A Prediction Model Based Energy Efficient Data Collection for Wireless Sensor Networks

[1]Balakumar D and [2]Rangaraj J
[1]Department of ECE, Annamalai University, Chidambaram, Tamil Nadu, India.
[2]Department of ECE (Deputed to GCT Coimbatore), Annamalai University, Tamil Nadu India.
[1]dbalakumarr@gmail.com, [2]srd_jsrd@yahoo.co.in

Correspondence should be addressed to Balakumar D : dbalakumarr@gmail.com.

Article Info
Journal of Machine and Computing (http://anapub.co.ke/journals/jmc/jmc.html)
Doi: https://doi.org/10.53759/7669/jmc202303031
Received 28 January 2023; Revised from 18 May 2023; Accepted 16 June 2023.
Available online 05 October 2023.

Abstract – Many real-time applications make use of advanced wireless sensor networks (WSNs). Because of the limited memory, power limits, narrow communication bandwidth, and low processing units of wireless sensor nodes (SNs), WSNs suffer severe resource constraints. Data prediction algorithms in WSNs have become crucial for reducing redundant data transmission and extending the network's longevity. Redundancy can be decreased using proper machine learning (ML) techniques while the data aggregation process operates. Researchers persist in searching for effective modelling strategies and algorithms to help generate efficient and acceptable data aggregation methodologies from preexisting WSN models. This work proposes an energy-efficient Adaptive Seagull Optimization Algorithm (ASOA) protocol for selecting the best cluster head (CH). An extreme learning machine (ELM) is employed to select the data corresponding to each node as a way to generate a tree to cluster sensor data. The Dual Graph Convolutional Network (DGCN) is an analytical method that predicts future trends using time series data. Data clustering and aggregation are employed for each cluster head to efficiently perform sample data prediction across WSNs, primarily to minimize the processing overhead caused by the prediction algorithm. Simulation findings suggest that the presented method is practical and efficient regarding reliability, data reduction, and power usage. The results demonstrate that the suggested data collection approach surpasses the existing Least Mean Square (LMS), Periodic Data Prediction Algorithm (P-PDA), and Combined Data Prediction Model (CDPM) methods significantly. The proposed DGCN method has a transmission suppression rate of 92.68%, a difference of 22.33%, 16.69%, and 12.54% compared to the current methods (i.e., LMS, P-PDA, and CDPM).

Keywords – Wireless Sensor Networks, Adaptive Seagull Optimization Algorithm, Extreme Learning Machines, Dual Graph Convolutional Network.

I. INTRODUCTION

WSNs are independent sensing devices that are spatially distributed and monitor physical or environmental parameters. It has numerous uses, such as disaster management, traffic control in smart cities, and environmental monitoring. Power utilization and stability are significant issues and crucial challenges in WSNs due to sensor node batteries' limited capacity and frequent battery replacement impracticality. The most vital component influencing power consumption is data extraction and transmission of packets. This is primarily due to the nodes' mandate to obtain all sensor readings continuously and precisely. These nodes require much power throughout each stage of data extraction, accumulation, and transmission [1].

As the number of sensor nodes (SN) in WSNs grows, several issues emerge, such as excessive power consumption, long network transmission delays, poor transmission quality caused by data transmission congestion, and data transmission blocking caused by partial node failure. Data loss/abnormality is a common occurrence in WSNs [2-5], as the financial requirements of sensor nodes can often lead to node/link failures. Data prediction is an analytical technique for dealing with these challenges [6], in which predictive actions are performed utilizing historical data acquired by sensors. There is no requirement for continuous transmission of data measured by sensor nodes while utilizing this technique [7]. **Fig 1** shows the WSN network model example.

These programmes might be practical in a manner that aids us in reflecting on our surroundings and their flaws. In terms of data prediction, the purpose is to create an algorithm that minimizes transfer rate and power consumption and maximizes sensor network durability. By this, enough data is gathered to group the present nodes in the configuration and a neural network node is utilized to find a suitable CH [8]. Time series forecasting approaches have been developed to capitalize on temporal correlation in WSNs. Each node in these systems uses a predictor to execute prediction operations based on prior sensor readings. The predictor prohibits sensor readings that deviate from the expected reading by a margin below a predetermined error limit from being transmitted.

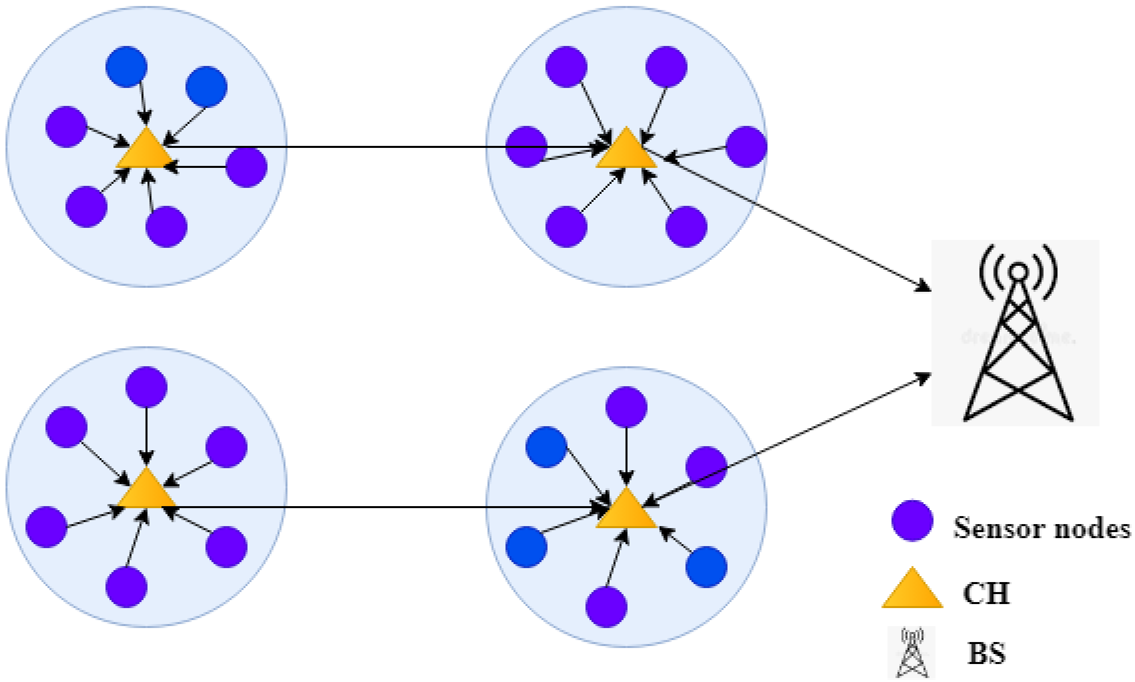

Fig 1. Network Model Example

To the greatest extent of the authors' knowledge, the time series of sensor readings is considered to be stationary and linear in all WSN approaches, and a linear time series approach (including ARIMA, ARMA, or AR) is employed for forecasting time series. However, the time series created by SN amid the process of detecting ambient conditions is typically non-stationary and nonlinear, and linear predictors cannot accurately predict non-stationary and nonlinear time series [9-10]. Simple strategies are employed in several previous studies, such as [11, 12], to create predictors for sensor networks to send data from the complete sensor array to the base station. Conversely, if there are considerable and constant changes in data values, the forecasting methods used in these works may need to be revised. Local prediction utilizing sensor network clustering may be an effective technique for addressing this issue.

Because the shortest routing path is used to send sensor data, a local predictive algorithm is cost-effective. Still, cluster-based local prediction confronts various problems. The primary issue is related to the high cost of training predictors, which is influenced by the disparity between computation and communication. Another area for improvement is sensor data's dynamic nature, primarily when predictive models perform poorly on inconsistent datasets.

The energy minimization is accomplished in two phases in this research effort, so data prediction is performed in the first stage, and statistical predictive modelling is performed in the second stage. The Wireless Sensor Network uses data prediction to forecast future data from all active nodes. An aggregator is a detection node in a network that gathers data and transfers it to the other nodes. Rather than delivering all acquired data upon processing it employs proper data prediction approaches, aggregator nodes only communicate the required quantity of data. Furthermore, substantial data reduction is accomplished in the second step by identifying neighboring nodes that generate data on a regular basis employing a statistical data forecasting model.

The primary purpose of proposed data forecasting techniques for cluster based WSNs is to minimize radio transmission power consumption by minimizing the transmissions between the transmitter and the receiver. To accomplish this, researchers must execute compelling data sampling predictions across wireless sensor networks during clustering and aggregating data at every cluster head to save overhead.

The critical issues also prompted us to choose cluster-based wireless sensor networks in our study. This work's major contributions are outlined below:

- This work developed a data prediction technique using ASOA, ELM, and DGCN approaches to minimize needless data transmission and power usage. The approach leverages a small number of sensor nodes for prediction-based data collecting and processing within the cluster.
- The suggested method consists of two procedures: choosing the CH and data prediction. The nearest Neighbor Node is used to construct the cluster. ASOA is used to select the CH. The fitness is determined in the CH selection procedure employing RER and distance.
- In the proposed framework, ELM filters data associated with each node to generate a tree to cluster the data collected by the sensors.
- The dual graph convolutional network (DGCN) is an analytical technique that predicts future patterns using time series data.
- The paper presents a MATLAB simulation using a practical demonstration of the suggested approach to measuring the transmission of data packets and utilization of energy in sensor nodes in networks with varying numbers of distributed sensor nodes.

