SMART WASTEWATER MANAGEMENT: A HYBRID DEEP WAVELET NETWORK MODEL FOR IMPROVED WATER QUALITY FORECASTING

J. Priskilla Angel rani J[1,*], Dr.C.Yesubai Rubhavathi[2]

[1] Assistant Professor, Department of CSE, Francis Xavier Engineering College, Tirunelveli, India.

[2] Professor, Department of CSE, Saveetha Engineering College, Chennai, India

[*]Corresponding E-mail: *pricy.angel@gmail.com*, tel: 9600013402

ACCEPTED MANUSCRIPT

ABSTRACT

The process of predicting the water quality is done by various predictions algorithms using Neural network based on Artificial Intelligence and Machine Learning methods. Several water quality detecting indicators are used to measure the purity level of water with the amount of usage based on topography and area of utilization. Its objective is to lessen the water scarcity in the region by giving valuable guidance and solutions. In order to handle wastewater treatment and forecasting, a hybrid wastewater forecasting model has been presented. Among many other factors, this prediction procedure takes into account pH, COD, BOD, ammonia, pressure, and humidity. Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and Discrete Wavelet Transform (DWT) are all integrated in a Hybrid Deep Wavelet Network (HDWN). In HDWN model, CNNs retrieve spatial information, LSTMs model temporal dependencies, and DWT breaks down input signals. This integration makes it possible to make precise forecasts and insights, which makes it easier to monitor and optimize treatment systems effectively and lessen South Tamil Nadu's water scarcity. This model performance evaluated using metrics MSE is 0.068, RMSE is 0.2325, MAPE is 0.55 MAE is 0.1945 and R^2 is 0.932. The comparative analysis of the tested results based on Integrated Neural Network LSTM model shows best predictions. Hence this HDWN model is useful for prediction and forecasting the wastewater quality.

Keywords: Prediction Algorithms, Wastewater treatment, Water Quality Parameter, Machine Learning, Convolution Neural Networks, Long Short-Term Memory Algorithm.

ACCEPTED MANUSCRIPT

1. Introduction

The Urbanization leads to the need of water demand in various purposes of agriculture, large factories, household, manufacturing unit etc. The various sources of water are pre-owned by many human activities which in turn lead to the rise of water scarcity in many areas of our country. Several factors such as increase in man power, development of cities and factories and domesticated life style makes all the renewable resources depleted in future years. The authors guvava, M et al. (2019) insisted the essential usage of pure water in adequate quantity with measurable quality range in various levels of applications in many districts need plenty of fluid resources. It is possible to reuse the treated water from industry and costal water. Ziadi, A. et al. (2019) This research investigation in the sources of pollution and mineralization in the Lebna, Cap Bon, Tunisia's coastal water sources. In order to determine the anthropogenic and natural elements influencing water quality, hydro chemical features were observed. The research revealed serious pollution caused by saline intrusion and farming practices, endangering the area's water supplies. Zhou, Y. et al. (2020) The authors say the research evaluated methods for reducing pollution in coastal waterways that are significantly impacted by human activity. It made clear that in order to effectively prevent pollution, coordinated methods integrating technological, ecological, and regulatory measures are required. Many physical factors are analysed to measure the water purity level by predicting the amount of usage and wastage. The water purified by the process of waste water management by applying biological factors, physical factors and chemical treatments. Many upcoming methodologies are discussed in various research regarding the water quality prediction and aim in measuring the pure water and its availability in various sectors. In order to forecast the Coastal Water Quality Index (WQI), Uddin et al. (2022) assessed a number of machine learning models, such as Decision Trees, Random Forests, and Extreme Gradient Boosting. According to their research, tree-based models fared noticeably better than other models in terms of prediction

ACCEPTED MANUSCRIPT

accuracy. These models were used to assist manage water quality in coastal habitats by lowering uncertainty and offering a solid foundation. This method represents a shift from conventional pollution control techniques, such those proposed by Tayeb et al. (2015), to predictive, data-driven water management. Chaukura, C. et al. (2020) provides a detailed review of current concerns about the presence and elimination of disinfection byproducts in the consumption of water. This research discusses the challenges and developments in addressing the issues. In 2020, Álvarez-Ruiz, R. and Picó, Y. conducted carried out research analyzing the contaminants and new methods affecting different types of water resources. It emphasizes the importance of water treatment and monitoring the difficulties of treatment process. Ahmed, U. et al. (2019) examined effective water quality prediction by means of supervised machine learning algorithms. Effluent waste water treatment is essential for the treating the waste water to improve the quality and quantity of usage in several applications. The process includes eight stages of purification first is the screening technique of the sample followed by the effective screening and clarification process with aeration to avoid unwanted effluents in water and supply of chlorination to avoid disinfection and water testing is done for better results which result in disposal of major and minor wastes and avoids loss of water quality enabling better purified water as the end product. Many sensing techniques are used to measure the rate of water evaporated or transpiration occurred may cause loss of wate apart from being contaminated. Industrial wastes, household waste and other solid and liquid waste gets mixed with the available water and reduce the level of availability of pure water. Measuring the water purity is based on pressure and this includes two process one is primary process and another one is secondary process. These methods include adding chlorine and finally with ozone treatment based on pressure retention that separates pure water and removes other impurities which makes 99.8 percent of purity measure and this percentage differs based on topography. This ozone can treat germs better than chlorine. This water can be taken as samples from various soil texture for analysing and predicting the level of purity based on factors mentioned

ACCEPTED MANUSCRIPT

above. Arepalli, Naik, and Amgoth (2024) suggested an Internet of Things-based framework utilizing the Cluster-Guided Temporal Flexible Network (CGTFN) model to improve the precision of water quality analysis and categorization in treatment procedures. Because it incorporates ambient and water characteristics to enhance classification accuracy and guarantee energy-efficient, real-time monitoring, this framework is ideal for smart aquaculture systems.

1.1 RELATED WORK

In recent decades, both mechanical and non-mechanical way of water quality predictions were applied for water quality prediction purposes. Ahmed, U. et al. (2019) When predicting water quality, a non-mechanical method can yield forecasts that are more accurate than those produced by conventional methods. Based on regression and random forest WQI is predicted. Liu, P. et al. (2019) and Hu, Z. et al. (2019) both investigated and provide the prediction of water quality using deep learning algorithms. The two experiments demonstrate how LSTM networks may enhance water quality prediction and management. LSTM model is used to predict the PH and COD variable with historical values for prediction. The model performed better than the ARIMA and SVR models, with a mean squared error (MSE) of 0.0017. By applying a data-mining technique, Asami, H. et al. (2021) simulated the Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD) and Total Suspended Solids (TSS) particles in wastewater treatment facilities. The research used advanced data-mining techniques to model and forecast important wastewater treatment parameters. The outcomes showed how these methods could be used to maximize the effectiveness and efficiency of wastewater treatment systems. likewise, data mining has been effectively applied to environmental evaluations of wastewater treatment plants in order to estimate missing data and reduce monitoring expenses.

Boyd et al. (2019) and Chaoui et al. (2023) managed missing data in their wastewater dataset using an ARIMA-based imputation technique, and then employed an ANN and RNN models for prediction. Their enhanced approach yielded a 90% increase in oxygen injection and a 37%

ACCEPTED MANUSCRIPT

reduction in energy use throughout the evaluation period, demonstrating that machine learning may greatly optimize resource use in wastewater treatment. Rani et al. (2022) and Furqan Rustam et al. (2022) developed an efficient artificial neural network (ANN) model for predicting water quality and usage. The model outperformed earlier techniques and shown resilience while processing data on water quality. However, an unbalanced dataset that can lead to overfitting hampered the study. Future studies should focus on expanding and balancing the dataset in order to improve the generalizability of the model. Vijay Anand et al. (2023) developed a neural network model that uses images of water samples to predict water quality using CNN, TensorFlow, and Keras. The program evaluates water compatibility based on color and overall quality, with an initial testing accuracy of 85%. Despite showing promise for the initial screening, more data integration and algorithm improvement are needed to enhance its efficacy. In 2022, Junhao Wu and Zhaocai Wang introduced a novel ANN-WT-LSTM model that uses wavelet transform and time-series imputation to reduce noise.Across all model this model outperforms. An IoT-enabled aquaculture pond's water quality can be classified using a Flexible Temporal Network (FTN) model based on deep learning, according to Arepalli and Naik (2025). Their method allows for precise real-time monitoring by efficiently capturing temporal fluctuations in environmental and water data. However, the accuracy of the model as a whole may be impacted by the quality and dependability of the sensor data.

Similarly, to assess aquatic water quality data, Arepalli and Khetavath (2023) created a Time-Series Convolutional Neural Network (TS-CNN) model. The TS-CNN is appropriate for dynamic aquatic environments because it can detect short-term trends in water quality measurements with little preprocessing. Notwithstanding its advantages, this model has trouble identifying long-term relationships, which restricts its capacity to forecast long-term trends in water quality.

The below table shows presents a comprehensive review of the wastewater quality prediction studies, highlighting the methodologies used in each paper

ACCEPTED MANUSCRIPT

S.No.	Author(s) and Year	Method / Algorithm	Task / Application	Dataset / Context	Advantage(s)	Limitation(s)
[1]	Asami et al. (2021)	Data Mining	Estimate missing data, reduce monitoring cost	Sensor networks	Improves system efficiency	Requires preprocessing
[2]	Chaoui et al. (2023)	ARIMA, RNN	WWT plant control	Time series data	Captures temporal trends	Vanishing gradients; linear assumptions
[3]	U. Ahmed et al. (2019)	Regression, RF, MLP	WQI/WQC prediction	Surface water data	Classification + regression	Limited generalizability
[4]	Rustam et al. (2022)	ANN	Water use & quality prediction	Mixed sensor + external data	Dual target modeling	Overfitting risk
[5]	Vijay Anand et al. (2023)	CNN	Water quality classification	Local features from sensors	Effective spatial feature extraction	Imbalanced and limited features
[6]	Hu et al. (2019)	LSTM	Forecasting water quality	Historical time-series data	Captures sequence dependencies	Needs large data; high compute cost
[7]	P. Liu et al. (2019)	LSTM, DNN	Quality prediction	Single-variate time series	Deep learning effectiveness	Needs tuning and optimization
[8]	Wu & Wang (2022)	WT, ANN, LSTM	Single-point reservoir quality prediction	Reservoir data	Combines multiple techniques	Lacks spatial generalization
[9]	Arepalli & Naik (2025)	Flexible Temporal Network (FTN)	IoT-enabled pond water classification	Sensor-driven aquaculture	Real-time prediction	Sensor dependency
[10]	Arepalli & Khetavath (2023)	Time-Series CNN (TS-CNN)	Water quality trend detection	IoT time-series	Strong short-term pattern recognition	Poor at long-term learning

Table 1.1. Survey on Water Quality Prediction

The working model discusses in detail about the process of purification followed by water sample analysis.

2. Working Methodology:

ACCEPTED MANUSCRIPT

2.1 Water Purification by Ozone at various pressure levels

Ozone air is pumped into the sample in the sample tank and the ozone is not directly taken by sample. It is good to pump ozone with a maintained level of 40 – 50 Pascal unit of pressure at optimal rate. The small amount of water taken 12.4 gallon per meter is pumped in to transparent tube with electrical charges plated at the ends and ozone is pushed into the tube for purification and the particles are dissociated as then the water components are separated from impure components and the wastewater is removed in gaseous state. The oxygen combines with ozone and hydrogen is released. This removes the waste and the germs from the water and a pressure is applied backward to retain the purification of water till the process completes.

