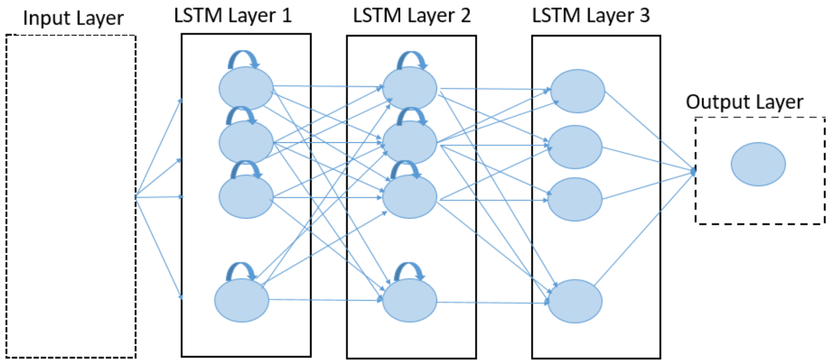

Figure 2. LSTM structure.

used to forecast anomalies in complex underwater environments.

3.6 Anomaly Detection

Reconstruction within an LSTM-based anomaly detection model is a capability of the model to reconstruct the input it received during its training due to the temporal dependencies and patterns that the model learnt. The LSTM layers process the input sequence and encode the data into a latent value. The encoded data is then decoded to get the original sequence, giving an output close to the information present. The error made in reconstructing the output is defined as the difference between the input and the reconstructed output and is usually measured using such metrics as Mean Squared Error. Data that lies on normal points and conforms to learned patterns creates low reconstruction errors, and data that does not deviate significantly creates high reconstruction errors. These errors are compared to an established threshold to mark data points as being normal or anomalous. When trained, the LSTM model is used to detect anomalies in the test dataset. The first step in the process of mode selection involves determining the reconstruction error at each data point, which serves as a measure of how well the model can reproduce the inputs. Having a high reconstruction error means there is a high chance that the given data point is not following what is expected and may be an aberration. To discriminate between the normal data and outliers, the threshold is established regarding the distribution of reconstruction errors. Data points that have error values that are more than this limit are considered anomalies, and those points that have error values less than this value are considered normal. Lastly, several measures are used to assess the usefulness of the anomaly detection process, such as accuracy, precision, recall, and standard deviation.

3.7 Comparison with Traditional Models

The LSTM model performance is contrasted to some other conventional machine learning models, such as RF, DT, KNN, LR, SVM, and NB, based on a similar set of evaluation measures. The difference in this comparison is that the LSTM shows better performance in terms of remembering the previous data temporal dependencies, which is the key to capturing the complex patterns and anomalies over time. As compared to traditional models, which might find it challenging to deal with the sequential nature of time-series data, the architecture of the LSTM is specifically formulated to work with the temporal dependency properties of time-series data and therefore performs better in the detection of outliers.

3.8 Cross Validation

A 2-fold cross-validation is utilized to evaluate the performance and the LSTM model's generalizability. In that method, the dataset is divided into two subsets: one subset trains the model, whereas the other one checks the validation. The experiment is then reset with the roles of the two subsets being interchanged, enabling the model to be tested against other data sets. This provides valuable insights into how the model handles unobserved data. Although 2-fold cross-validation is a convenient and helpful way of determining whether the model is reliable, it might fail to absorb all the variability in performance that the higher k-fold cross-validation would provide.

3.9 Proposed Solution

Outlier identification is essential in UWSNs. It is based on an LSTM model because it is used to decide sequential data and incorporates the temporal nature of data, as shown in Figure 3. This is done by selecting an appropriate dataset, proceeding through data preprocessing, which involves removing outliers,