The rest of this work is structured in the following order: Section II discusses the associated work. Section III discusses various preparation approaches as well as the essential modelling process. Section IV compares the performance of the suggested technique to that of other relevant techniques, and Section V concludes the work.

II. LITERATURE SURVEY

Syed Ahmed Suleiman et al. (2020) suggested a hybrid approach using autoregressive integral moving average (ARIMA), Kalman filter (KF) decision tree (DT), and decision tree (DT) approaches to anticipate the data sampling demands of SNs to decrease needless data aggregation. Data clustering and aggregation are employed for each cluster head to efficiently perform sample data prediction across WSNs, primarily to minimize the processing overhead caused by the prediction approach. The model's performance was assessed at several epochs and with varying numbers of nodes. Based on experimental results, the proposed method surpasses existing relevant methods regarding the accuracy of predictions and energy consumption [13].

GC Jagan et al. (2022) suggested a three-phase architecture for an efficient data collection method, with the phases being modified LEACH, Bi-LSTM, ELM, and adaptive Kalman filter (KF). The outcomes of this investigation outperform existing methods. The findings show that the suggested data aggregation approach surpasses the existing IDAD2DC, EEDP, SDNAELWA, and READP techniques [14].

Wang Haibin et al. (2021) created a reliable dual predictive data reduction strategy for wireless sensor networks. This method reduces data in the Data Prediction Phase (DPP) and the Data Reduction Phase (DRP). DRP's primary goal is to decrease the transmissions between SNs and sink nodes, thus reducing power usage. Additionally, it detects and discards erroneous data at sensor nodes. According to simulation results, the suggested technique is efficient and successful regarding data reliability, data reduction, and power utilization [15].

Sathyapriya Loganathan et al. (2020) suggested an energy-efficient self-diagnosing clustering technique for WSNs. Rather than choosing CHs at random, an initial cluster head is chosen, and then clusters are formed to achieve more excellent performance than previous approaches. The suggested approach determines cluster heads depending on the weighted metrics of SNs. The sensor nodes then adapt automatically by rendering the right judgements in real time, utilising the observed data. However, the detected information is frequently erroneous owing to mechanical, wireless, and battery difficulties. Compared to LEACH, the suggested approach approximately doubles the lifetime, is twice as excellent as QLEACH and ECH, and is 51% better than the ad-hoc method [16].

Khushboo Jain et al. (2022) investigated a combinational data prediction model (CDPM) capable of generating previous data to regulate latency and predicting upcoming data to decrease unwanted data transmission. A genuine WSN-based programme is simulated by employing a real dataset to test the effectiveness of our suggested CDPM data prediction algorithm. The effectiveness of proposed method is also compared to that of the HLMS, ELR, and P-PDA algorithms. The findings reveal that the suggested CDPM outperforms a single forward or backward method regarding data transmission reduction, power efficiency improvement, and latency regulation [17].

Jainism et al. (2020) suggested a two-vector model using the ECR approach for synchronizing prediction data in transit within a cluster in order to minimize cumulative mistakes in continuous data forecasts. During the data collecting cycle's initialization step, it develops an approximation of future data and determines its prediction error. ECR is a primary, unstructured, lightweight, and scalable data-predicting approach. It minimizes data transfer and battery utilization while retaining precision, although it is difficult. [18].

Jainism et al. (2021) suggested an ELR approach that excludes the Sensor Node from broadcasting significant amounts of data for a predetermined time, during which the Base Station predicts upcoming data values, consequently reducing the

WSN's power usage. ELR is an energy-efficiency approach for reducing data transmission and prolonging the longevity of networks; however, it does not consider using cluster topologies, scalability, or control latency [19].

Al-Qurabat et al. (2020) suggested a DGAST technique that collects sensor information regularly and segments the network into rounds. Every DGAST round consists of four phases: data aggregation, data accumulation, selective transfer, and adjusting the frequency of samples acquired for SN. The suggested model saves electricity and increases the lifetime of sensor networks; however, it requires complex computations and a lot of memory [20].

Agarwal et al. (2021) suggested a Data prediction technique using linear regression (DP-LRM) model that eliminates redundant data transfer by establishing a model of regression of linear descriptors on sequentially observed data values along with constructing any data aggregation approach. It employs a buffer using a linear filtering technique to compare and correlate all incoming information.

DP-LRM is an energy-saving approach that effectively decreases data transmission costs while maintaining reliability and accuracy at low data volumes. However, it is analytically complicated and fails to consider flexibility [21].

Nelson et al. (2021) proposed combining HFBLMS and QKLMS to create HFQKLMS filters. The HFBLMS approach was created by incorporating FC theory with the HLMS method. For data aggregation, the prediction procedure employs the HFQKLMS filter technique. This technique is energy-efficient, preserves accuracy on tiny information, and increases network longevity, but it is analytically challenging and needs to consider scalability [22].

Wang et al. (2021) suggested a data reduction (DR) method employing the KF. The approach reduces data in two stages: data prediction and data reduction. This practical and successful DR method is dependable, energy-efficient, and increases system longevity. However, it has a significant computational overhead and ignores using cluster topology and scalability of the network [23].

Jain et al. (2021) presented Data transmission reduction method (DTRM) execution in Cluster Heads. This research disables temporal redundancy and data reading correlation, allowing SNs to send only a small number of data values, boosting data transmission efficiency and lowering power consumption. However, DTRM is based on single-value comparisons and delivers data correctness, decreased data transfer, low complexity cost, lightweight processing, restricted memory consumption, robustness, and efficiency [24].

Famila et al. (2021) hypothesised that RCHST-IETSMP uses a Hyper-Erlang procedure to combine two major energy-defining parameters and a confidence parameter in order to pick a Hyper-Erlang distribution using process-integrated Semi-Markoc predictions successfully. This approach is dependable and extends the lifetime of wireless sensor networks, but it is computationally complicated because it needs to consider scalability and control delay. [25]. **Table 1.** Analyses All of the Previously Stated Data Transmission Techniques in WSNs With Widely Recognized Features.

Table 1: WSN comparison of existing data transmission approaches

Reference	Techniques used	Contribution	Limitation
[18]	Extended cosine regression (ECR)	Prediction model that is simple and lightweight	Not compatible with the cluster-based structure and fails to control delay
[19]	ELR	Data transmission is reduced, and network life is extended.	It is not scalable
[20]	Data collection and aggregation combined with selective transmission	Conserve energy and extend the life of periodic networks	Complicated computations and extensive use of memory
[21]	DP-LRM	Retains accuracy and minimizes the transmission cost	The complexity of algorithms is high
[22]	HFQKLMS	Efficient use of energy, Excellent accuracy	Fails to control delay
[23]	Kalman filter-based data reduction method	Data transmission and consumption of energy are reduced	High computing overhead
[24]	DTRM	Minimal complexity costs, Reliable and efficient	Single value comparison
[25]	RCHST-IETSMP	WSN longevity is extended	There is no effective CH selection, and the algorithmic complexity is high

Major Contribution

Numerous techniques for minimizing data transfer in WSNs have been developed, although control latency has not been introduced yet. Compared to the approaches and techniques outlined above, the suggested algorithm has the advantage of regulating latency and minimizing power consumption by attaining significant data transmission denial and lower RMSE (better data quality).

III. PROPOSED METHODOLOGY

This work uses WSN to provide an adequate data aggregation technique for WSN operating machine utilization prediction. ASOA, ELM, and DGCN are utilized to anticipate the following three stages, as illustrated in **Fig 2**.

Network Model

The sensors are linked and distributed randomly and can shift dynamically from one spot to another. The sink is in the middle of the network. The suggested ASOA optimization algorithm selects the most suitable CH on the sink. CH can be joined by its nearest neighbor nodes. CH takes data from cluster neighbors and transmits aggregated data to the BS.