The quality of water is important for various usage based on where the water is obtained by us for our needs and several factors determine the water gets affected by impure substances, germs, viruses and harmful chemicals which makes it not suitable for usage of the human [Zhou, Y. et al. (2020), Holkar, C.R. et al. (2016)]. A large number of neural network models have been created to forecast the long-term quality of water. These models make use of neural units that can handle various datasets with different water samples according to a number of physical characteristics. They are categorized, trained to produce desired results, and often their performances are compared [Guedes-Alonso, R. et al. (2020)]. Ahmed, U. et al. (2019) done his research on efficient quality prediction in wastewater using supervised ML algorithms. Several purification approaches based on water quality and quantity obtained at raise or fall of level in source are evaluated by many valuable factors that physical (pressure, humidity, evapotranspiration, temperature etc) or chemical (acid rain, polluted water wastes). However, this research relies on large dataset for training the model and also which could restrict its application in areas with discontinuous or poor water quality data. In our research proposal a simulated and comparative studies are done based on ANN-LSTM, DWT-LSTM and CNN-LSTM to improve the efficacy of the neural models and comparative tables and graphs are generated using python code;

2.2. Treatment of Water Sample

ACCEPTED MANUSCRIPT

The water treatment can be carried out based on chlorination techniques in which sample of water gets added to the different forms of chlorine such as sulphates, hydrates which are capable of removing germs and many hard substances found in water samples and convert it to a suitable form for drinking. The water quality is the most powerful element considered as the best outcome among the threats created by population and pollution in recent era. The physical parameters we had addressed in the research are pressure and evapotranspiration.

In order to improve water management, Ashwini, K. et al. (2022) carried out research to create an in telligent model for predicting water quality. The LSTM working suggests many quality factors for prediction that are evaluated with improved correlation mapping when analysed with other neural models.

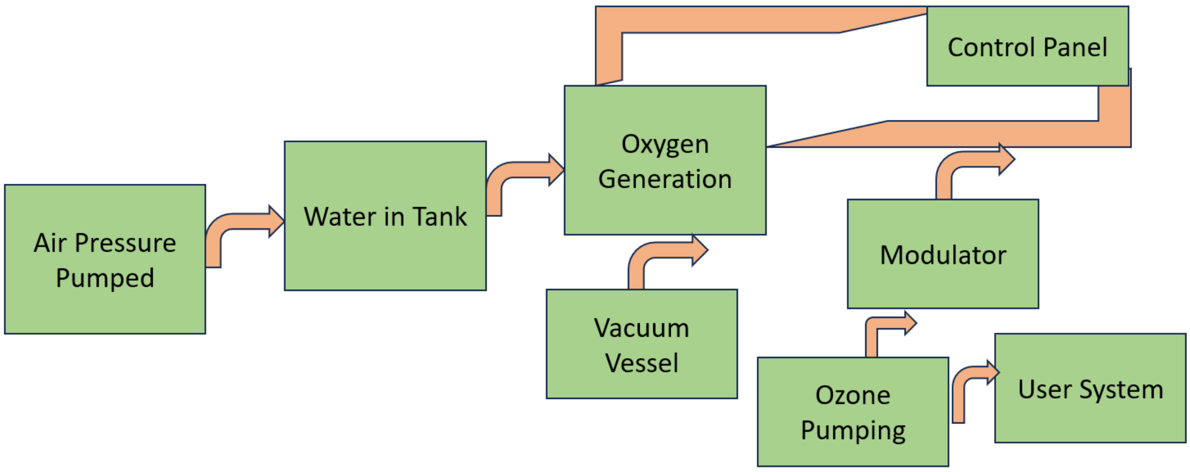


Figure 2.2. Model of Ozone purification

Some oxygen elements are extracted from the air when it is pressured and pumped into the vacuum vessel on a great force then when ozone is pumped in to the water sample in the tank which is capable of purifying the sample. The whole system is supported by the modulator and managed by the controller by the control panel, based on various atmospheric pressure the process is repeated again and again for better accuracy.

2.3 Waste Water Treatment Process

The main objective of wastewater treatment methods is to improve the accuracy of water treatment. There are two basic stages in the waste treatment that can be basic and additional technique, which can be essential for treating the waste samples to a purified water. In the basic step, the unwanted material is sedimented at the bottom and selector removes those wastes by residual activation and additional methods includes sludge handling and buffering and the process is repeated to remove the aerobic digestion from pre thickening of the sediments and there by the water gets purified effectively. Chen, H.J. et al. (2013) looked into the effectiveness of treatment and the microbial community. subsequently explored the relationship between microbial diversity and the effectiveness of treatment.

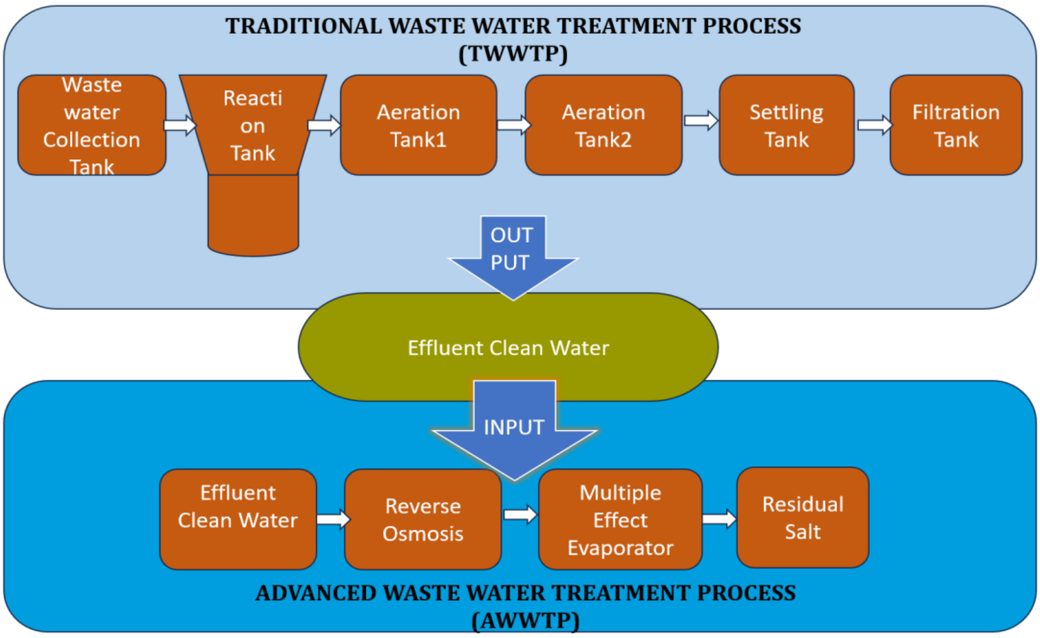


Figure 2.3. Waste Water Treatment Process

2.4 Effects of Water and water quality in ground water

There are several factors for water loss and majority of the water are dirty and not suitable for usage for various purposes. Wang, M. et al. (2017) researched on how effectively wastewater treatment works to prevent the spread of antibiotic resistance in the environment. To safeguard human health, the ecosystem's integrity, and the economy's sustainability, it is imperative to have a complete understanding of the effects on water quality. Contaminated water poses serious risks to human health, harms aquatic ecosystems, and impedes agricultural and industrial operations. Thus, in order to address the shortage of water and guarantee that different water-dependent areas have access to high-quality water, water treatment is important for the purification and can avoid the rate of evaporation from the soil is predicted by the radiation level of the soil top to maintain the water level. There by we can retail the water level and quality at a considerable rate on the living areas. The fertility of the soil is depleted by crops as they develop quickly; this is dependent on the cultivated area. Crop growth inhibition lowers soil water loss because it decreases evapotranspiration. The rate of water loss is noticeably higher in mature trees, nevertheless. The transpiration covered from the layer of sand, can be the major process. Water obtained is measured as the amount got per unit area (litre/ meter). This amount refers to the water vapour from sources of water by the above two process. The heat of evaporization (Hevap) denotes the hotness used for evaporation of water to vapour. If 1ml of water is vaporised then energy can be in vaporised form based on the regions either hot, cod or medium in temperature. In the area of humid climate or rainfall region the evaporation is less compared to the high temperature regions.

The table below gives the temperature and water loss of several regions in Tamil Nadu . The region as classified as tropical, sub-tropical, humid, arid area and temperature varies accordingly and the rate of water evaporated differs based on many aspects.

Regions	**Avg Temp (°C)**		
	Cold	**Medium**	**Hot**
	~5°C	**25°C**	**> 40°C**
Regions of Tropical			
- humid / sub	4– 5	3.5–4.5	6.5- 8
-arid area	1.8–3.6	4.0–5.6	7– 8.1
Region of Temperate			
- humid area	0.9– 1.8	2.3 – 5	3.6 – 6.8
-arid/ semi	1.2– 3.8	3.7 – 6.7	7.5 – 10

Table 2.4. shows the regions and temperature changes that affect the water evapotranspiration

ACCEPTED MANUSCRIPT

The scarcity of water in dry areas or the excess intake of water by living organs (plants, animals or humans) can also affect the cycle of inhabitation and the process of evapotranspiration. Management of water supply to the crops the intake and usage of living beings can also considered for improving the water cycle rate based on evaporation, transpiration, precipitation and so on. The method of growing crops can be also taken into account along with the climate, type of plants, soil nature, and area of cultivation. The rainfall, wind, temperature also has serious factors to be considered. Water demand rises as a result of substantial effects on different areas caused by decreases in water levels from sources like groundwater or rainfall. One of the main causes of water scarcity in all its manifestations is still the dryness of the soil surface. One of the main indicators of water scarcity is the persistence of dry soil surfaces. Various irrigation techniques, such as drip systems, bore wells, and traditional methods, can be used to decrease water loss in cultivated lands in order to address this issue. The type of plants grown determine the rate of water loss by evaporation or by transpiration. Every factor to be considered is to minimise the water loss. Prasad, A.N. et al. (2015), Mohammed, H. et al. (2018), and Singh, P. et al. (2017) researched a variety of methods for monitoring and predicting water quality. To maintain the ecosystem wellbeing waste water can be treated and water quality must be maintained and properly channelized for the proper usage of water in all conditions.

3. Materials and methods

3.1 HANDLING DATA SETS

3.1.1 Data mining with Normalization

Normalization results bring the output of many different working models in neural networks that are scaled and classified and the correlation derived gives a normalized value based on experimentation with diffused samples of water based on parameters of air pressure or evapotranspiration. The samples mean is calculated with maximized or minimized samples as

$$X\ meam = Xi\ - X\ min\ /\ Xmax - Xmin \qquad (1)$$

ACCEPTED MANUSCRIPT

The mean value is obtained by the above formulation or equation (1) based on normalization which gives the extract of datasets based on data variability.

3.2 Proposed Methodology for Water Quality Prediction

The diagram below shows the quality prediction based on HDWN models, which helps to determine the samples been tested and then trained for various predictions. Initially data samples were collected from Tamil Nadu regions and the model process are given blew

Data Preprocessing:

Data processing is the first step to dealing with missing values, outliers, other anomalies in the data, preliminary data cleaning and preprocessing are necessary. Using the Discrete Wavelet Transform (DWT), the input time series data is divided into various frequency bands.