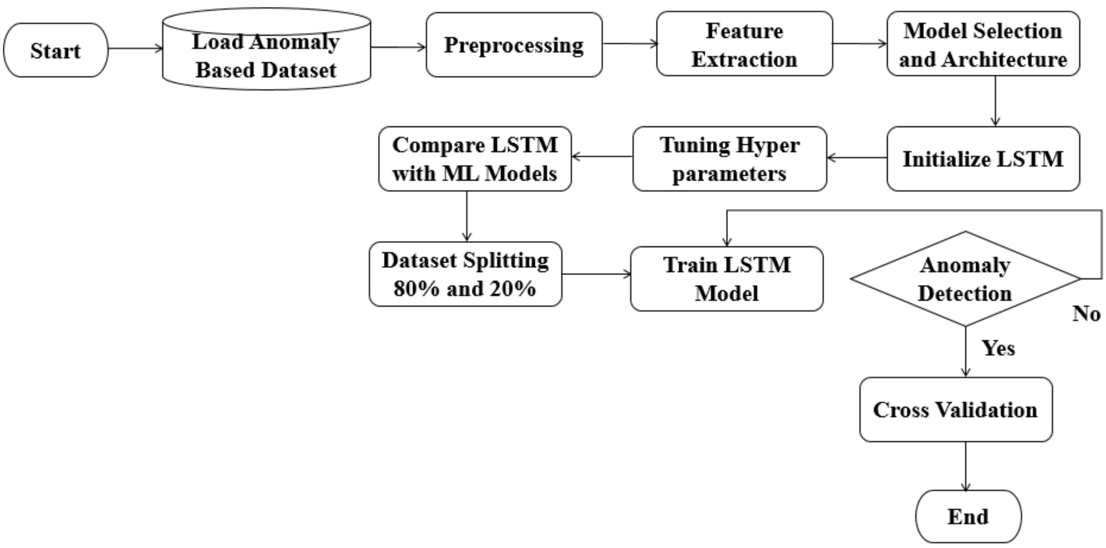

Figure 3. Proposed solution.

normalization, and labeling. The input is then enriched in terms of highlighting both temporal patterns and domain-specific insights using feature extraction. In the model selection process, several LSTM models are tested and compared to classic machine learning models. The data is split into an 80% training set and a 20% testing set. The reconstruction errors are used to detect anomalies, and the model performance is ensured against 2-fold cross-validation. This is to increase the level of accuracy of outlier detection and reliability within the UWSNs.

4 Experiments

In this section, experiments are performed to detect outliers in underwater sensor data. The dataset is well-explained at the beginning, involving the most critical environmental measurements, such as water temperature, depth, salinity, and pressure, and including an anomaly label for outliers. It also highlights the key tools and libraries needed to manipulate data and create models, such as Pandas, NumPy, TensorFlow, and Keras.

4.1 Network Simulation Setup

The simulated network is in a 100 m cubic space, and it comprises 250 nodes (sensors) and five nodes (sinks) in a mesh topology. The simulation will take 30 minutes, and it will be conducted based on the MATLAB R2015a system. The acoustic medium in which one operates, i.e., the underwater environment, presents specific considerations as acoustic waves present a set of distinct behavioral attributes not found in electromagnetic waves. Use the propagation delay with the speed of sound (1500 m/s). The energy consumption model introduced the loss of the incorporated signal path. Expanded acoustic range to represent normal ranges in an acoustic medium. Illustrated outlier nodes and energy used in an acoustic environment. To simulate, the data set will consist of several sensor measurements (water temperature, depth, pressure, salinity, pH, dissolved oxygen, turbidity, signal strength, timestamp, and level of disturbance). Anaconda is Python-based and integrated with MATLAB to collect data, preprocess data, and process the data further. The sensor information is shared along the mesh network and processed with the help of a model, which an LSTM represents. The main goal of this simulation is to detect outliers and hence make the outlier detection system of the sensor networks under the water much more accurate. The simulation settings, such as node details, acoustic medium, and dataset properties, have been summarized in Table 1 and can be referred to as the simulation setting descriptions.

Table 1. Parameters in simulation setup.