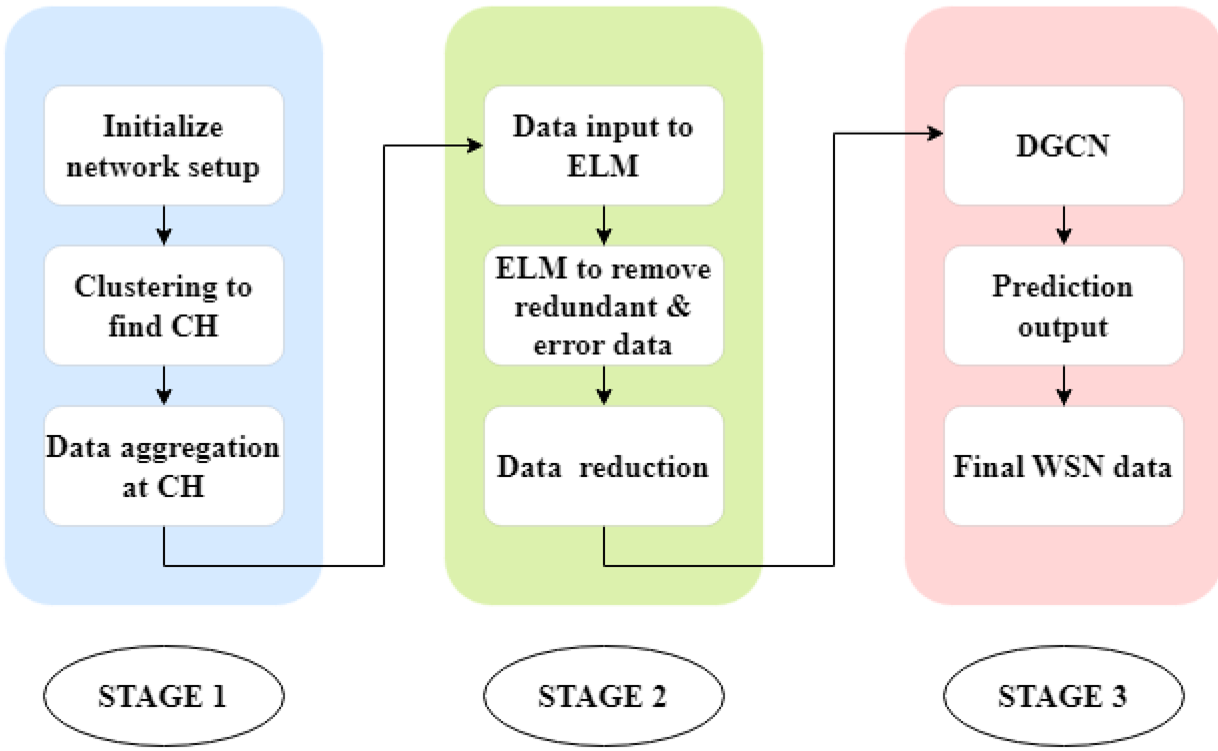

Fig 2. Stages of the Proposed Algorithm

Adaptive Seagull Optimization Algorithm Based Clustering Algorithm

Seagulls are a group of extremely smart creatures that live on the Planet and can be found in various shapes and sizes. Seagulls are migratory birds that travel in response to seasonal weather variations to find sufficient nourishment sources. The seagull optimization algorithm's fundamental idea is to imitate the seagulls migratory and attack behavior and to discover the ideal explanation by constantly modifying the position of the seagull.

Biological Characteristics

Gulls are the foremost familiar seabirds in many seaside cities; they typically live in clusters and use their brilliance to locate and attack the bait. Seagulls have two distinct behaviors: migratory behavior and aggressive behavior. As the term implies, migration refers to the transportation of animals from one location to another in a certain manner in response to changes in the weather to grab enough food resources to retain their vitality production. In the course of migration, gulls generally fly in groups, each individual bird flying at a different place along the trip to prevent collisions between seagulls. Each seagull in a group can change its location by travelling in the direction of the optimal location; simultaneously, the seagulls will execute the assault behavior necessary to obtain nutrition; while attacking, the seagulls will constantly move in a spiral motion, similar to flying [26].

Bio Mathematical Modeling

A random seagull is selected according to the biological conditions of the seagull population, and a corresponding mathematical representation is utilized to define, describe, and execute its migratory and migratory behavior.

Migration Behavior

The approach replicates the manner in which gull populations migrate from one place to another during migration. At this point, every seagull must meet the following conditions:

Collision Avoidance: To avert collisions with nearby seagulls, the technique estimates the updated location of the particular seagull employing an additional parameter X. The parameter X controls the avoidance of collisions between nearby seagulls.

$$\vec{N}_s = X \times \vec{p}_s(T) \tag{1}$$

$$X = f_c - (T \times (f_c/Max\_iteration)) \tag{2}$$

In Equ (1) and (2), $\vec{N}_s$ represents the updated location gained by the seagull activity, and the resulting location is unlikely to conflict with other seagull locations. The seagull's current position is represented by $\vec{p}_s(T)$. T is the current number of iterations of the algorithm, and $Max\_iteration$ represents the maximum number of iterations of the algorithm in the process. X refers to the additional metric stated to prevent collisions with nearby seagulls (other seagulls). f_c represents a hyper-parameter assigned by the technique, and its variable is fixed to 2, ensuring that the variable X declines linearly from 2 to 0 when the number of iterations T is repeated.

Migrate in the path of the optimum neighbor: Individual seagulls move toward their optimum neighbor to prevent collision with neighboring seagulls.

$$\vec{D}_s = Y \times (\vec{p}_{bs}(T) - \vec{p}_s(T) \tag{3}$$

$$Y = 2 \times X^2 \times Rn \tag{4}$$

In Equ (3) and (4), $\vec{D}_s$ represents the optimal seagull's direction and $\vec{p}_{bs}(T)$ represents the optimum neighbor's (other seagull's) location direction. Rn is a random number in the interval [0, 1], and it can interfere with the technique's execution by making independent random modifications within the interval of values selected, averting the technique from slipping into a local optimum throughout implementation. Y is another random number using the variable X, and the variable Y is utilized to determine the global or local search of the technique.

The optimum seagull position is approached as follows: After travelling to the optimum neighbor's direction, every seagull will shift to the globally optimal path and eventually arrive at an alternative location.

$$\vec{C}_s = \left| N_s + \vec{D}_s \right| \tag{5}$$

Where $\vec{C}_s$ incorporates the requirements for seagulls to prevent collisions and migrate to the optimal individual, which can be described as another updated location attained by the seagull and the spacing between the current seagull and the optimal seagull.

Attacking Behavior

Seagulls are very intelligent creatures that may use their past and experience to find food. Because of the necessity for long-term hunting and repeated attack behaviors during migration, seagulls rely on wings and weight to retain a high degree of stability when attacking prey, and the specific spiral movement behavior in flight is continually changing. Its altitude and angle of attack. When a seagull attacks, it moves in the 3D a, b, and c planes, as shown below.

$$a = R \times \cos(k) \tag{6}$$

$$b = R \times \sin(k) \tag{7}$$

$$c = R \times k \tag{8}$$

$$R = u \times e^{kv} \tag{9}$$

where R is the flight radius of the spiral when the seagull assaults, k is a random number in the range $[0, 2\pi]$, u and v are constants explaining the spiral shape, which are commonly defined as 1, and e refers as the base of natural logarithms.

$$\vec{p}_s(T) = \left(\vec{C}_s \times a \times b \times c\right) + \vec{p}_{bs}(T) \tag{10}$$

Above, $\vec{p}_s(T)$ is an updated phrase that determines the final seagull search position by combining migratory and assault behaviors. Flowchart of adaptive seagull optimization algorithm is shown in **Fig 3**.

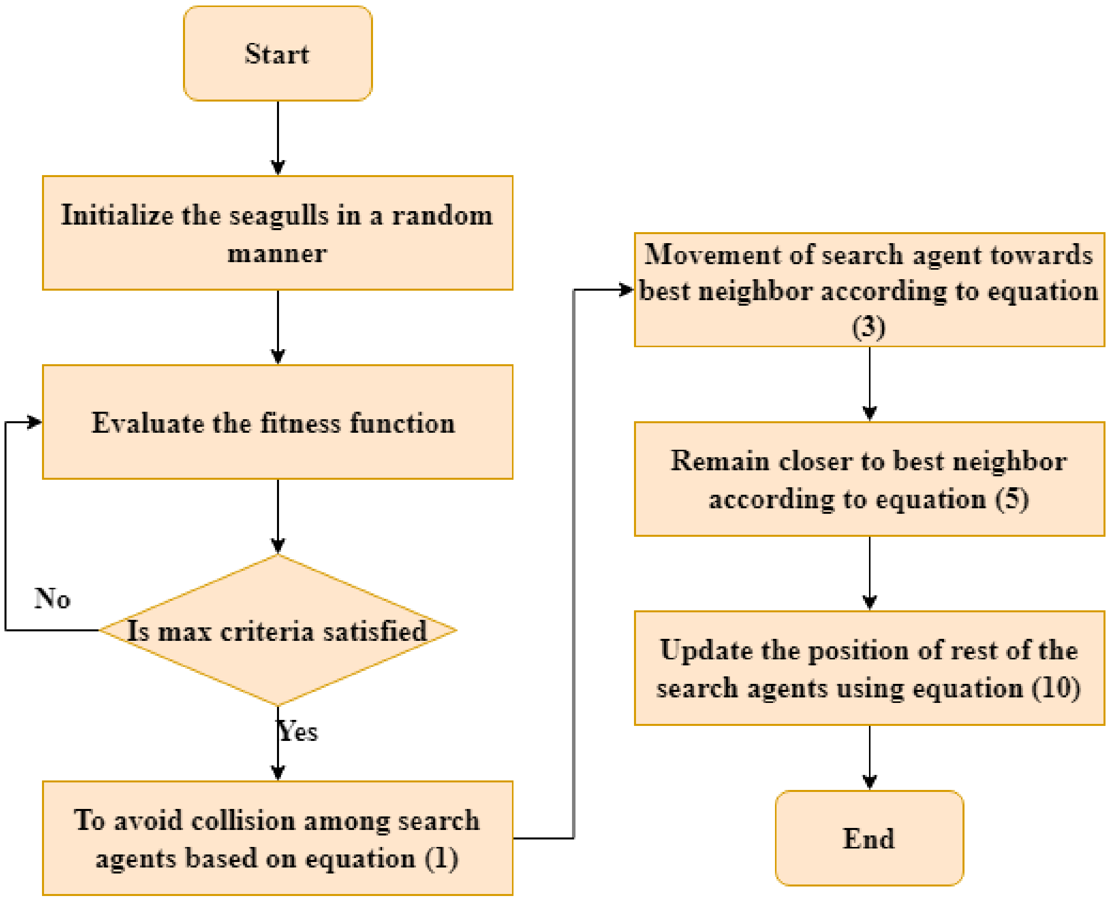

Fig 3. Flowchart of Adaptive Seagull Optimization Algorithm

Extreme Learning Machines (ELM)