Feature Selection:

Convolutional Neural Networks (CNNs) are used to extract spatial characteristics from the date set. Even though the data set contains many features and select the feature which are more helpful in the prediction process. Within the wastewater treatment settings, convolutional layers with filters are used to capture spatial correlations and local trends.

Modeling with LSTM:

Using Long Short-Term Memory (LSTM) layers to model temporal dependencies and identify long-term trends in time series data. LSTM learns the hierarchical information about the data in prediction process.

Model Architecture:

Decomposed input signals are accepted by the input layer.

DWT layers: Use DWT to separate the input signals into their component frequencies.

CNN layers: Convolutional layers are used to extract spatial information from the decomposed data.

LSTM layers: Recognize long-term trends in the data and model temporal dependencies.

Output Layers Produces forecasts for future wastewater parameters at the output layer.

ACCEPTED MANUSCRIPT

Training and Testing:

Eighty percent of the data in this suggested technique were utilized for training the model, and twenty percent were used for assessing the model's predictive capabilities. MSE, RMSE, MAE, and MAPE have all been used to evaluate the model.

Finally measured the quality and quantity of water based on the sample parameters considered for various water collected from several areas based on Ph, COD, BOD, Total Nitrogen temperature, humidity, soil texture they are analysed and tabulated and finally the comparative results show the best water prediction models and the factors to be considered in future evaluations. The detailed process of the HDWN model is given below in Algorithm 1.

Algorithm 1: Hybrid Deep Wavelet Network (HDWN) Wastewater Quality Prediction Model

---

1: Input data D= $\{d_1, d_2, \ldots\ldots d_n,\}$ // ex: Ph, turbidity, COD, BOD, etc.,

2: Preprocessing: handling missing data and Normalization

    D_{Clean} = Normalize (Clean (D))

3: Wavelet Decomposition

    Wavelet Function: W= Wavelet(type)

    Decompose data: C=Decompose (D_{clean}, W)

    where C = $\{C_{low}, C_{high}\}$.

4: Feature Extraction

    F=Extract_Features(C) // Extract from both time and frequency domains.

5: **Model Initialization**

    M=HDWN_Model ()

    Design the hybrid model using convolutional (CNN and LSTM).

6: **Model Training**

    M_{train} =Train (M, F, L, loss="MSE', optimizer="Adam")

7: **Validation**

    S_{val}=Validate (M_{train}, D_{val}).

8: If S_{val} > threshold

9: Tune hyperparameters H=$\{h_1, h_2, \ldots, h_n\}$ // Ph, BOD, COD etc.,

10: **Prediction**

    P = Predict (M_{train}, D_{test}) // Generate predictions for future wastewater parameters.

11: **Performance Evaluation**

ACCEPTED MANUSCRIPT

$E = \text{Evaluate}\ (P, D_{actual}, \{RMSE, MAE, R^2\})$

Assess model accuracy using performance metrics.

12: **Output Results**

Output: Predicted parameters P, trends, and visualizations.

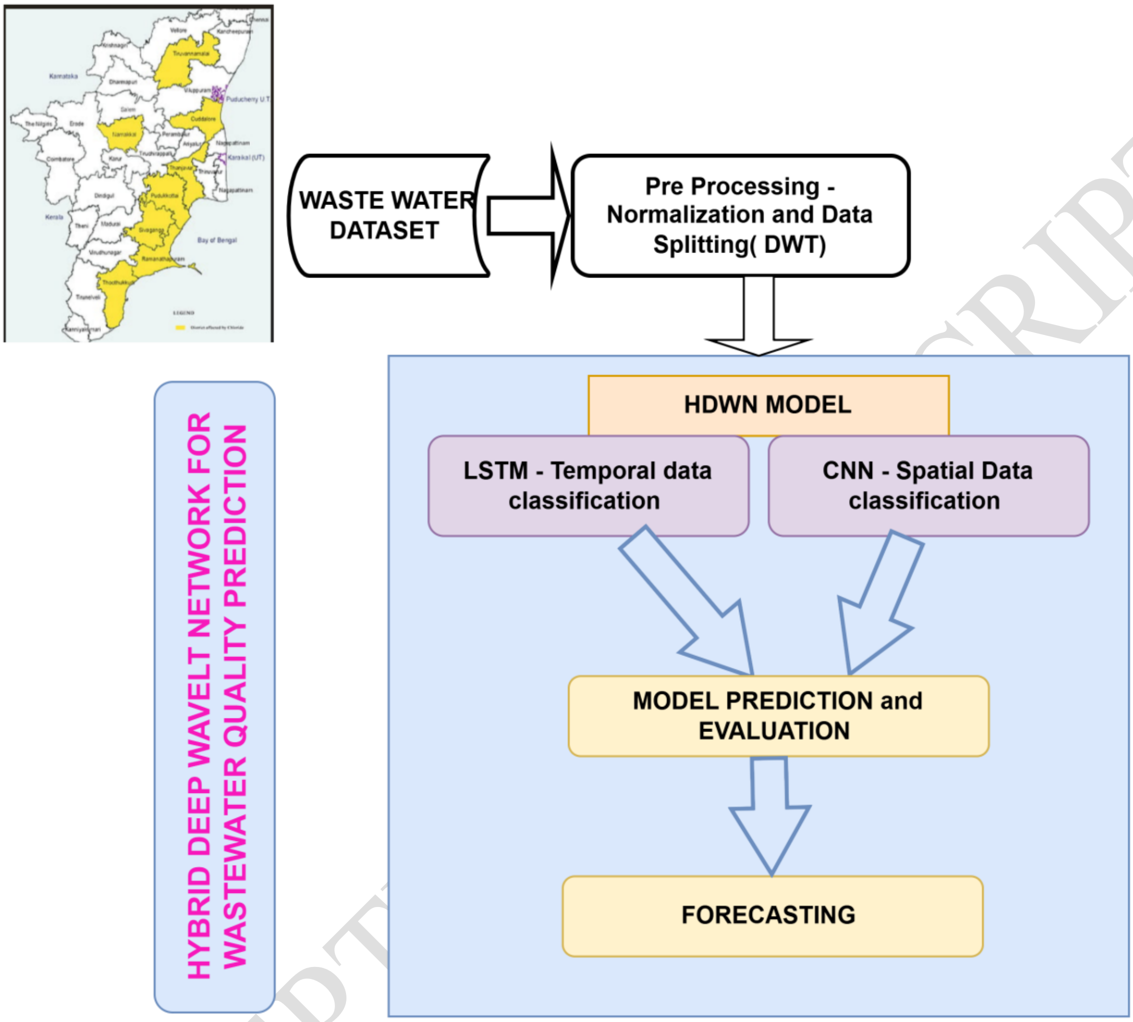


Figure 3.2. Architecture of Hybrid Deep Wavelet Network Model

3.3 Data Flow Diagram

The dataflow diagram is the pictorial representation of workflow of how the water sample are analysed and predicted based on several long- and short-term memory process and the classification and predictions are made with best accuracy and the output is compared with several iterative analysis and predictions made many times which is capable of providing many precise outputs. The Figure 3.3 shows the data flow diagram is with the process of prediction and comparison.

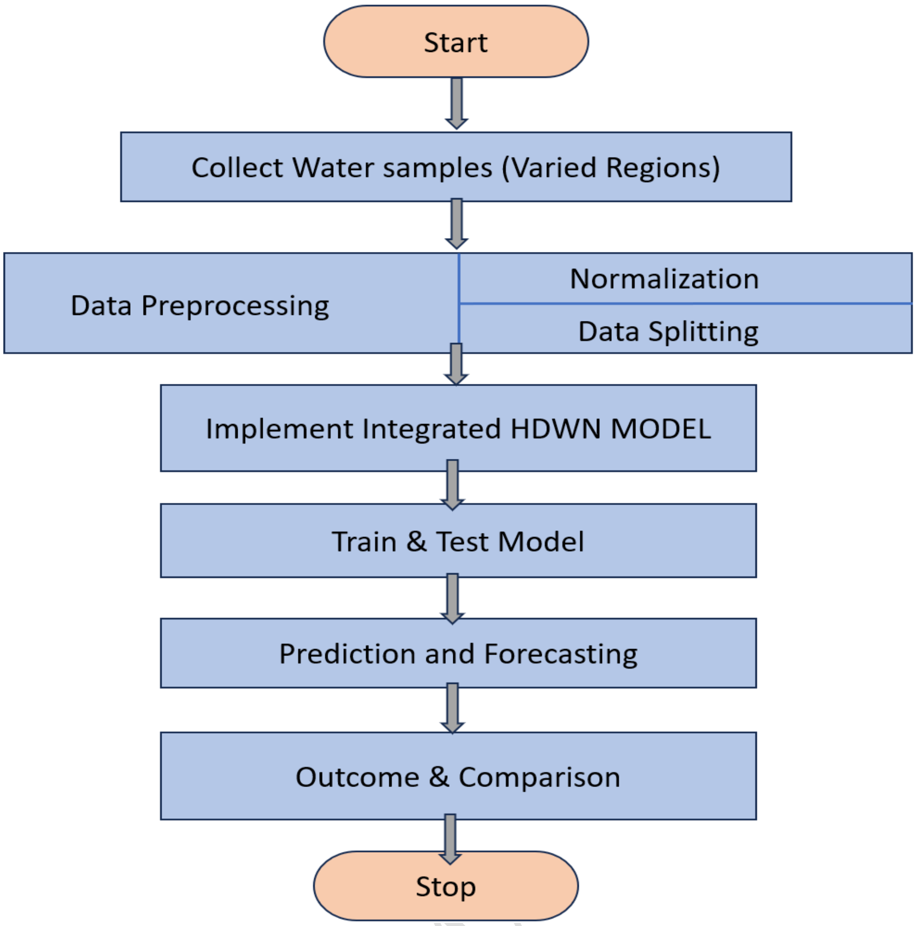


Figure 3.3 Data Flow Diagram of prediction and comparison of water samples.

3.4 Neural Network Structure

The Neural models in machine learning which includes three basic units called the input layer, the middle layer and output layer, where Xi is the input sample and middle layer denoted as Xim and output layer Yi. The model diagram below shows the neural network with n number of nodes.