S.No	Parameters	Values
1.	Area Size	100m
2.	Number of Nodes	250
3.	Topology	Mesh
4.	Duration	30 minute
5.	Communication Medium	Acoustic
6.	Simulator	MATLAB R2015a
7.	Programming Language	Python
8.	Tool for Programming	Anaconda
9.	Frequency	25KHz
10.	Energy	100 Joules

4.2 Description of the Dataset

The data consists of environmental readings taken at different moments, which allows for the identification of anomalies. Essential properties are the time stamp of the measured data (time of measurement), depth in meters, temperature of the water in Celsius, salinity in parts per thousand, pressure in atmospheres, and strength of signal in dBs. There is also a battery voltage measured in volts, pH that measures the acidity or alkalinity of water, the concentration of dissolved oxygen is measured in milligrams per liter, and finally, water turbidity is measured in Nephelometric Turbidity Units. It has other features like: CHL-a concentration in micrograms/liters, NO3 concentration in milligrams/liters, and current velocity in meters/seconds. Any data entry is also marked as having an anomaly, and the values of 1 and 0 are assignments to indicate anomaly or normal data, respectively. The combination of these characteristics provides a detailed picture of the environmental situation in the water, enabling easy identification of anomalies.

4.3 Description of Features

Its dataset comprises principal characteristics that measure the most relevant environmental and sensor performance parameters of UWSNs, which are very insightful about different dimensions of the aquatic environment, as can be seen in Table 2. Such characteristics include the temperature of the water, the depth, the pressure, the salinity, the pH, the dissolved oxygen, the turbidity, and the strength of the signal. Moreover, timestamps are also present so that the time of data collection is accurate and the intensity of disturbance is measured to evaluate the activity of the environment. Every aspect also helps to reveal the thermal conditions and chemical composition, as well as the biological health of the underwater ecosystem. The data can also help determine the quality of data transmission and verify the data distributed by the sensors, which altogether make up a complete picture of the situation underwater, to monitor the environment effectively.

Table 2. Description of features.

S.No	Features	Descriptions
1.	Temperature	Measures the water temperature.
2.	Time Stamp	Records the time when sensor data is collected
3.	Depth	Measures the depth or distance from the water surface.
4.	Pressure	Measures water pressure at a specific depth.
5.	Salinity	Measures the concentration of salt in water.
6.	Signal Strength	Measures the strength of underwater communication signals.
7.	Potential of Hydrogen	It checks the acidity and alkalinity of the water.
8.	Dissolved Oxygen	It measures the amount of oxygen dissolved in the water.
9.	Turbidity	Measures the cloudiness or haziness of water
10.	Level of Disturbance	Indicates the degree of disturbance or activity in the environment.

4.4 Parameters for Model Implementation

The outlined table of parameters to follow. The process of implementing the LSTM model to improve outlier detection within UWSNs explains what has to be used to develop and train the model. Such parameters are the number of hidden layers and the number of units on each layer that define the depth and ability of the network to represent complex structures. Activation functions used by the hidden layer and the output layer influence the learning of nonlinear relationships and the estimation of precise class probabilities by a model. Basic parameters like optimizer, loss functions, and evaluation parameters play a very crucial role in guiding the process of training the model to achieve its best performance and to test the accuracy, precision, and recall of the model. Other parameters, such as the epoch number and the split of the validation, accomplish practical training and confident validation of the model that can generalize to previously unseen data. The set of extensive configurations of parameters is necessary to build a substantial LSTM model aimed at finding anomalies in a complex underwater environment.

4.5 Parameters of the Model with Values

The identified parameter list enables us to organize the configuration in a simple list, displaying the values used to compare the model performance of various parameters in the ML models. The DL, LSTM model hyperparameters include the number of hidden network layers (1, 3 in the search space), units per layer (64, 128, or 256), and the choice of either ReLU or sigmoid activation functions, which provide flexibility in architecture design. The training parameters set shall have 10 epochs and a 0.2 validation split, and this practice would result in good training and evaluation. The two folds are used to cross-verify numerous machine learning classifiers, such as RF, DT, KNN, LR, SVM, and NB. Such precise specification can help effectively compare the performance of the model and its ability to generalize, and with the help of this comparison, define the best possible option for detecting the outliers in UWSNs. As can be seen in Table 3 below, the parameters and their corresponding values are given in detail.

Table 3. Parameters of the model with values.