The ASOA output is routed through extreme learning machines to remove redundant and error-prone information. The extreme learning machine is a feed-forward neural network with distinct learning stages, as shown in **Fig 4.** The projection stage is not trainable, and the input values are determined randomly. There is no need for iterative calculations. This feature shortens training model computation time; however, random bias and weight selection can result in unstable predictions. It is proposed to merge the Mahala Nobis distance-based radial basis function (MDRBF) with the ELM network to address ELM's inadequacies.

Dual Graph Convolutional Network (DGCN)

DGCN is built on a symmetrically interlinked network framework, and it receives a single raining input as X and predicts its non-raining variant Y. Network comprises two feature extraction layers, several basic units, and two reconstruction layers. The feature extraction layer uses a conventional 3 x 3 convolution operation to extract shallow features from the rain input. As illustrated in **Fig 5**, the two shallow traits are spread to deeper layers via interlink [27, 29].

The network is made up of various fundamental modules, including a multi-scale convolution with a dilation module and a global GCN module for capturing local and global input. To transport shallow features to deep layers, use symmetric layer skip connections. We build fundamental components based on the proposed module to assemble the proposed dual-graph convolutional network for predicting data. In particular, this work initially designed a spatial GCN module in each unit to acquire global spatial information from the preceding unit.

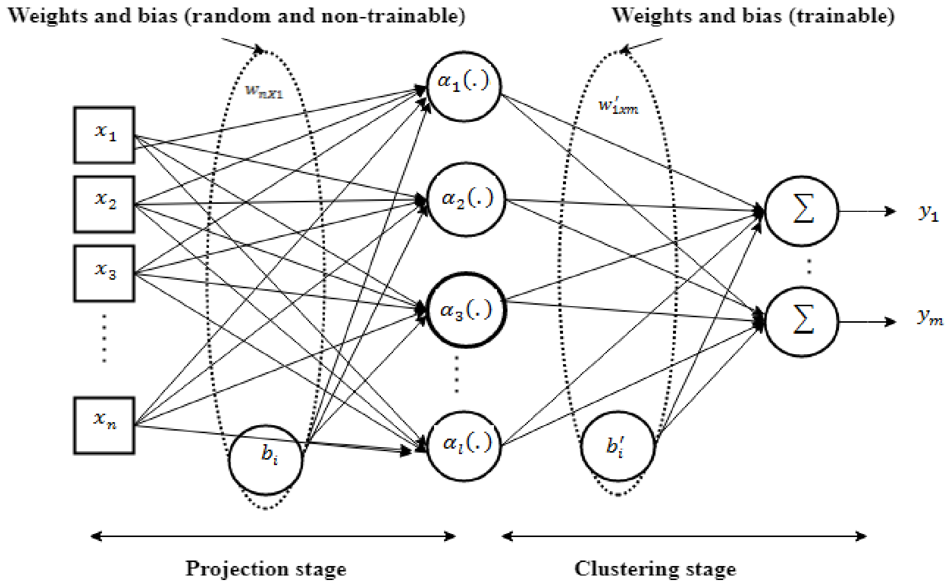

Fig 4. Structure of ELM Network

This global spatial information is then supplied into the convolution with a dilation module, which aids in the extraction of local spatial features from several scales. Finally, to get content information in addition to spatial information, the technique employs the Channel GCN module to investigate the association between features with rich global and local spatial illustrations. The flow of the proposed model's fundamental unit is described above:

$$f_{unit} = f_{in} + cGCN(DCM(sGCN(f_{in}))) \tag{11}$$

Where *DCM(•), cGCN(•),and sGCN(•)* represent the dilated convolutional module, channel GCN module, and spatial GCN module, respectively. This work uses 1x1 fundamental units to construct the de-raining network. Furthermore, symmetric skip connections are used in this work to link deep and shallow layers and prevent the gradient from fading.

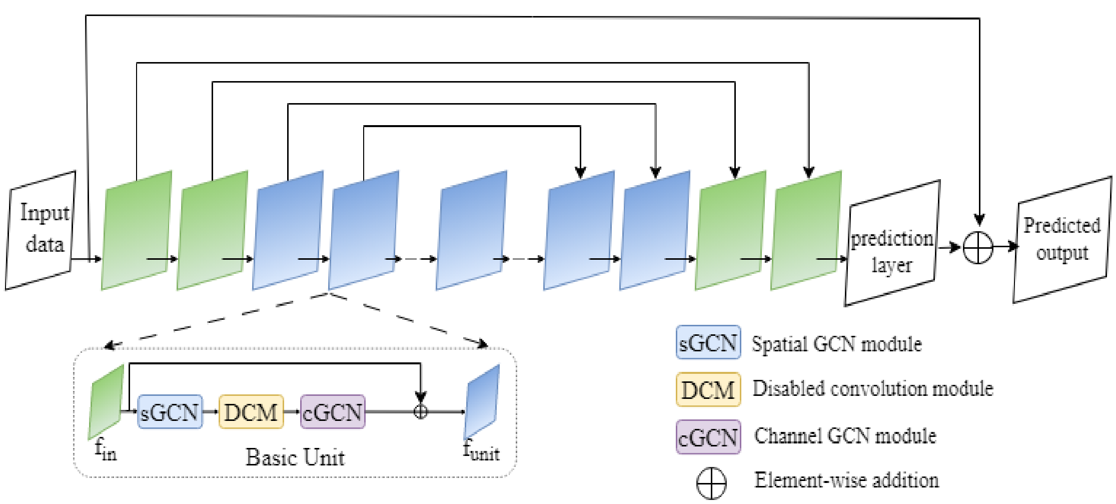

Fig 5. The Overall Architecture of Proposed Dual Graph Convolutional Network

Loss Function

Mean squared error (MSE) is the most commonly utilized loss function for training a neural network. However, because of its' penalty, MSE frequently produces over-smoothed results. To overcome this issue, the work uses the mean absolute error (MAE) to strike a compromise between data prediction and detail preservation.

$$L = \frac{1}{m} \sum_{i=1}^{m} \left\| y_i - y_{gt,i} \right\|_1, \tag{12}$$

Where m denotes the total number of training data, y_i represents the predicted data and y_{gt} represents the actual data. Activation functions process the input data in both the forward and backward layers, and the outcome is produced as an output.

IV. RESULTS AND DISCUSSION

A simulation based on the utilization of the Internet of Things is used to assess the suggested method in **Table 2**. To train the proposed model, the Annamalai University Next Generation Laboratory collects roughly 2,400 real-time sensor reading samples every day. Using publicly available sensor datasets, we chose a set of temperature and humidity measurements recorded every 31 seconds from our laboratory sensor implementation. The suggested DGCN's performance is also compared to the LMS [30], P-PDA [31], and CDPM [17] algorithms.

Table 2. Simulation Parameters

Parameter	Value
Initial energy (E_0)	1J
Sensor Nodes (N)	1500
Prediction Threshold ($^{\delta}$)	0.5,1.0
Data packet (D)	500 bytes
Network Area	1000mx1000m
Transmission energy (E_{TX})	150 nJ/s for 1- bit,10m
Simulation time interval (T)	150 s
Aggregation energy (E_{DA})	5(nJ/bit)/s
Reception energy ((E_{RX})	50nJ/s for 1-bit
Free space amplifier energy (ϵ_{fs})	10(pJ/bit)/m^2

*Transmission Suppression **(TS)***

Transmission suppression is a calculation of the ratio of sent data using any data prediction model to actual sensed data without applying any data prediction technique.

$$\text{TS\%} = \left(\frac{\text{Transmitted data by using prediction algorithm}}{\text{actual sensed data}}\right) \text{x } 100 \tag{13}$$

This work calculates the TS% of the four algorithms for the SN's mean temperature and mean humidity. The greater the TS%, the less data is transported and the lower the power consumption. The TS% of the four algorithms for the SN's mean temperature and mean humidity are depicted in **Fig 6 and 7**, respectively, and are shown in **Table 3**. The suggested DGCN TS% is constantly more excellent than the LMS-TS, P-PDA, and CDPM techniques' thresholds for any round of communication.