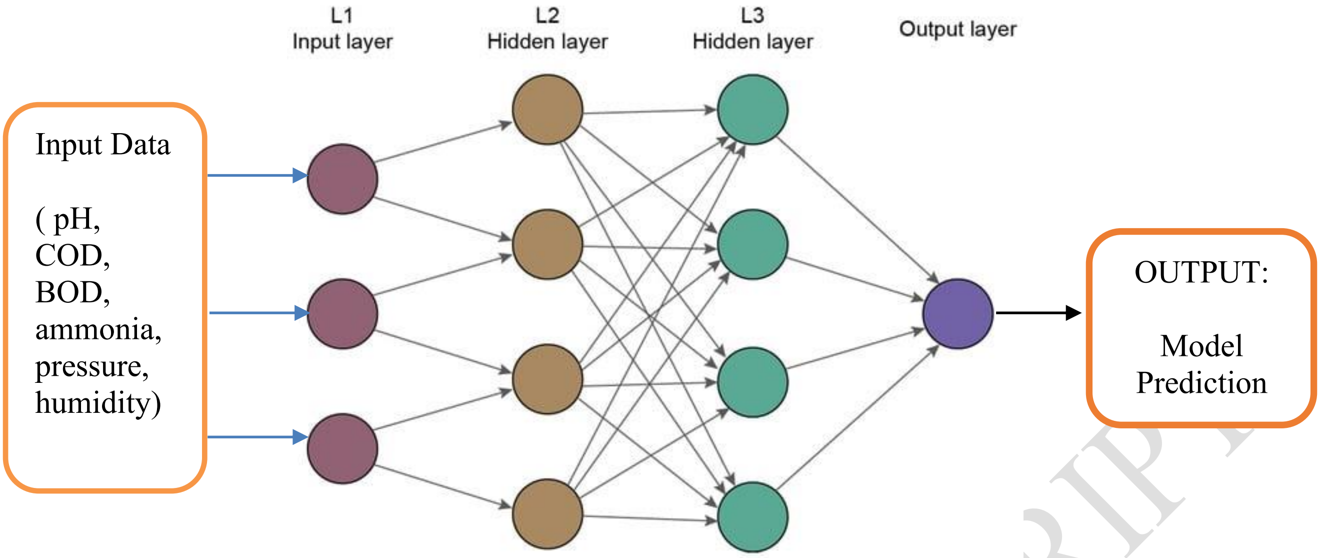


Figure 3.4 Structure of the Neural Network Model

The neural model's structure is depicted in Figure 3.4.1. It analyzes X input samples, sends them to many hidden layers, and then outputs Y samples.

3.5. Formulation in Wavelet Transform

Transformation of Wavelet series holds samples based on time, error value (mean/sample) or other factors for processing the outcomes. The Fourier transform in series which super imposes the input with varied frequencies. This wavelet transforms based on CNN or DWT-CNN with support of LSTM helps in predicting the values with high efficacy depending on varied frequency range. This wavelet transformation divides the waves considered as input in many bands of signals with varied feature namely frequency, time, error, and bandwidth. This is capable of enhancing the signal localization to overcome its basic limitations. An CNN-LSTM with Attention mechanism for forecasting effluent quality is presented by Li Y et al. in 2023.To improve prediction accuracy, attention mechanisms with CNN and LSTM networks. Their results point to promising developments in effluent wastewater quality prediction, which can greatly enhance wastewater treatment procedures and guarantee water quality standards.

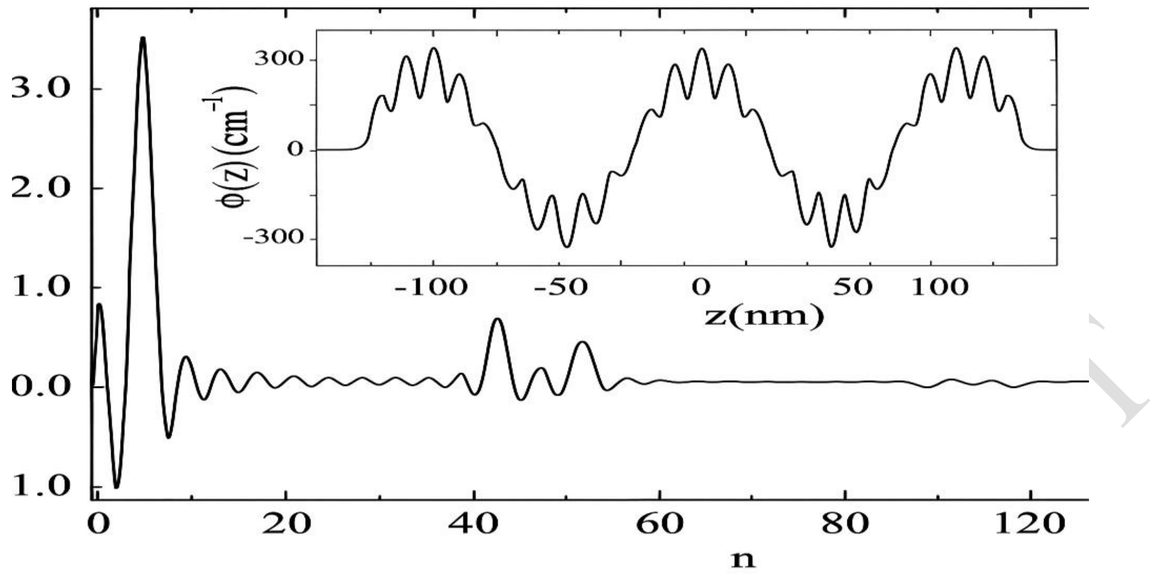


Figure 3.5 shows the wavelet transform generation based on n samples with Fourier formulation

3.6 Convolution neural networks with LSTM

Convolution neural networks with LSTM in Figure 3.6.1 is based on LSTM has many filtrations of samples in parallel at all possible levels of working gives best trained samples as the output. The LSTM with every other information corresponding to the state are processed across several times with factor of time added in turn to the input samples considered for analysis. These different models have predictions made very easy even though handling many datasets as samples in an iterative.

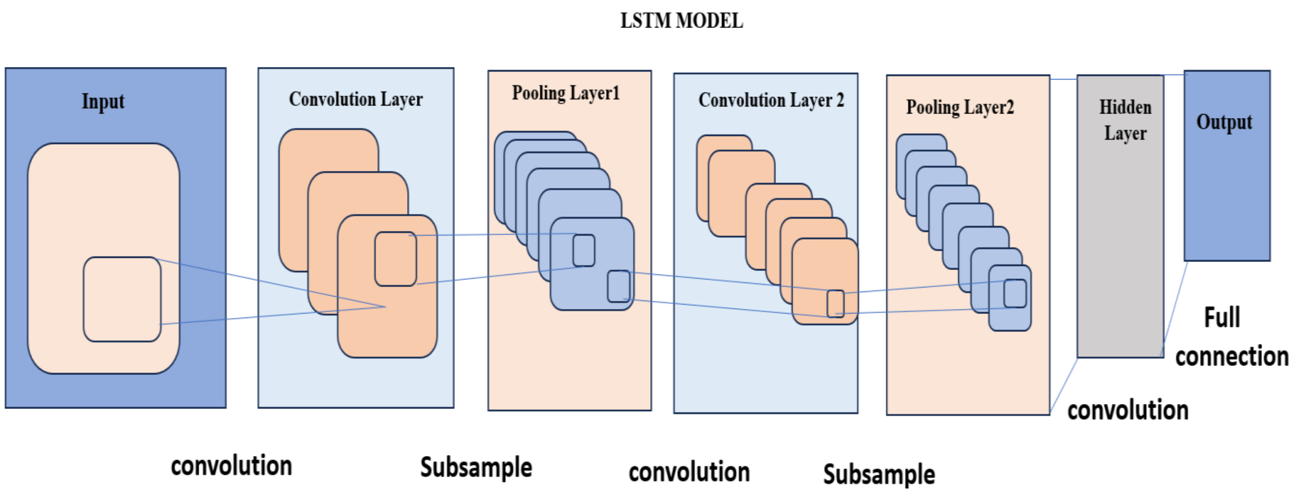


Figure 3.6.1. shows the LSTM model

The input samples of water with measured quantity in liters or milliliters are collected and convoluted as sub samples and then pooled together and the process gets repeated several times and passed through the hidden layers and finally combined as a single output which best results based on accuracy. Figure 3.6.2 shows the Process of LSTM with convolution model for data samples collected from different regions and the data has been transmitted to different layer before reaching the output layer.

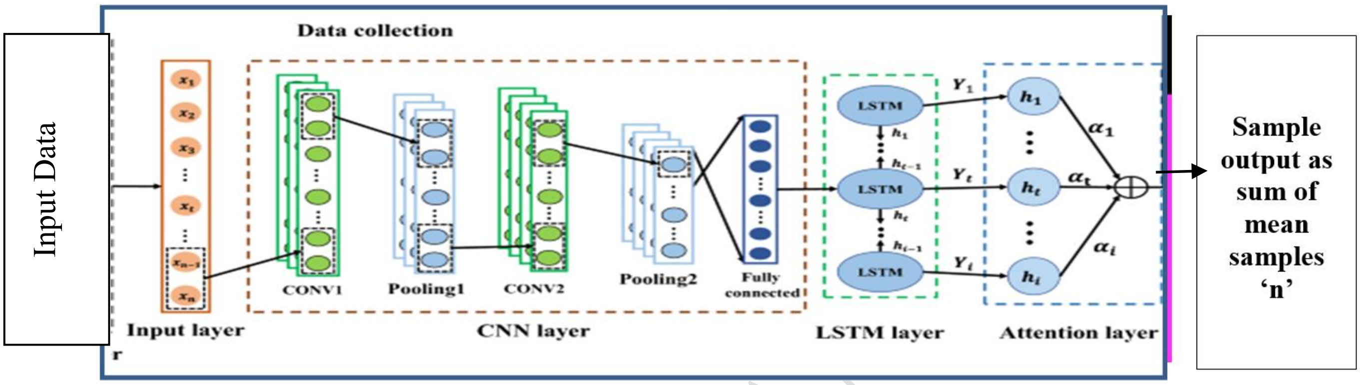


Figure 3.6.2 shows the LSTM with convolution model handling n different samples

3.7 Mechanistic Insight: DWT's Role in Enhancing CNN/LSTM:

The proposed HDWN architecture includes a preprocessing step using Discrete Wavelet Transform (DWT) to address the main shortcomings of traditional deep learning models like CNN and LSTM.While CNNs are adept at extracting spatial features, they can overfit noisy datasets and cannot handle temporal variability.Although LSTMs are designed to capture temporal dependencies, they struggle with non-stationary signals and long-term dependencies when high-frequency fluctuations or irregular trends are present.

To improve learning efficiency and enhance signal quality, the input time-series data stored in CSV format were subjected to Discrete Wavelet Transform (DWT). The dataset was tabular, but the temporal nature of parameters like pH, BOD, COD, ammonia, pressure, and humidity enabled effective decomposition. DWT divides each signal into low-frequency (approximation) and high-frequency (detail) components, separating noise from significant trends. Thanks to this denoised, multiresolution representation, Convolutional Neural Networks (CNN) can extract spatial features that are more robust, and Long Short-Term Memory (LSTM) units can model long-range temporal dependencies with greater accuracy. DWT reduces input volatility and filters noise, which improves the model's signal-to-noise ratio, speeds up convergence, and enhances generalization performance across various environmental conditions.

3.7.1 Performance Contrast Between HDWN and Contemporary Models

Provide a clear comparison of HDWN with 2–3 latest models from the literature survey regarding architecture, pre-processing, and prediction robustness.