S.No	Parameter	Values
1.	Number of Folds for Cross-Validation	2
2.	LSTM Architecture Hyperparameters	Hidden Layers: [1, 2, 3] Units per Layer: [64, 128, 256] Activation Function: ['ReLU', 'sigmoid']
3.	Training Parameters	Epochs: 10 Validation Split: 0.2

4.6 Evaluation Metrics

Performance of DL and LSTM models is measured with several key metrics, which give a complete picture of their performance. Such metrics include accuracy, which measures the general correctness of a model's predictions, precision, which measures the fraction of correct positives among identified outliers, and recall, which measures the model's ability to capture all actual outliers. Collectively, these metrics provide valuable insights regarding the precision, dependability, and efficiency with which anomalous activities are detected in the underwater environment under the complex dynamics of the environment.

4.7 Standard Deviation

The standard deviation (σ) measures the variation or spread of values for metrics such as accuracy, precision, and recall. It is calculated by using the formula in Equation 1:

$$\sigma = \frac{\sqrt{\sum (x_i - \mu)^2}}{N} \tag{1}$$

where x_i represents each value of the metric, μ is the mean value of the metric, and N is the total number of observations, which is 9 in this observation. This formula helps quantify how much the metric values deviate from the mean, providing insight into the consistency of the model's performance across the observations.

4.8 Model Architecture With Sigmoid

LSTM uses a sigmoid activation function, with one, two, or three hidden layers, each containing one or more units corresponding to 64, 128, or 256 units. The performance measurement based on these variations of architecture is explored, and how this affects performance in terms of detecting outliers in UWSNs. Importantly, the smoothness of the gradient is the most characteristic feature of the sigmoid activation function, and this attribute of the activation function also makes it most applicable to binary classification problems and solving non-linear data interconnections. This study aims to achieve the balance between model complexity and model performance by systematically modifying the number of hidden layers and the units in each hidden layer, in single-layer, two-layer, and three-layer models. The research focuses on the influences of architectural complexity on the model resiliency and performance in distinguishing outliers in complex underwater settings. Finally, it provides some information about the impact of the depth of the layers and unit structures on the generalization power of a model and its successful functioning under the conditions of the real world.

4.9 Evaluation Metrics for Sigmoid Activation

The overall assessment of the model settings shows different performance indicators on the basic parameters, i.e., accuracy, precision, and recall. The 1st and 2nd are well-balanced precision-recall ratios, but their accuracy is a bit lower, implying that they will be appropriate with simpler data structures. The 3rd and 4th setups demonstrate greater accuracy, with the 4th model showing an increased accuracy but with a slightly lower recall. The 5th setup is an appropriate combination of precision and recall, which is why it is a good choice in most cases when detecting outliers is necessary. The 6th one attains the most fantastic accuracy but experiences a slight decrease in the recall, which can limit its application in situations where strong anomaly detection is required. On the other hand, the 7th configuration scores the worst in all the metrics, and this implies that it is not well suited to work with complex data. In the 8th configuration, high recalls are observed in an expert at par with high accuracy. Interestingly, the 9th configuration returns the best balance results overall, or simply said, it has a combination of the highest accuracy, precision, and recall. This qualifies it to be the perfect solution for the thorough searching of outliers in underwater sensor networks. The detailed summary of performance metrics of different LSTM model configurations is shown in Table 4.

Table 4. Performance metrics for sigmoid activation.

Model iteration	Accuracy	Precision	Recall
1st	0.87	0.88	0.89
2nd	0.88	0.88	0.87
3rd	0.90	0.89	0.90
4th	0.91	0.89	0.88
5th	0.90	0.91	0.90
6th	0.92	0.90	0.89
7th	0.89	0.87	0.86
8th	0.92	0.89	0.90
9th	0.92	0.92	0.91

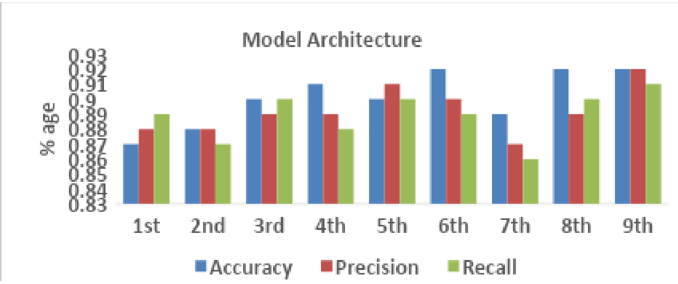

Figure 4. Performance comparison: sigmoid activation.