Table 3. TS% of SNs' Average Temperatures and Humidity

S.No	Threshold	Average temperature of SNs				Average Humidity of SNs			
		LMS	P-PDA	CDPM	Proposed	LMS	P-PDA	CDPM	Proposed
1	0.05	34.99	46.89	59.79	72.85	43.52	52.95	67.97	75.86
2	0.10	46.70	55.54	68.69	75.64	54.66	59.83	74.98	88.87

3	0.15	59.67	66.33	73.76	83.75	59.89	64.79	78.96	91.64
4	0.20	69.77	75.65	79.62	87.88	66.88	70.93	80.69	93.88
5	0.25	75.76	79.42	82.35	92.68	69.99	77.78	82.98	94.64

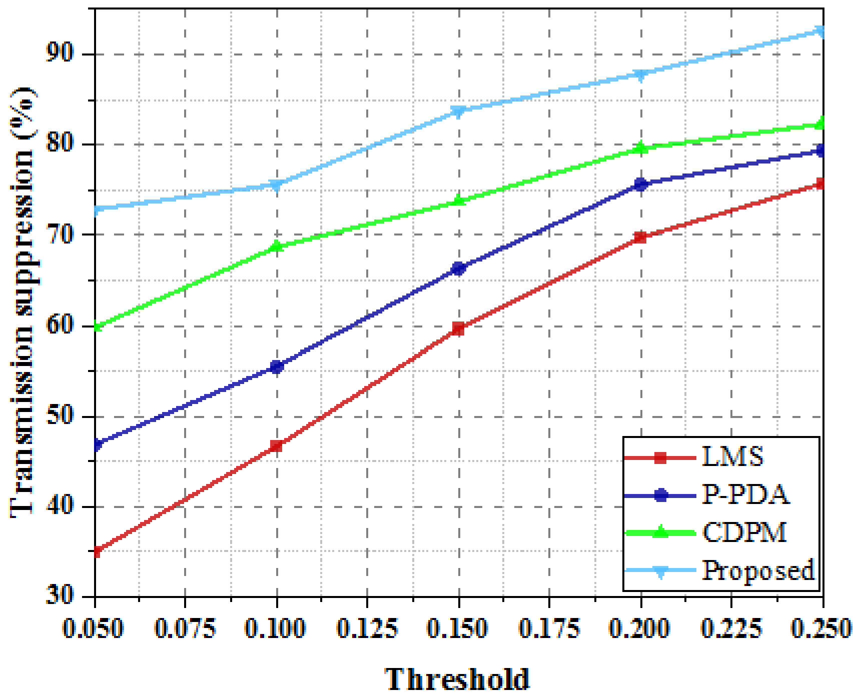

Fig 6. Transmission Suppression for Average Temperature

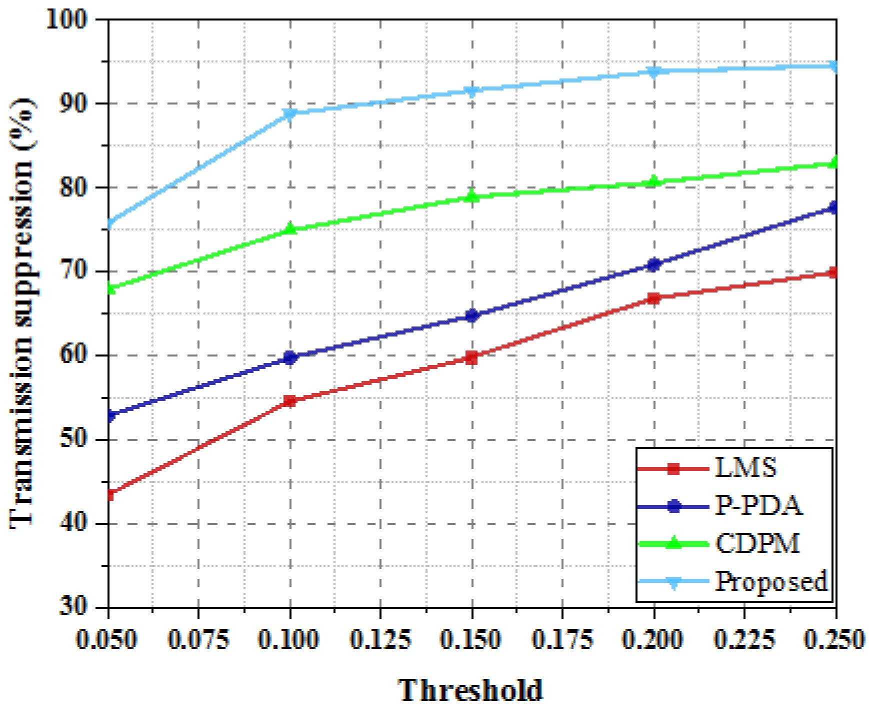

Fig 7. Transmission Suppression for Average Humidity

Energy Consumption

The quantity of energy utilized in a wireless sensor network is proportional to how much radio communication the SN does. Reducing the total data transferred to the Base Station will significantly extend the wireless sensors network's lifetime. The greater the TS, the less data is transmitted, and the less power is consumed in **Table 4.**

Table 4. Energy Consumption % of Average Temperatures and Average Humidity of SNs

S.No	Threshold	Average temperature of SNs				Average Humidity of SNs			
		LMS	P-PDA	CDPM	Proposed	LMS	P-PDA	CDPM	Proposed
1	0.05	32.56	22.34	13.32	9.10	11.20	7.05	5.36	4.10
2	0.10	33.53	20.62	11.25	7.01	11.31	7.10	5.31	3.65
3	0.15	31.16	18.34	10.40	6.34	10.15	6.34	4.07	3.02
4	0.20	27.17	16.21	9.34	5.02	9.26	6.00	4.43	2.01
5	0.25	21.52	14,35	7.33	5.61	8.14	5.75	3.67	1.67

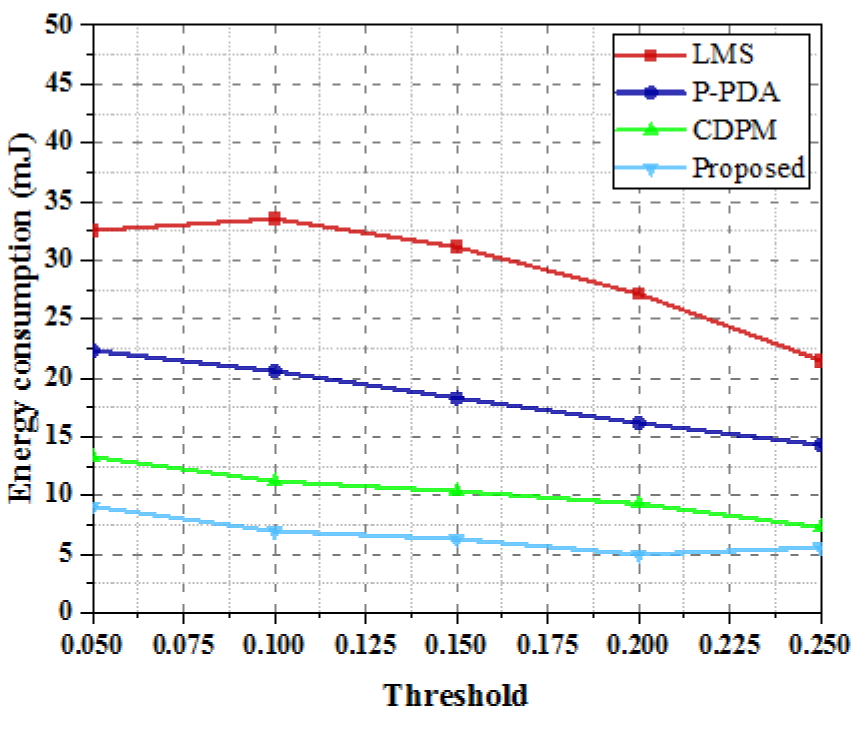

Fig 8. Energy Consumption for Average Temperature

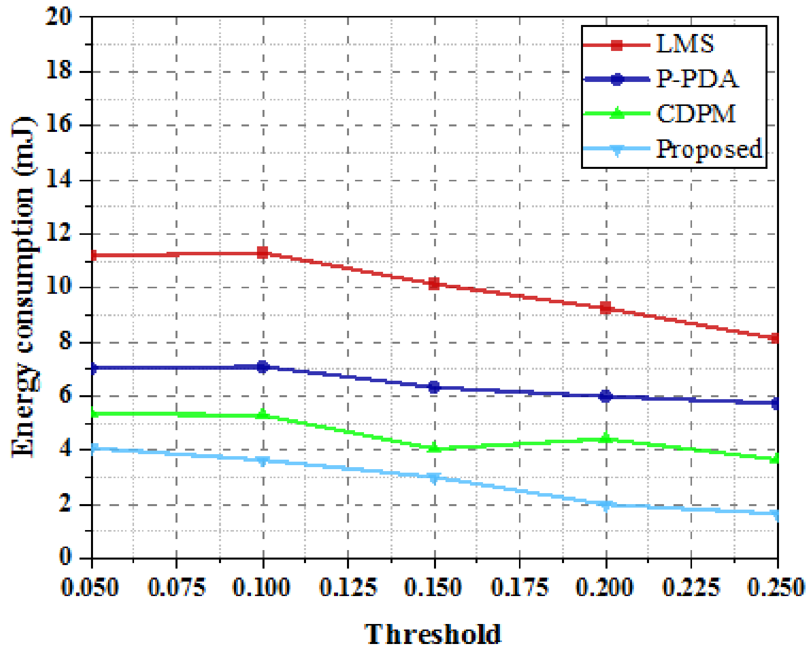

Fig 9. Energy Consumption for Average Humidity

Since most data prediction methods diminish data throughput, this work compares DGCN energy consumption to that of the LMS, P-PDA, and CDPM algorithms. DGCN is used in conjunction with these three algorithms to collect data from 10 rounds of communication, with every round having a configurable threshold ranging from 0.05 to 0.25 and a step function of 0.5. This work calculated the energy utilization using an energy model of the SN's mean temperature and humidity. **Fig 8 and 9** demonstrate the power usage of the four techniques for SN's average temperature and humidity.