Table 1.2. Performance Contrast Between HDWN and Contemporary Models

S.No	Model	Limitation	HDWN Resolution
1	CNN	Ignores temporal dynamics; sensitive to noise	DWT filters noise; LSTM captures temporal dependencies
2	LSTM	Struggles with non-stationary signals	DWT decomposes signals into more stationary subbands
3	TS-CNN (2023)	Noisy data degrades performance	Pre-DWT denoising improves CNN's feature extraction
4	ARIMA (2023)	Poor non-linear pattern learning; high bias	HDWN captures non-linear trends via LSTM and DWT's multi-resolution input
5	ANN-LSTM (2022)	Overfitting due to lack of preprocessing and signal control	DWT smooths input variation, improving generalization and preventing overfitting
6	FTN (2025)	Depends heavily on sensor quality and lacks inherent noise filtering	DWT pre-processing mitigates sensor noise and improves signal robustness
7	Rustam et al. (2022)	Generalizability concerns with ANN architecture	HDWN's hybrid design balances feature richness and sequence modelling

4. Results and Discussion

The water samples of different physical features are taken as input and the nature of water based on chemical factors are evaluated and every pH, salt, hydrates, sulphates are analysed and are tested and trail values are excavated from various predictions and finally the best quality of water samples with low salt, pH values are taken to get best water as the output of purification.

ACCEPTED MANUSCRIPT

The water quality value is multiplied by the factors affecting the water and then divided by the water factors to determine the water quality index. The given variation shows the quality of water from number of variables, including turbidity, pH, dissolved oxygen, and nutrient levels. After being gathered from diverse locations, these samples are tested and purified to provide a dataset with varying degrees of water quality metrics. This connection assists in the visual representation of the heat map of water quality parameters, which shows variations in water quality over various locations and time periods visually. Each parameter's value is correlated with a particular color intensity

The samples taken are send to various stages when the input is passed to several hidden layer the unwanted information is not considered for evaluation are left and the reaming considerable factors are sent to the other level of predictions and each state holds the message that re considerable and several un wanted details are removed while working with long short term memory model.

ACCEPTED MANUSCRIPT

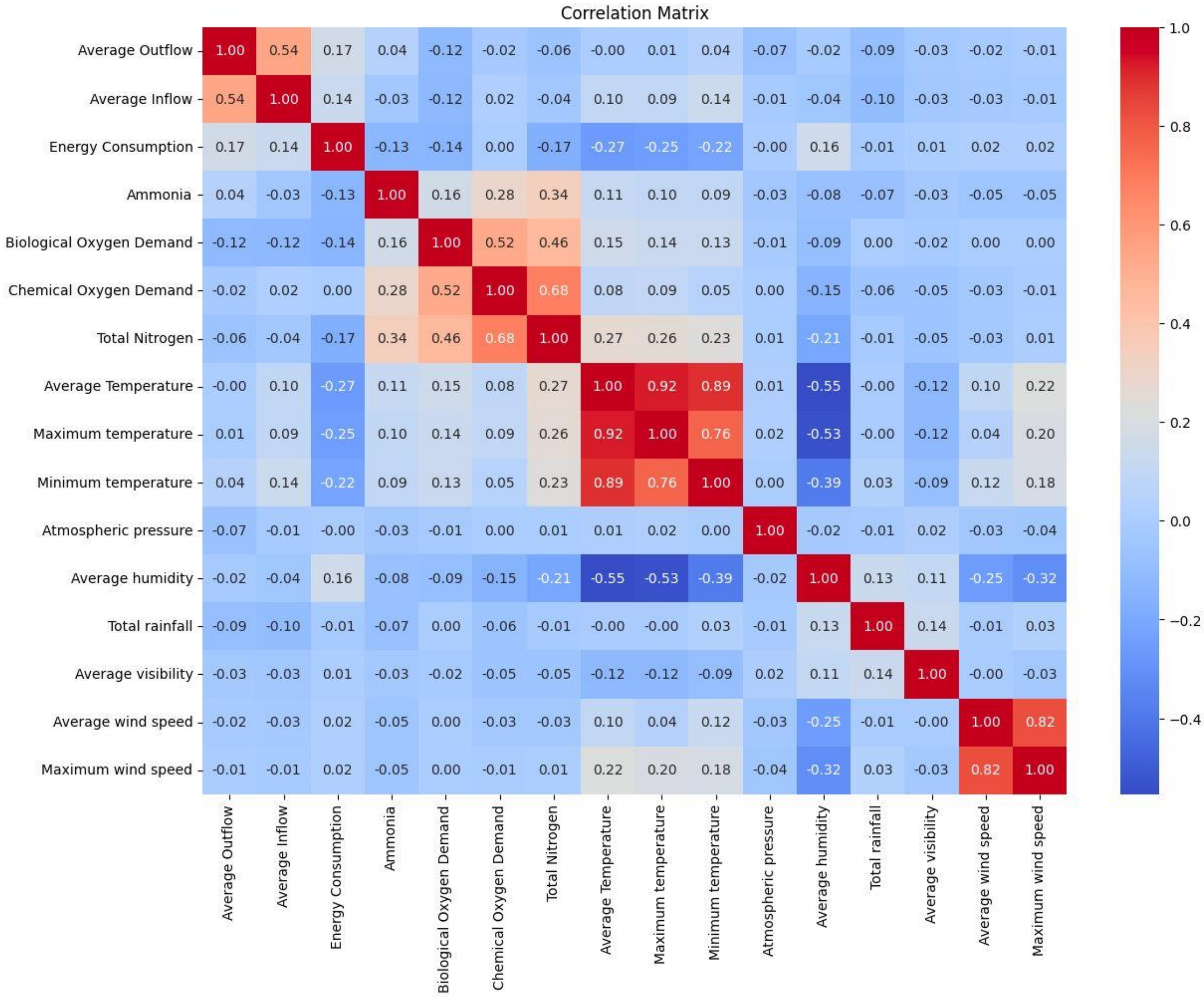
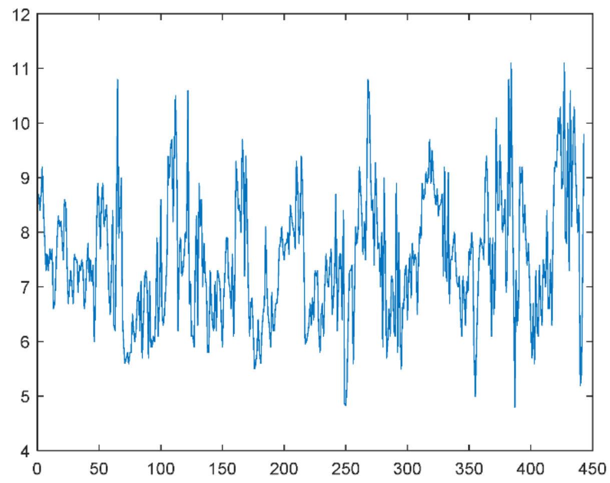


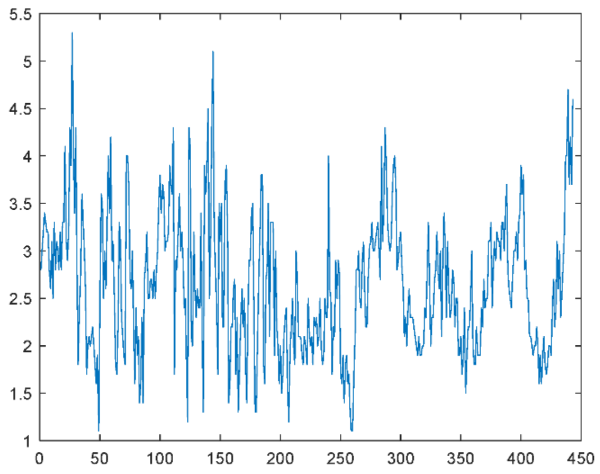
Table 4.1.1 Heat map diagram of various parameters considered for the sample water predicted.

ACCEPTED

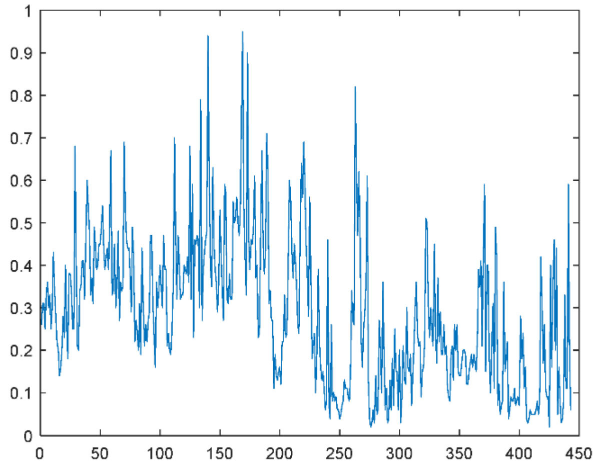
The graph below shows the variation of several prediction algorithm used based on many water samples taken and the predicted. The outcome gives the best accuracy of the samples evaluated. The LSTM with conventional model or based on wavelet transform gives better results when compared to other ANN models of neural working methodologies. Then chemical on demand, the pH values, the salinity, the hardness, solid wastes, organic carbonates also determine the water quality affecting factors and the purification level should be high for the best water samples to evolve and makes the water quality and quantity better compared with other methods of testing and training. The below Figure 4.1.2 shows the original date values from the water samples.



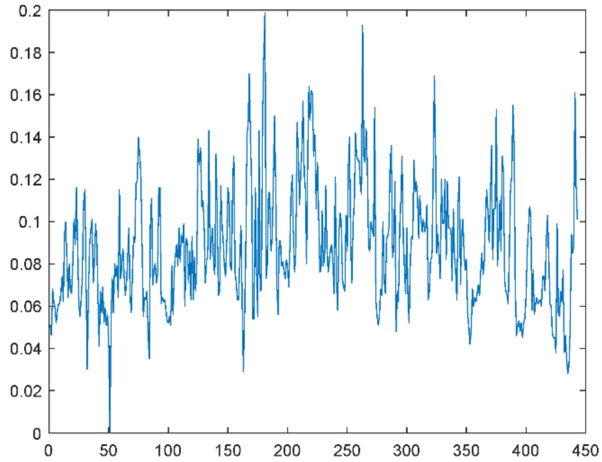
(a)



(b)



(c)



(d)

Figure 4.1.2 Sample Data set (a) Shows the pH level measured with the water sample (b) the COD of the water level (c) shows the BOD in water sample (d) The Total Nitrogen level in the water sample.

The conventional model was built to evaluate the water quality by machine learning methods of different formulations with support of regressive variables and co-efficiency considered for better practices.

Predictive circumstances are affected by temperature variations in oxygen concentration and chemical composition during the intervals between water sample collection. The sample mean formula is used to measure training performance, and the index shows the graphical yield after taking into account different parameters. Both long and short memory techniques are used to

ACCEPTE

generate the variables, and to improve results, convolutional support and wavelet transformation-based LSTM analysis are used. Comprehensive measures are taken to evaluate the prediction model's accuracy across large training and testing samples.

It is essential for preserving valuable water for sustainable use in the face of diminishing water resources. Current variations in water quality and quantity, seen in regions such as Tamil Nadu, India, and many other places across the world, show different times of need. Analyzing oxygen demand from evapotranspiration is made easier by using a neural network such as LSTM with convolutional analysis; wavelet transform is used to identify compounds in water samples and forecast carbon need. This predictive technique is strengthened by large datasets covering multiple parameters, which allows for more precise predictions and is reinforced by purifying procedures.

4.1 Prediction of Chemical parameter of Water Quality

The DWT-CNN-LSTM model prediction of various parameters from the data sample collected. The data consist of Ph, COD, BOD, TN, Ammonia etc. There samples are pre-processed and given the model for prediction. The below fig from 4.1.a to Fig 4.1.e shown the prediction accuracy level of the data which is given to the model.