In testing different LSTM model design tasks, it can be seen that when increasing the model complexity, i.e., introducing more hidden layers and a larger number of units, the model performance in terms of accuracy, precision, and recall tends to improve. Remarkably, the 9th configuration is seen to be the best-scoring model since it has the best scores on all measures, as shown in Figure 4. This setup is the best when it comes to improving outlier detection in UWSNs. The choice of the best model configuration can be made according to the needs of the task, whether the task requires higher accuracy, precision, recall, or a trade-off between them.

4.10 Model Performance Metrics

The elaborate summary of the architecture offers a variety of LSTM model settings for outlier detection, all of which use the ReLU activation function. 1st architecture with a single hidden layer of 64 units is best suited to simple tasks. One hundred twenty-eight units further expand it in the 2nd architecture, which enables the model to learn more complex patterns. The 3rd architecture also offers improved capacity of 256 units, which is more suitable for the identification of complex outliers. The 4th architecture adds two hidden layers composed of 64 units and allows hierarchical feature learning and improved ability to work with complex temporal dynamics. The 6th and 5th architectures have greater depth and unit capacity, and the 6th model works specifically well with a more complex dataset because it has a greater layer size. The 7th architecture grows to three hidden layers, each with 64 units, and this enhances the ability of the model to extract hierarchical patterns. In the 8th architecture, the units per layer are expanded to 128, and thus feature learning on more complex settings is augmented. Lastly, the 9th architecture, consisting of three nested hidden layers, with 256 units in each of them, provides the most significant capacity and depth, and is optimal to use when modeling the detailed patterns and finding the subtle outliers.

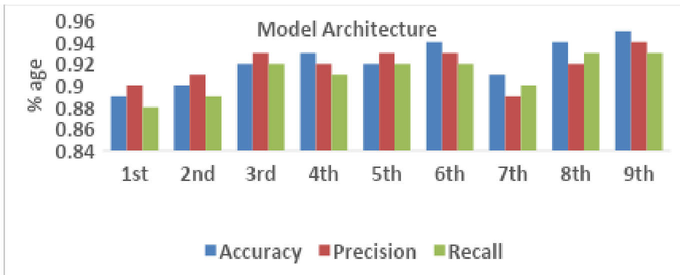

Figure 5. Performance comparison: ReLU activation.

4.11 Evaluation Metrics with ReLU Activation

When comparing the model configurations in Figure 4 in detail, it is possible to observe that the performance of different architectures varies. The 1st architecture provides significant accuracy with somewhat less recall and is usable in case of lighter datasets. The 2nd architecture will be more accurate and have more recall, and still have more precision, which is a good combination for moderate tasks. The 3rd architecture can be said to perform well in terms of high accuracy, precision, and recall, thus works well with moderately complex data sets. The 4th architecture is the most accurate architecture of the initial models; however, the precision is insignificantly less, yet it is highly effective. The 5th architecture has an equal performance to the 3rd one but has a bit higher precision. The 6th architecture is the most fantastic with significant precision as well as recall, and it's the most appropriate architecture to use when handling complex data. Although the 7th architecture displays less precision and accuracy, it is better on recall, and does not work well on precision-critical tasks. The

8th architecture offers good accuracy and recall and a reasonable precision, and is thus a reasonable and acceptable multifaceted architecture in the case of complex datasets. And at last, the 9th architecture performs the best in this general task, as it has the highest accuracy and precision; hence, it is the most appropriate to perform a complete outlier detection. Table 5 shows performance results of different LSTM model settings.

Table 5. Performance metrics for ReLU activation.