Data Equality

Data quality is vital in determining excellence in the Wireless Sensor Network. This work previously discussed Root Mean Squared Error (RMSE) as a means to reduce the error of data perceived by any SN.

$$\text{RMSE}=\sqrt{\frac{1}{M}\sum_{i-1}^{M}(Ei)^2} \tag{14}$$

where $E_i = P_i$ -$\widehat{\text{P}}$i; Pi is the actual data of SN and $\widehat{\text{P}}$i is the predicted values of SN.

Table 5. RMSE % of Average Temperatures and Average Humidity of SNs

S.No	Threshold	Average temperature of SNs				Average Humidity of SNs			
		LMS	P-PDA	CDPM	Proposed	LMS	P-PDA	CDPM	Proposed
1	0.05	0.020	0.015	0.012	0.01	0.023	0.021	0.016	0.013
2	0.10	0.023	0.017	0.013	0.011	0.030	0.027	0.021	0.017
3	0.15	0.025	0.019	0.015	0.012	0.037	0.030	0.025	0.023
4	0.20	0.027	0.025	0.021	0.019	0.040	0.036	0.033	0.027
5	0.25	0.037	0.029	0.026	0.023	0.052	0.047	0.040	0.036

Fig 10 and 11 show that changing the threshold value (0.05 to 0.25) reduces the root mean squared error of the four humidity and temperature algorithms, resulting in greater data accuracy. However, the suggested approach's data accuracy is always superior because it has the lowest root mean squared error value of all thresholds. As a result, while maintaining high data accuracy, the suggested method has a more effective data suppression rate and improved power efficiency.

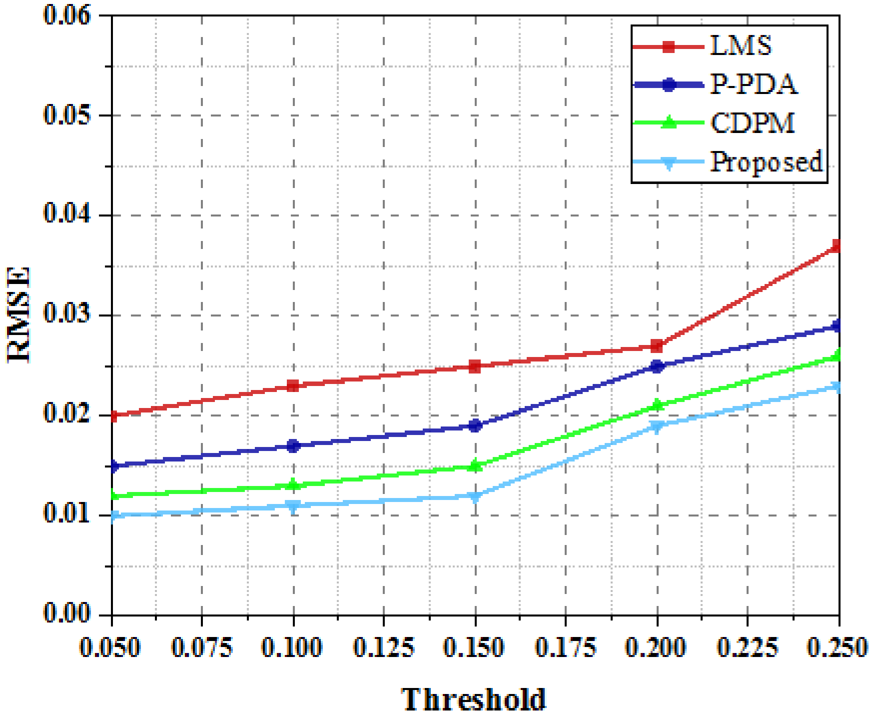

Fig 10. RMSE for Average Temperature

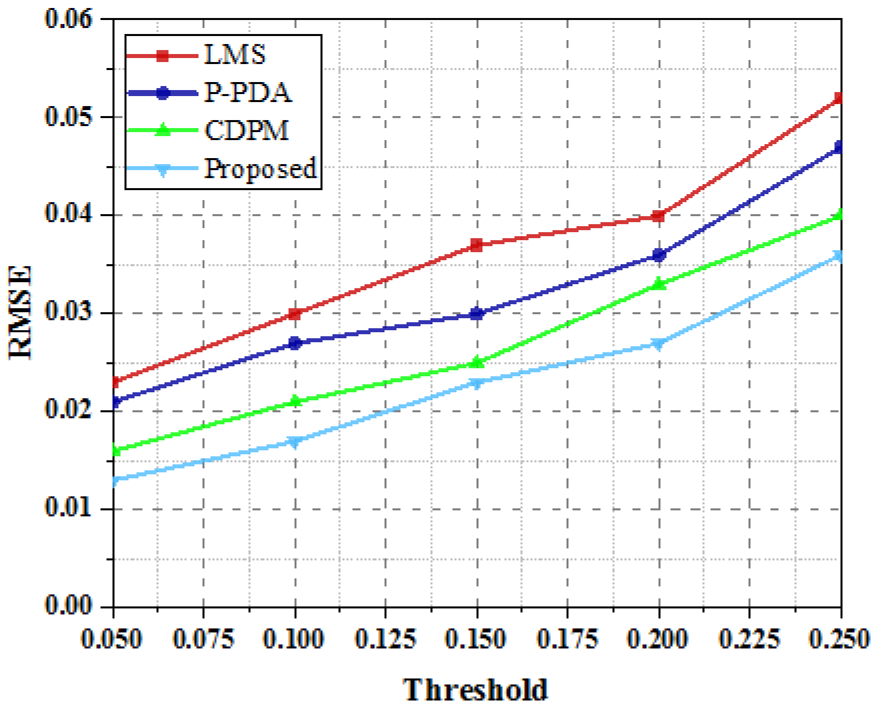

Fig 11. RMSE for Average Humidity

Positive Prediction

By adjusting the threshold and adding the successful prediction data supplied in every round, the overall number of successful predictions is computed, and the percentage of correct predictions is computed over various criteria in **Table 6.**

Table 6. Positive Predictions

S.No	Threshold	LMS	P-PDA	CDPM	Proposed
1	0.1	99.64	100.76	115.77	158.90
2	0.2	101.97	105.78	134.98	180.41
3	0.3	102.5	120.98	160.98	210.80
4	0.4	109.9	145.44	190.54	240.42
5	0.5	122.69	156.63	200.57	270.89

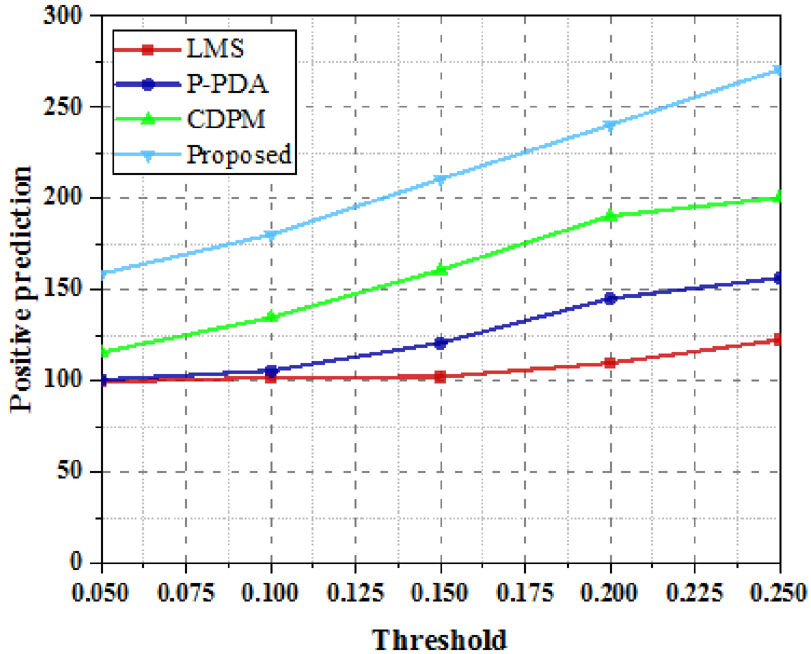

Fig 12. Evaluation of Positive Predictions

Table 6 shows that the proposed DGCN algorithm outperforms LMS, P-PDA, and CDPM regarding the number and percentage of positive predictions. The case of DGCNs with different thresholds, the total number of positive predictions is relatively high. As illustrated in **Fig 12,** the number of successful predictions nearly doubles compared to other existing algorithms. This ensures that the proposed algorithm is correct and reliable.