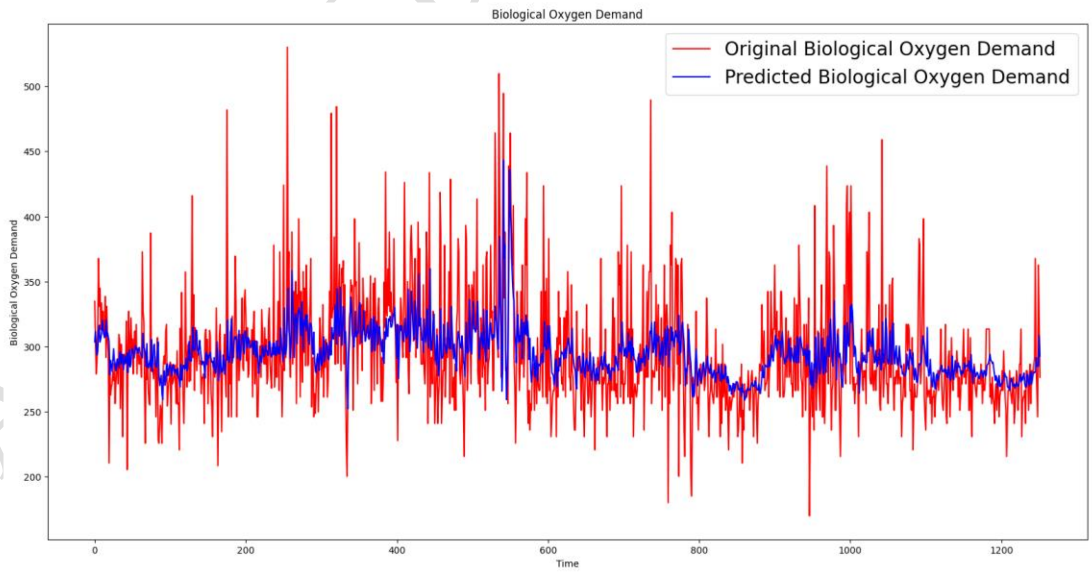


Fig : 4.1 a Actual Vs Predicted BOD

Fig 4.1 a show the precited values for BOD from the given data set original BOD for the last 50 days. The prediction accuracy is much better than the other algorithms. Fig 4.1.b Show the

ACCEPTED MANUSCRIPT

predicted value of COD from the given data set. Fig 4.1 c & Fig 4.1.d are the prediction graph of Ammonia and Total Nitrogen.

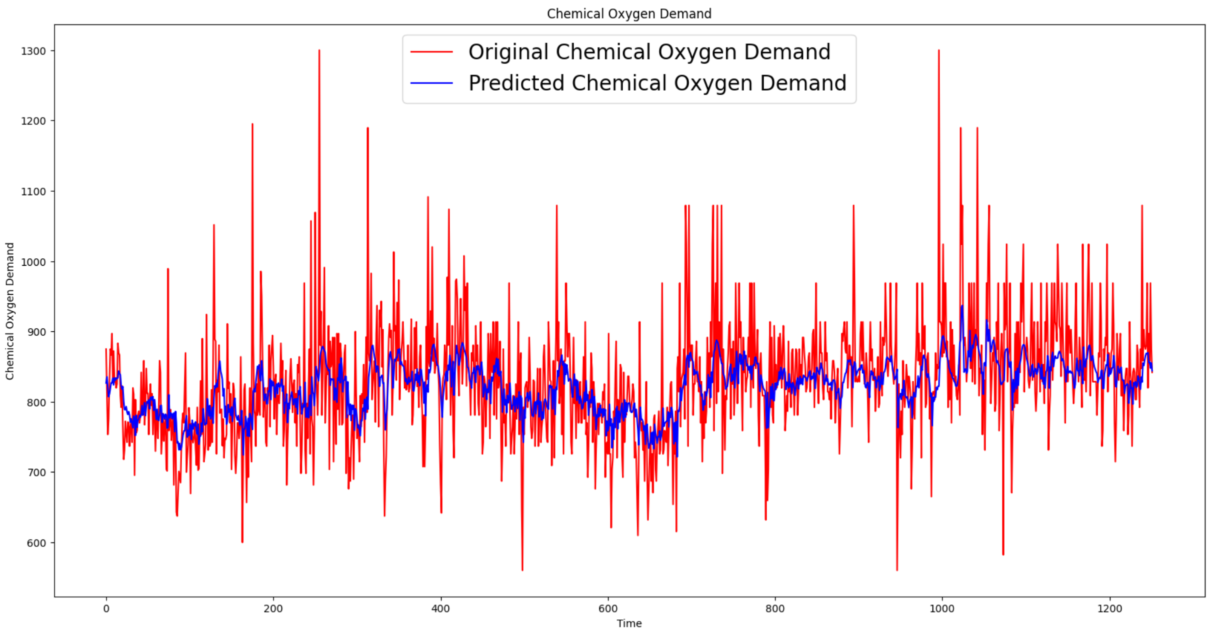


Fig : 4.1 b Actual Vs Predicted COD

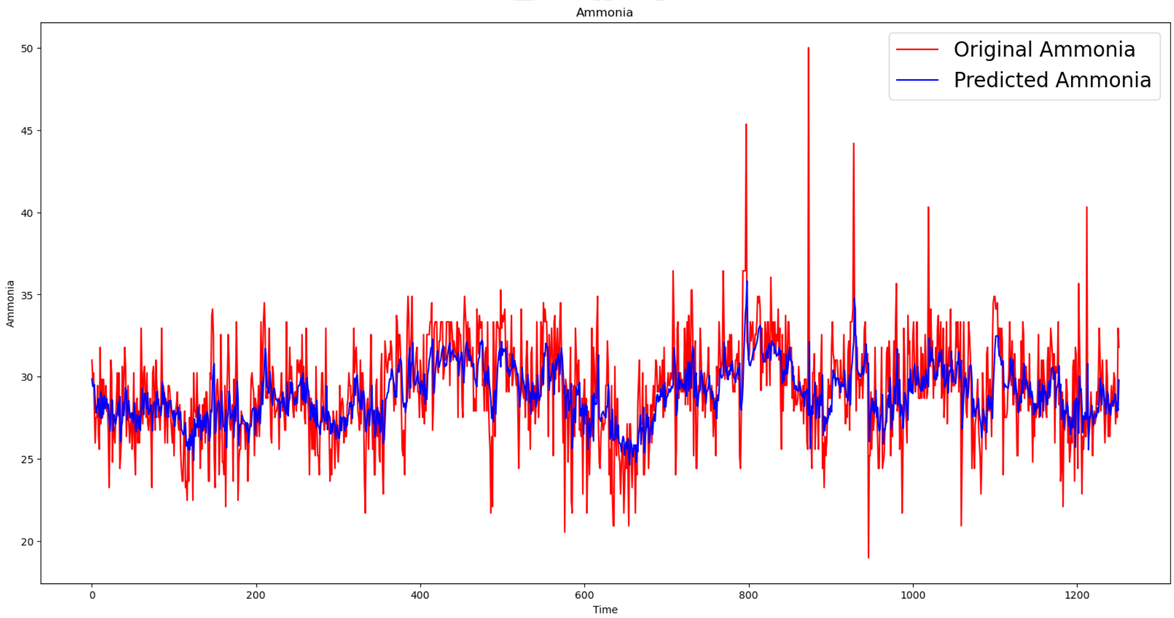


FIG: 4.1 c Actual Vs Predicted Ammonia

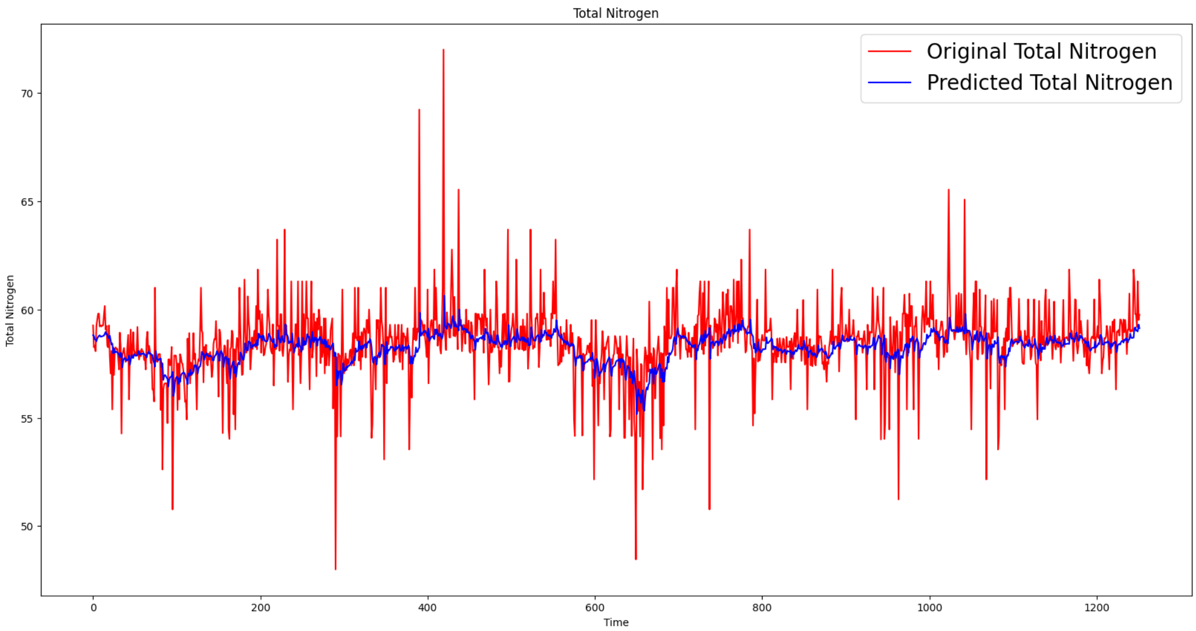


Fig : 4.1. d Actual Vs Predicted Total Nitrogen

4.2 Prediction of Physical parameter of Water Quality

Wastewater can have physical attributes that are quantifiable and visible without changing the water's chemical composition. These measurements contain vital information regarding the general state and properties of the water. In our model the physical parameter temperature of the water sample can be consider for the prediction of Waste Water treatment plant process efficiency. Fig 4.1e shows the predicted values of temperature. Temperature has an impact on a number of biological and chemical

wastewater treatment processes and may be a sign of thermal pollution or industrial discharges. It may cause the misprediction if the water temperature is more.