Model Iteration	Accuracy	Precision	Recall
1st	0.89	0.90	0.88
2nd	0.90	0.91	0.89
3rd	0.92	0.93	0.92
4th	0.93	0.92	0.91
5th	0.92	0.93	0.92
6th	0.94	0.93	0.92
7th	0.91	0.89	0.90
8th	0.94	0.92	0.93
9th	0.95	0.94	0.93

The comparison of various LSTM model settings indicates that more complex models achieve better performance. The 9th architecture stands out in terms of accuracy and precision, and with its high recall, it becomes the best model for detecting outliers in UWSNs. In choosing the model, trade-offs should be considered between accuracy and precision, as well as recall, to optimize the model for the particular task of outlier detection.

4.12 Mean Calculation

The mean (μ) for each metric is calculated as in Equation 2:

$$\mu = \sum \frac{x_i}{N} \tag{2}$$

For example, for sigmoid accuracy:

$$\mu_{accuracy} = \frac{0.87 + 0.88 + 0.90 + 0.91 + 0.90 + 0.92 + 0.89 + 0.92 + 0.92}{9} = 0.9011 \tag{3}$$

4.13 Standard Deviation Calculation

The standard deviation (σ) is calculated for each metric. For Sigmoid Accuracy:

$$\sigma = \frac{\sqrt{\sum (x_i - \mu)^2}}{N} = 0.0173 \tag{4}$$

Similarly, calculations are performed for all metrics:

The calculated standard deviations for Sigmoid and ReLU activation functions are summarized below in Table 6.

Table 6. SD of performance metrics for sigmoid and ReLU activation functions.

S.No	Metric	Sigmoid Std Dev	ReLU Std Dev
1.	Accuracy	0.0173	0.0187
2.	Precision	0.0147	0.0152
3.	Recall	0.0152	0.0166

Table 6 draws a comparison in terms of the standard deviations of Accuracy, Precision, and Recall given the Sigmoid and ReLU activation functions. In all these, Sigmoid users exhibit lower standard deviation values (0.0173, 0.0147, and 0.0152) more frequently than ReLU users, who have higher measures (0.0187, 0.0152, and 0.0166), albeit somewhat more consistently.

4.14 Model Performance Comparison for Sigmoid Activation

Comparison of different outlier detection models used in detail shows their significant differences concerning their performance. The non-trivial accuracy, precision, and recall realized by the proposed LSTM model with the sigmoid activation function instead of using other activation functions present in the literature make this the most effective model in managing complex patterns in underwater sensor data. Instead, ML models, such as the RF model, although compelling enough, are not as good as LSTM in describing complex outlier patterns and, as such, demonstrate both lower accuracy and precision. The overall performance of the DT model is the lowest, as it has problems handling complicated datasets. The KNN model is more accurate than Decision Trees and Naive Bayes. Still, it is less efficient than LSTM, primarily due to its sensitivity to distance metrics and high computational cost. Like other simple models, Logistic Regression does not predict well when it comes to dealing with complex outliers. SVMs are pretty good, but still not able to compete with LSTMs. Lastly, NB is the least efficient because it assumes the independence of features and therefore cannot work well whenever complex outliers have to be detected. Table 7 below provides a comparison of the different machine learning (ML) models along with the best-performing LSTM model with sigmoid activation function concerning the three performance indicators, namely Accuracy, Precision, and Recall.

Table 7. Comparison of performance metrics for sigmoid activation.

Models	Accuracy	Precision	Recall
LSTM (Best, Sigmoid)	0.92	0.92	0.91
Random Forest	0.85	0.84	0.83
Decision Tree	0.80	0.79	0.78
K-Nearest Neighbors	0.82	0.81	0.80
Logistic Regression	0.81	0.80	0.79
Support Vector Machine	0.83	0.82	0.81
Naive Bayes	0.78	0.77	0.76

4.15 Model Performance Comparison for ReLU Activation

Table 8. Comparison results of various models.