Number of Packets Transmission

Table 7. Comparison Between the Overall Packet and Node Counts

No.of nodes	Without prediction	With prediction			
		LMS	P-PDA	CDPM	Proposed
200	1.5	0.3	0.25	0.20	0.06
400	2.2	0.6	0.36	0.28	0.10
600	2.5	0.9	0.5	0.32	0.18
800	2.8	1.10	1.0	0.6	0.4
1000	3.0	1.58	1.35	1.10	0.7
1200	3.3	2.0	1.78	1.5	1.4
1400	3.9	2.5	2.32	1.8	1.56

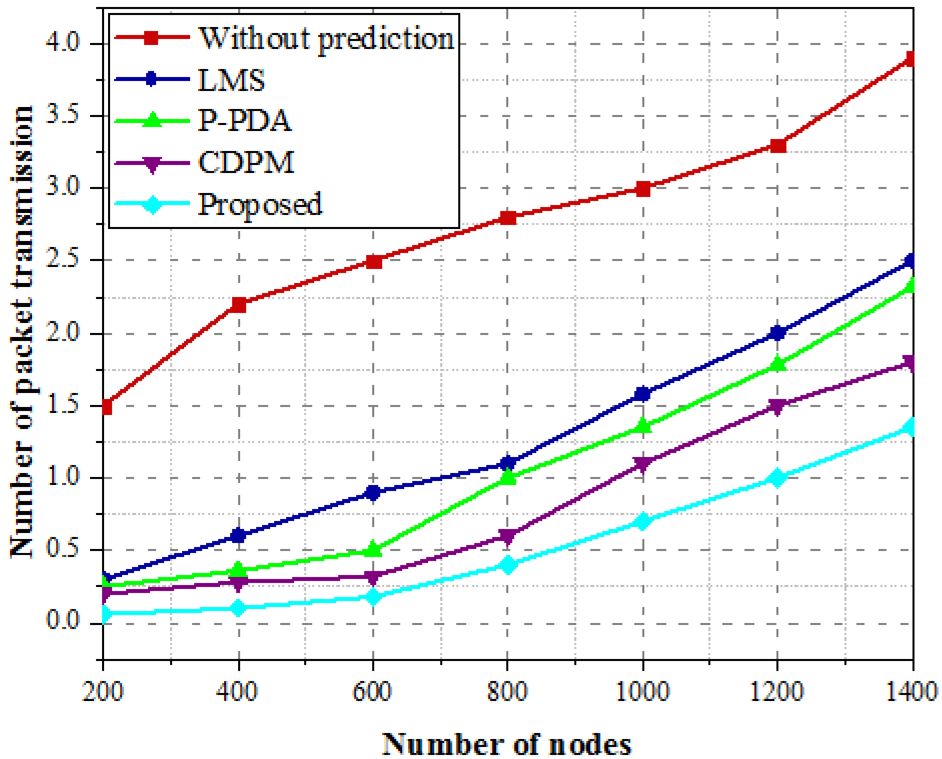

Fig 13. Evaluation of Number of Packet Transmission

Fig 13 depicts various approaches and the network's scalability for real-time data sets. The LMS method of transferring data to the recipient has a high communication cost. Furthermore, advanced computations, such as computationally expensive aggregations, complicate P-PDA. Our model is capable of predicting and updating local data. As a result, the proposed algorithm processes raw data at sensor nodes or intermediate nodes using data aggregation techniques, which reduces packet transmission and saves power while improving performance.

Performance Evaluation of All Algorithms

$$\mathrm{MAE}(\mathrm{x},\hat{x}) = \frac{1}{n}\sum_{i=1}^{n} \left| x_i - \hat{x}_i \right| \tag{15}$$

$$\mathrm{MAPE}(\mathrm{x},\hat{x}) = \frac{100\%}{n}\sum_{i=1}^{n} \left| \frac{x_i - \hat{x}_i}{x_i} \right| \tag{16}$$

$$R^2 = 1 - \frac{S_{residual}}{S_{total}} \tag{17}$$

$$SS_{residual} = \sum_i (x_i - \hat{x}_i)^2 \tag{18}$$

$$SS_{total} = \sum_i (x_i - \bar{x}_i)^2 \tag{19}$$

In the preceding equations, x_i , $\hat{x}_i, and$ x represent the true, predicted, and average values, respectively. The number of samples is denoted by n. MAPE denotes the ratio of prediction bias to the actual value. Because each data type has a different data range, the calculated error varies significantly between them. The R^2 can be defined as the ratio of the predicted mean square error to the variance of the data. It denotes the fitness of the predicted and actual values. Table 8 displays the calculated evaluation indicators.

Table 8. Predictive Evaluations of Multiple Models

Algorithms	MAPE	MAE	R^2	Computing time (ms)
LMS	1.1765	0.1125	0.9841	0.10
P-PDA	0.6080	0.1065	0.9856	0.09
CDPM	0.3287	0.0887	0.9888	0.07
Proposed	0.2062	0.0545	0.9925	0.05

Fig 14 depicts a loss function performance comparison of LMS, P-PDA, CDPM, and the suggested DGCN. It can be seen that the suggested DGCN has the lowest MAPE, MAE, and R^2. The DGCN MAE for the following sensor data is 0.0545 for the given temperature dataset.

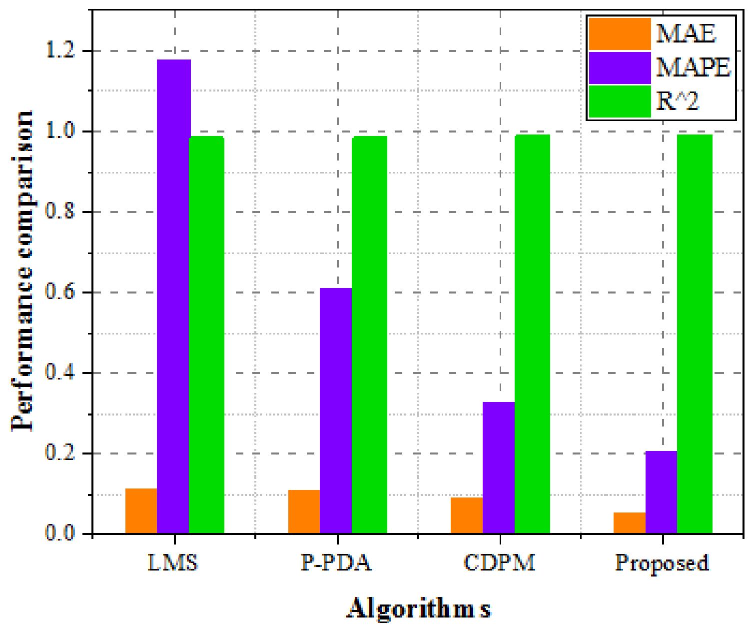

Fig 14. Performance Comparison of Proposed Prediction Among All Algorithms

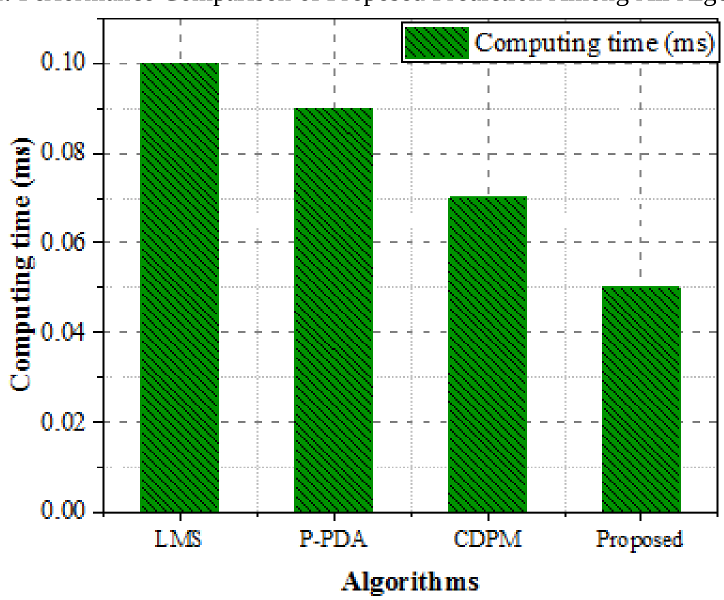

Fig 15. Computing Time of Proposed Prediction Among All Algorithms

Fig 15 shows computational time comparisons for predicting individual sensor values. It can be seen that LMS has the longest computation time of 0.010 ms, while the suggested DGCN has the shortest computation time of 0.05 ms. The suggested algorithm outperforms existing algorithms.