ACCEPTED MANUSCRIPT

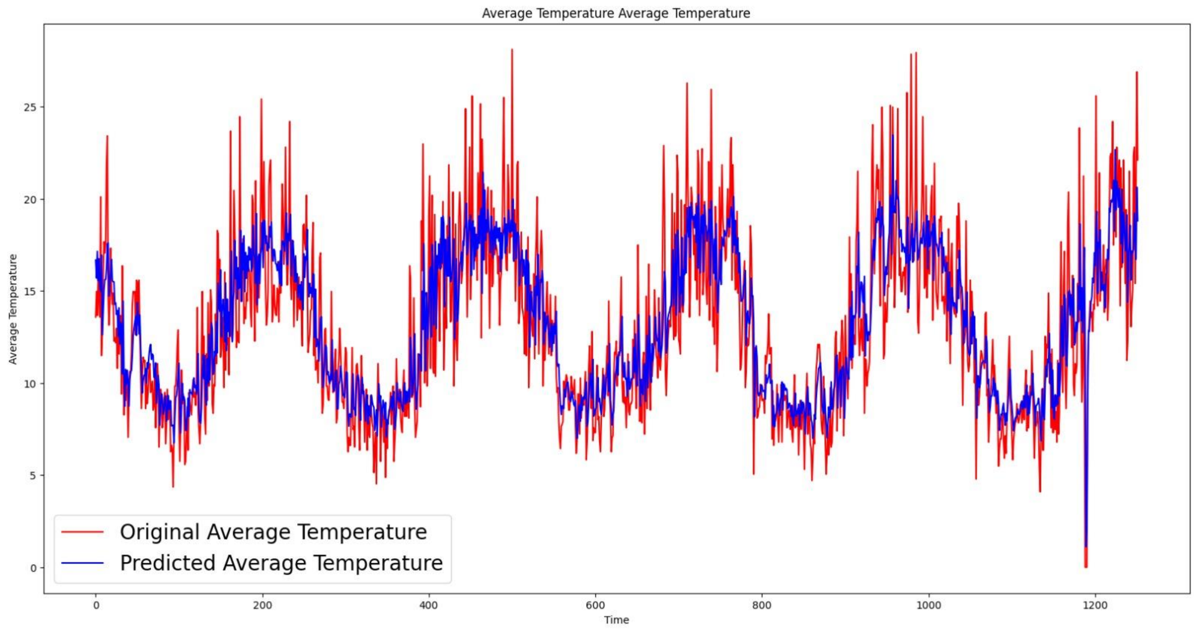


Fig: 4.1 e Actual Vs Predicted Temperature

4.3 Prediction of Energy Consumption of WWTP

The necessity for several physical, electrical, and thermal processes involved in wastewater treatment results in a considerable energy consumption in wastewater treatment facilities (WWTPs). WWTPs can improve overall sustainability, minimize operational costs, and lessen their environmental impact. Fig 4.2.1 Shows the energy consumption prediction of Waste water treatment. From the prediction we can regulate the sustainability, overall operational cost in terms of all resource's usages. It can help to reduce the overall impact of environmental pollutants.

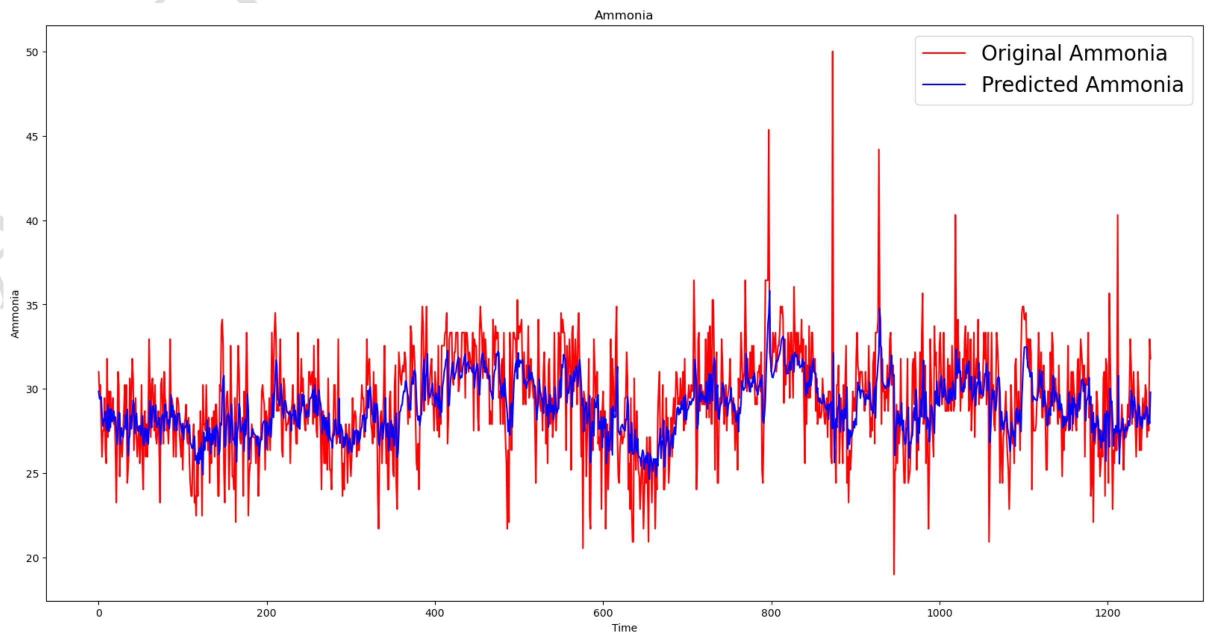


Fig : 4.2.1 Actual Vs Predicted Energy Consumption

This process is repeated iteratively and many predictions are made based on very important factors considering the accuracy as 98.8 percentage while handling large or small amount of dataset in real.

4.2. Comparison HDWN with other models of ML

The comparative graph was generated with several modes in which convolution-based LSTM model shows better results on calculating the mean sample valve predicted when compared with the n samples of water from several regions of hot, humid, rainy, moderate climate where the rate of pressure and evapotranspiration varies from time to time [43]. The mean rate of 110 is obtained for CNN-LSTM model and long and short-term model with 95 mean rate and artificial model with 80 and persistent forecast of 70 and so on as the output of the several predictions analysed form the given data set.

Parameters / Model	**MSE**	**RMSE**	**MAPE**	**MAE**	**R^2**
HDWN	0.068	0.2325	0.55	0.1945	0.932
FTN	0.3251	0.273	0.43	0.179	0.6749
TS-CNN	0.5612	0.23	0.36	0.1798	0.4388
CNN-LSTM	0.6515	0.2405	0.344	0.18975	0.3485
ANN -LSTM	0.7745	0.64	0.87625	0.42275	0.2255
ARIMA	1.03525	0.7395	0.29075	0.2595	0.0101

Table 4.2.1 Comparison of various Model Evaluation using metrics MSE, RMSE, MAPE MAE and R^2.

Table 4.2.1 show the model prediction accuracy by the model evaluation. The Model Performance Evaluation is based on the metric MSE, RMSE, MAP MAE and R^2. Our proposed model achieved 98.8% accuracy than other models.

ACCEPTED MANUSCRIPT

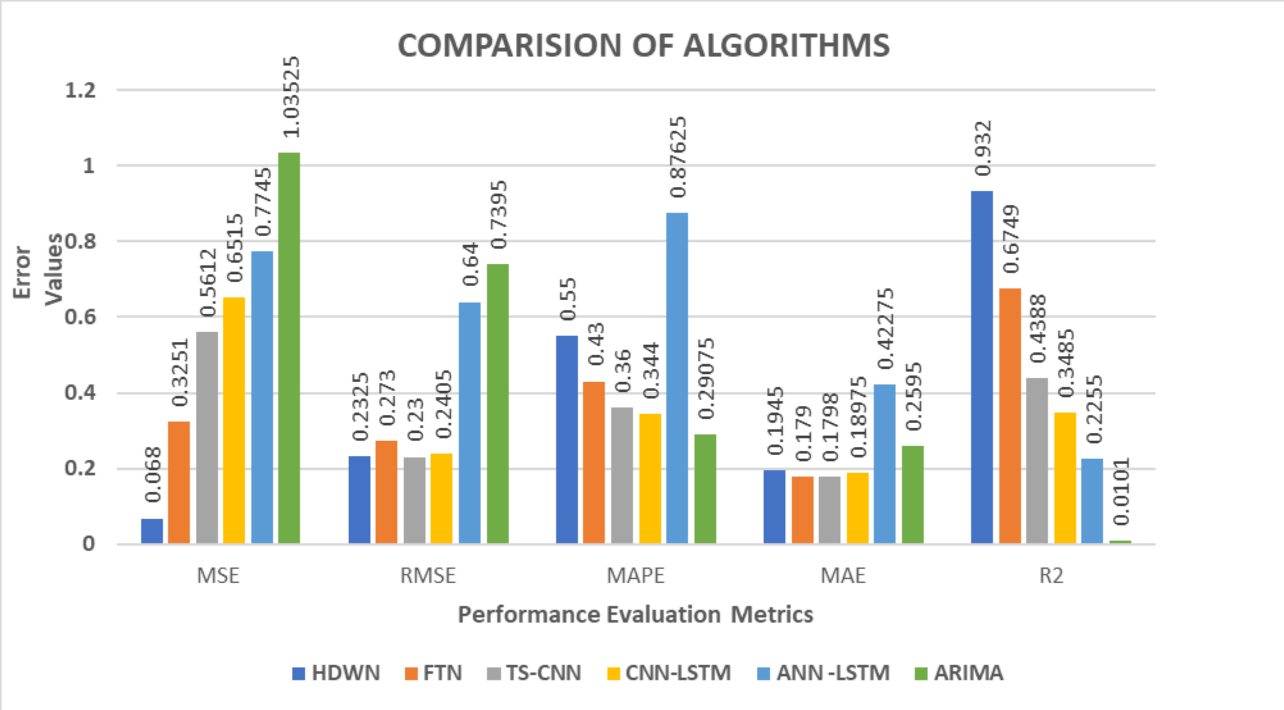


Figure 4.2.1 Comparison of Proposed Model Evaluation with other algorithms.

In our model we achieved better performance that can be shown as a graph. In Figure 4.2.1 Comparison of the Proposed Model Evaluation with other algorithms is given. Where MSE is 1.03, RMSE is 0.74, MAPE is 0.87 , MAE is 0.42 and R^2 is 0.71 the higher error rate by other algorithm but by our proposed method is very less than the other model MSE is 0.068, RMSE is 0.23, MAPE is 0.55 ,MAE is 0.19 and higher R^2 0.94. Our model performance is increased than the other model. The various water samples are collected as input and those data collections are sent to the convolution layer and the predictions are memorised in pools and iteratively it is tested and trained several times based on LSTM with convolution and finally all the predicted samples are summed together to get the mean optimal samples with utmost trustful data output.

The comparative study analyses many best models of prediction and this research can use ARIMA model, wavelet-based Fourier model along with LSTM model for many levels of predictions and can handle larger volumes of data collection.

4.2.1 Innovation in Comparison with Existing Techniques:

ACCEPTED MANUSCRIPT

This work's innovation is the HDWN model, which effectively extracts both spatial and temporal information from water quality data by combining the Discrete Wavelet Transform (DWT) with a hybrid CNN-LSTM architecture. The suggested model has improved predictive accuracy and robustness for coastal water quality forecasting, as evidenced by its much-reduced error rates (MSE: 0.068, RMSE: 0.2325, MAPE: 0.55%, MAE: 0.1945, R^2: 0.94) in comparison to existing models like FTN, TS-CNN, and ARIMA. Our method differs from earlier research because it combines deep sequential learning with wavelet-based preprocessing.

Water quality prediction is enhanced by the HDWN model's integration of DWT, CNN, and LSTM, which combines their respective advantages. By breaking down and eliminating noise from the data, DWT assists in identifying significant frequency components. While LSTM learns long-term temporal connections necessary for time-series forecasting, CNN extracts significant spatial characteristics from these modified signals. In comparison to conventional models, this integration improves accuracy, resilience, and generalization by enabling the model to handle complicated, noisy, and non-linear water quality data.