Model	Accuracy	Precision	Recall
LSTM (ReLu)	0.95	0.94	0.93
Random Forest	0.85	0.84	0.83
Decision Tree	0.80	0.79	0.78
K-Nearest Neighbors	0.82	0.81	0.80
Logistic Regression	0.81	0.80	0.79
Support Vector Machine	0.83	0.82	0.81
Naive Bayes	0.78	0.77	0.76

The detailed analysis of models for Outlier detection reveals that the proposed LSTM model with a ReLU activation function outperforms all other models, achieving the highest accuracy. A close examination of the models presenting outlier detection indicates that the suggested LSTM model with the ReLU activation function demonstrates the most significant accuracy, precision, and recall compared to all the remaining models. This model excels in detecting complex patterns and dependencies in underwater sensor data, making it particularly effective for outlier detection. Comparatively, the RF model compares well and fails to live up to the effectiveness of the LSTM model, especially in grasping complex outlier patterns. The DT model performs worse in all the measures and is unable to perform complex tasks because of overfitting and a poor pattern-recognition ability. KNN outperforms DT and NB, but remains not as good as LSTM, as it encounters problems associated with sensitivity to distance and heavy computational workload. The LR is a simple method applicable to simple sets of data, but it is not very effective in complex outlier detection. SVM achieves fair performance through careful parameter tuning, but it falls short of LSTM. Last of all, NB, due to its assumption of feature independence, does the poorest, and hence is least helpful in cases dealing with complex outlier detection. Table 8 below illustrates the comparison of different models with their efficient use in identifying the outliers.

5 Conclusion

Underwater sensor networks (UWSN) play a crucial role in surveillance and the examination of various marine environment properties. The stimulating environments in UWSN may generate unreliable sensor data, specifically outliers that can significantly impact the data quality. Anomaly detection has a critical role in data privacy and quality in UWSNs. This paper presents an LSTM-based outlier detection model in critical UWSN environments. It captures temporal patterns, making it very suitable, as far as the complexity of the underwater data is concerned. The model achieved a high accuracy of 95% in identifying various outliers. These findings have a high potential real-world implication and can be utilized in areas such as marine monitoring, infrastructure inspection, and hazard detection. In the future, real-time detection, advanced architectures of deep learning, and transfer learning will be used to provide better adaptability and performance to different environments in the underwater world.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Taher, R. J., & Mohsen, K. K. (2024). Underwater wireless sensor networks. In *BIO Web of Conferences* (Vol. 97, p. 00023). EDP Sciences. [Crossref]

[2] Masood, F., Khan, M. A., Alshehri, M. S., Ghaban, W., Saeed, F., Albarakati, H. M., & Alkhayyat, A. (2024). AI-based wireless sensor IoT networks for energy-efficient consumer electronics using stochastic optimization. *IEEE Transactions on Consumer Electronics*, *70*(4), 6855-6862. [Crossref]

[3] Mohamed, R. E., Saleh, A. I., Abdelrazzak, M., & Samra, A. S. (2018). Survey on wireless sensor network applications and energy efficient routing protocols. *Wireless Personal Communications, 101*(2), 1019-1055. [Crossref]

[4] Heidemann, J., Stojanovic, M., & Zorzi, M. (2012). Underwater sensor networks: applications, advances and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 370*(1958), 158-175. [Crossref]

[5] Jouhari, M., Ibrahimi, K., Tembine, H., & Ben-Othman, J. (2019). Underwater wireless sensor networks: A survey on enabling technologies, localization protocols, and internet of underwater things. *IEEE Access, 7*, 96879-96899. [Crossref]

[6] Amoli, P. V. (2016). An overview on current researches on underwater sensor networks: Applications, challenges and future trends. *International Journal of Electrical and Computer Engineering, 6*(3), 955. [Crossref]

[7] Fattah, S., Gani, A., Ahmedy, I., Idris, M. Y. I., & Targio Hashem, I. A. (2020). A survey on underwater wireless sensor networks: Requirements, taxonomy, recent advances, and open research challenges. *Sensors, 20*(18), 5393. [Crossref]

[8] Jiang, S. (2018). On securing underwater acoustic networks: A survey. *IEEE Communications Surveys & Tutorials, 21*(1), 729-752. [Crossref]

[9] Chartres, C., & Varma, S. (2010). *Out of water: from abundance to scarcity and how to solve the world's water problems*. FT Press.