5. Conclusion and Future work

This research emphases on prediction of water quality based on various samples from different area with several models based on LSTM with CNN or wavelet transform series as proposed, the water sample collected from various regions of south Tamil Nadu as sample input object and monitoring done based on pressure, humidity, evapotranspiration and effluent water treatment is done to remove the unwanted wastes from the water and made it useful for varied purpose both commercial and domestic

(1) The fluid nature of the water with different ranges trained and tested with many neural models gives effective accuracy and this could be enhanced with better outcome with different samples of water from other districts of Tamil Nadu based on soil topography

(2) The predicted samples with chemical oxygen demand are evaluated with models of convolution based long and short-term memory network and also based on wavelet transform neural models

and many correlated variables are analysed and predictions made with high level of accuracy. This research could be improved based on many factors that determine the quality of water both physical and chemical parameters.

(3) The LSTM with CNN along with ARIMA, ANN and Persistent models are compared and better outputs are given with better predicted quality of water when compared with other models such as ANN. This research considers small dataset of water taken from a defined area as input and predictions are made based on quality factors. This can be improved by extending the various classification and prediction algorithm in future. Also, the management of water quality if done to control the effects of humans as pollutions and wastage.

ACKNOWLEDGEMENT:

This research is purely based on my own idea and working and the methodology used for water purification is ozone and chlorine-based purification and the different samples of water are collected from several area of south Tamil Nadu. The physical condition determines the water loss and purification accuracy and the dataset are trained and tested based on neural algorithms and the comparative output are analysed for better efficacy.

ACCEPTED MANUSC

References

[1] U. Ahmed, R. Mumtaz, H. Anwar, A. A. Shah, and J. García-Nieto, "Efficient Water Quality Prediction Using Supervised Machine Learning," Water, vol. 11, no. 11, p. 2210, 2019, doi: 10.3390/w11112210.

[2] A. A. Chaoui, W. Elkhoumsi, M. Laaouan, R. Bourziza, and K. Sebari, "Application of ARIMA Imputation Model and RNN Forecasting Model – WWTPs Performances Optimization," Environment and Ecology Research, vol. 11, no. 3, pp. 456–466, 2023, doi: 10.13189/eer.2023.110305.

[3] R. Álvarez-Ruiz and Y. Picó, "Analysis of emerging and related pollutants in aquatic biota," Trends in Environmental Analytical Chemistry, vol. 25, p. e00082, 2020.

[4] M. Gollapalli, "Ensemble Machine Learning Model to Predict the Waterborne Syndrome," *Algorithms*, vol. 15, no. 3, art. no. 93, Mar. 2022, doi: 10.3390/a15030093.

[5] M. G. Uddin, S. Nash, M. T. M. Diganta, A. Rahman, and A. I. Olbert, "Robust machine learning algorithms for predicting coastal water quality index," Journal of Environmental Management, vol. 321, art. no. 115923, Nov. 2022, doi: 10.1016/j.jenvman.2022.115923.

[6] C. R. Holkar, A. J. Jadhav, D. V. Pinjari, N. M. Mahamuni, and A. B. Pandit, "A critical review on textile wastewater treatments: Possible approaches," Journal of Environmental Management, vol. 182, pp. 351–366, 2016.

[7] Geetha, S., Gouthami, S. Internet of things enabled real time water quality monitoring system. Smart Water 2, 1 (2016). https://doi.org/10.1186/s40713-017-0005-y.

[8] H. Asami, M. Golabi, and M. Albaji, "Simulation of the biochemical and chemical oxygen demand and total suspended solids in wastewater treatment plants: Data-mining approach," Journal of Cleaner Production, vol. 296, art. no. 126533, 2021, doi: 10.1016/j.jclepro.2021.126533.

[9] P. Guedes-Alonso et al., "Pharmaceutical and personal care product residues in a macrophyte pond-constructed wetland treating wastewater from a university campus: Presence, removal and ecological risk assessment," Science of The Total Environment, vol. 703, p. 135596, 2020.

[10] Y. Kangabam, S. D. Bhoominathan, S. Kanagaraj, and M. Govindaraju, "Development of a water quality index (WQI) for the Loktak Lake in India," *Applied Water Science*, vol. 7, no. 6, pp. 2907–2918, Oct. 2017, doi: 10.1007/s13201-017-0579-4.

[11] J. Jiang, S. Tang, D. Han, G. Fu, D. Solomatine, and Y. Zheng, "A comprehensive review on the design and optimization of surface water quality monitoring networks," *Environmental Modelling & Software*, vol. 132, art. no. 104792, Oct. 2020, doi: 10.1016/j.envsoft.2020.104792.

ACCEPTED MANUSCRIPT

[12] Y. Li, B. Kong, W. Yu, and X. Zhu, "An Attention-Based CNN-LSTM Method for Effluent Wastewater Quality Prediction," Applied Sciences, vol. 13, no. 12, p. 7011, 2023, doi: 10.3390/app13127011.

[13] H. Hu et al., "A water quality prediction method based on the deep LSTM network considering correlation in smart mariculture," Sensors, vol. 19, no. 6, p. 1420, 2019, doi: 10.3390/s19061420.

[14] P. Liu, J. Wang, A. Sangaiah, Y. Xie, and X. Yin, "Analysis and Prediction of Water Quality Using LSTM Deep Neural Networks in IoT Environment," Sustainability, vol. 11, no. 7, p. 2058, 2019, doi: 10.3390/su11072058.

[15] J. P. A. Rani, R. Sujeethra, and C. Y. Rubavathi, "Prediction of Water Quality using Artificial Neural Network," in Proc. 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2022, pp. 992–999, doi: 10.1109/ICSSIT53264.2022.9716346.

[16] A. E. Maxwell, T. A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: An applied review," *International Journal of Remote Sensing*, vol. 39, no. 9, pp. 2784–2817, 2018, doi: 10.1080/01431161.2018.1433343.

[17] A. N. Prasad, K. A. Mamun, F. R. Islam, and H. Haqva, "Smart Water Quality Monitoring System," in *Proc. 2nd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*, Fiji Islands, 2015, pp. 1–6, doi: 10.1109/APWCCSE.2015.7476234.

[18] S. Sakizadeh, "Artificial intelligence for the prediction of water quality index in groundwater systems," *Modeling Earth Systems and Environment*, vol. 2, no. 1, pp. 1–9, 2016, doi: 10.1007/s40808-015-0063-9.

[19] P. Singh and P. D. Kaur, "Review on Data Mining Techniques for Prediction of Water Quality," *International Journal of Advanced Research in Computer Science*, vol. 8, no. 5, pp. 396–401, May–June 2017.

ACCEPTED MANUSCRIPT

[20] R. Tayeb, M. R. Chellali, A. Hamou, and S. Debbah, "Impact of urban and industrial effluents on the coastal marine environment in Oran, Algeria," Marine Pollution Bulletin, vol. 98, pp. 281–288, 2015.

[21] J. Zhang, B. Yu, G. Yan, H. Zhu, X. Y. Xu, and L. Zhu, "Seasonal bacterial community succession in four typical wastewater treatment plants: correlations between core microbes and process performance," Scientific Reports, vol. 8, no. 1, p. 4566, 2018, doi: 10.1038/s41598-018-22683-1.

[22] A. H. Mohammed, I. A. Hameed, and R. Seidu, "Machine learning: based detection of water contamination in water distribution systems," in *Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO '18 Companion)*, Kyoto, Japan, July 2018, pp. 1664–1671, doi: 10.1145/3205651.3208235

[23] S. Okeyo, N. Nontongana, T. O. Fadare, and A. I. Okoh, "Vibrio species in wastewater final effluents and receiving watershed in South Africa: implications for public health," International Journal of Environmental Research and Public Health, vol. 15, no. 6, p. 1266, 2018.

[24J. Park, K. T. Kim, and W. H. Lee, "Recent advances in information and communications technology (ICT) and sensor technology for monitoring water quality," *Water*, vol. 12, no. 2, art. no. 510, Feb. 2020, doi: 10.3390/w12020510.

[25] N. A. Lakshmi, N. Samyuktha, and J. P. A. Rani, "A novel system for air pollution prediction using supervised machine learning algorithm," *Int. J. Recent Trends Eng. Res. (IJRTER)*, Special Issue 7, pp. 147–151, Mar. 2020, doi: 10.23883/IJRTER.CONF.20200315.025.GIEIA.

[26] P. Guvava, B. J. Abiodun, and F. Otieno, "Projecting drought characteristics over East African basins at specific global warming levels," Atmospheric Research, vol. 228, pp. 41–54, 2019.

[27] A. Ziadi, N. T. Hariga, and J. Tarhouni, "Mineralization and pollution sources in the coastal aquifer of Lebna, Cap Bon, Tunisia," Journal of African Earth Sciences, vol. 151, pp. 391–402, 2019.

ACCEPTED MANUSCRIPT

[28] Y. Zhou, L. Wang, Y. Zhou, and X. Mao, "Eutrophication control strategies for highly anthropogenic influenced coastal waters," Science of The Total Environment, vol. 705, p. 135760, 2020.

[29] R. Diaz and R. Rosenberg, "Spreading dead zones and consequences for marine marine ecosystems," Science, vol. 321, pp. 926–929, 2008.

[30] C. Chaukura et al., "Contemporary issues on the occurrence and removal of disinfection byproducts in drinking water—A review," Journal of Environmental Chemical Engineering, vol. 8, p. 103659, 2020.

[31] K. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A Review," Sensors, vol. 18, pp. 1123–1134, 2018.

[32] A. Arepalli, K. Naik, and J. Amgoth, "An IoT Based Water Quality Classification Framework for Aqua-Ponds Through Water and Environmental Variables Using CGTFN Model," International Journal of Environmental Research, vol. 18, no. 4, p. 73, 2024.

[33] P. G. Arepalli and K. Naik, "Water quality classification framework for IoT-enabled aquaculture ponds using deep learning based flexible temporal network model," Earth Science Informatics, vol. 18, no. 2, p. 351, 2025.

[34] P. G. Arepalli and J. Khetavath, "An IoT framework for quality analysis of aquatic water data using time-series convolutional neural network," Environmental Science and Pollution Research, vol. 30, no. 60, pp. 125275–125294, 2023.

[35] S. Geetha, S. Gouthami, "Internet of Things Enabled Real Time Water Quality Monitoring System," 2017, pp. 2690-2714.

[36] A. K. Ashwini, D. Diviya, J. Janice Vedha, and M. Deva Priya, "Intelligent model for predicting water quality," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, Paper ID: IJRASET44658, ISSN: 2321-9653, June 21, 2022.

ACCEPTED MANUSCRIPT

[37] M. Kumar Jha, R. Kumari Sah, M. S. Rashmitha, R. Sinha, and B. Sujatha, "Smart Water Monitoring System for Real-Time Water Quality and Usage Monitoring," 2018.

[38] N. Kumar Koditala and P. Shekar Pandey, "Water Quality Monitoring System Using IoT and Machine Learning," 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE), San Salvador, El Salvador, 2018, pp. 1-5, doi: 10.1109/RICE.2018.8509050.

[39] A. Hamzeh Haghiabi, A. Heidar Nasrolahi, and A. Parsaie, "Water quality prediction using machine learning methods," 2018.

ACCEPTED MANUSCRIPT