[10] Li, S., Qu, W., Liu, C., Qiu, T., & Zhao, Z. (2019). Survey on high reliability wireless communication for underwater sensor networks. *Journal of Network and Computer Applications, 148*, 102446. [Crossref]

[11] Otero, P., Chowdhry, B. S., Zeng, H., & Aamir, M. (2021). underwater communications and sensors technologies. *Wireless Personal Communications, 116*(2), 955-961. [Crossref]

[12] Felemban, E., Shaikh, F. K., Qureshi, U. M., Sheikh, A. A., & Qaisar, S. B. (2015). Underwater sensor network applications: A comprehensive survey. *International Journal of Distributed Sensor Networks, 11*(11), 896832. [Crossref]

[13] Awan, K. M., Shah, P. A., Iqbal, K., Gillani, S., Ahmad, W., & Nam, Y. (2019). Underwater wireless sensor networks: A review of recent issues and challenges. *Wireless Communications and Mobile Computing, 2019*(1), 6470359. [Crossref]

[14] Gola, K. K., & Gupta, B. (2017). Underwater sensor networks routing (UWSN-R): A comprehensive survey. *Sensor Letters, 15*(11), 877-890. [Crossref]

[15] Eleftherakis, D., & Vicen-Bueno, R. (2020). Sensors to increase the security of underwater communication cables: A review of underwater monitoring sensors. *Sensors, 20*(3), 737. [Crossref]

[16] Das, A. P., Thampi, S. M., & Lloret, J. (2020). Anomaly detection in UASN localization based on time series analysis and fuzzy logic. *Mobile Networks and Applications, 25*(1), 55-67. [Crossref]

[17] Rault, T., Bouabdallah, A., & Challal, Y. (2014). Energy efficiency in wireless sensor networks: A top-down survey. *Computer networks, 67*, 104-122. [Crossref]

[18] Cao, X., Liu, L., Cheng, Y., & Shen, X. (2017). Towards energy-efficient wireless networking in the big data era: A survey. *IEEE Communications Surveys & Tutorials, 20*(1), 303-332. [Crossref]

[19] Meleshko, A. V., Desnitsky, V. A., & Kotenko, I. V. (2020). Machine learning based approach to detection of anomalous data from sensors in cyber-physical water supply systems. In *IOP conference series: materials science and engineering* (Vol. 709, No. 3, p. 033034). IOP Publishing. [Crossref]

[20] Liu, N., Chen, D., Huang, H., Huang, X., Yang, Q., & Xiong, N. N. (2022, July). Anomaly Detection of Underwater Sensor Data Based on Temporal and Spatial Correlation. In *International Conference on Artificial Intelligence and Security* (pp. 239-252). Cham: Springer International Publishing. [Crossref]

[21] Saeed, N., Al-Naffouri, T. Y., & Alouini, M. S. (2018). Outlier detection and optimal anchor placement for 3-D underwater optical wireless sensor network localization. *IEEE Transactions on Communications, 67*(1), 611-622. [Crossref]

[22] Samara, M. A., Bennis, I., Abouaissa, A., & Lorenz, P. (2022). A survey of outlier detection techniques in IoT: Review and classification. *Journal of Sensor and Actuator Networks, 11*(1), 4. [Crossref]

[23] Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural computation, 31*(7), 1235-1270. [Crossref]

Sunbal Rani received the MS in Computer Science from Abasyn University, Pakistan, in 2025. She is currently serving as a Lecturer in the Department of Computing at Abasyn University. She has contributed to various national and international research publications. Her research interests include artificial intelligence, deep learning, underwater sensor networks, and anomaly detection. (Email: sunbal.rani@abasyn.edu.pk)

Saqib Shahid Rahim received the MS in Computer Science from Preston University, Pakistan, in 2017. He is currently serving as a Lecturer in the Department of Computing at Abasyn University. He has contributed to various national and international research publications. His research interests include ratification intelligence, networking, deep learning, and underwater sensor networks.