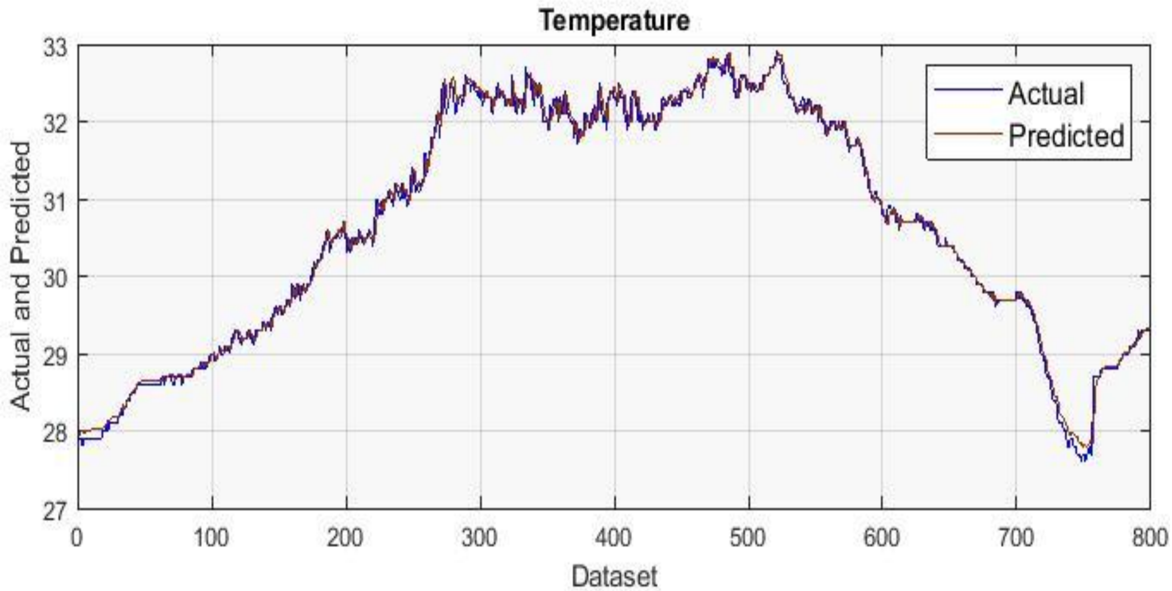

Fig 16. Prediction Accuracy for Various Data Sets

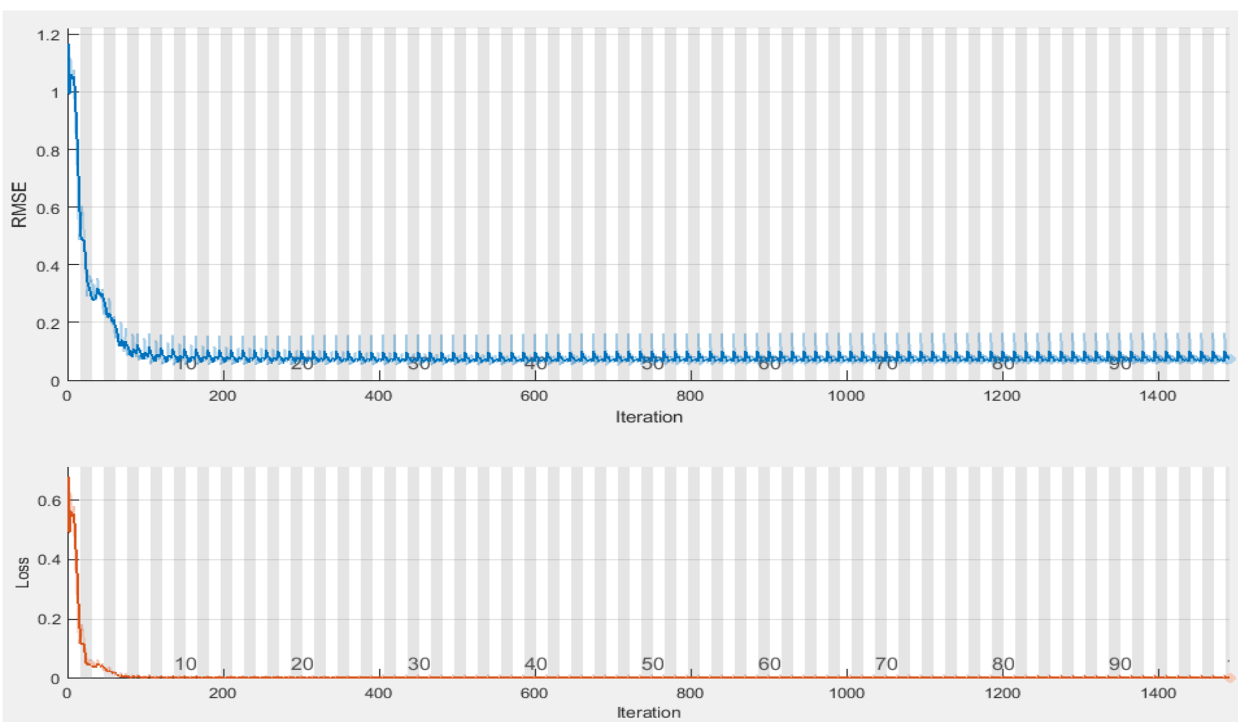

Fig 17. RMSE Prediction Error Using DGCN

Fig 16 and 17 show the predictive efficiency for non-dynamic data on both datasets, demonstrating the system's effectiveness.

V. CONCLUSION

This work suggests a method based on clusters and predictions for collecting energy-saving data. During the clustering stage, the sensor nodes form groups, and the heads of the groups gather and save the data estimated by the sensor nodes. The presented hybrid prediction model investigates the relationship between communication and prediction. The model's performance was evaluated by employing different numbers of nodes at different times. According to simulation results, the proposed model outperforms other related methods regarding energy efficiency and prediction accuracy. As a result, it may substantially decrease energy use for data accumulation in hierarchical networks and significantly increase network longevity, even when many clusters are allocated. The results show that the proposed data aggregation method surpasses existing LMS, P-PDA, and CDPM methods. The proposed DGCN method has a transmission suppression rate of 92.68%, a difference of 22.33%, 16.69%, and 12.54% compared to current methods (i.e. LMS, P-PDA, and CDPM). In future work, traffic generators, including LoRa network operators, can be incorporated into the suggested application.

Data Availability

No data was used to support this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding was received to assist with the preparation of this manuscript.

Ethics Approval and Consent to Participate

The research has consent for Ethical Approval and Consent to participate.

Competing Interests

There are no competing interests.

References

[1]. S. Thomas and T. Mathew, “Secure Data Aggregation in Wireless Sensor Network using Chinese Remainder Theorem,” International Journal of Electronics and Telecommunications, Jul. 2023, doi: 10.24425/ijet.2022.139886.

[2]. O. Ojuroye, R. Torah, S. Beeby, and A. Wilde, “Smart Textiles for Smart Home Control and Enriching Future Wireless Sensor Network Data,” Smart Sensors, Measurement and Instrumentation, pp. 159–183, Oct. 2016, doi: 10.1007/978-3-319-47319-2_9.

[3]. S. Z. Erdogan and T. T. Bilgin, “A data mining approach for fall detection by using k-nearest neighbour algorithm on wireless sensor network data,” IET Communications, vol. 6, no. 18, pp. 3281–3287, Dec. 2012, doi: 10.1049/iet-com.2011.0228.

[4]. S. Boubiche, D. E. Boubiche, A. Bilami, and H. Toral-Cruz, “Big Data Challenges and Data Aggregation Strategies in Wireless Sensor Networks,” IEEE Access, vol. 6, pp. 20558–20571, 2018, doi: 10.1109/access.2018.2821445.

[5]. X. Li, X. Tao, and Z. Chen, “Spatio-Temporal Compressive Sensing-Based Data Gathering in Wireless Sensor Networks,” IEEE Wireless Communications Letters, vol. 7, no. 2, pp. 198–201, Apr. 2018, doi: 10.1109/lwc.2017.2764899.

[6]. Ben Arbi, F. Derbel, and F. Strakosch, “Forecasting methods to reduce energy consumption in WSN,” 2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), May 2017, doi: 10.1109/i2mtc.2017.7969960.

[7]. Hongbo Jiang, Shudong Jin, and Chonggang Wang, “Prediction or Not? An Energy-Efficient Framework for Clustering-Based Data Collection in Wireless Sensor Networks,” IEEE Transactions on Parallel and Distributed Systems, vol. 22, no. 6, pp. 1064–1071, Jun. 2011, doi: 10.1109/tpds.2010.174.

[8]. S. C, S. D, and N. A, “Accurate Data Aggregation Created by Neural Network and Data Classification Processed Through Machine Learning in Wireless Sensor Networks,” Sep. 2021, doi: 10.21203/rs.3.rs-895195/v1.

[9]. Hongbo Jiang, Shudong Jin, and Chonggang Wang, “Prediction or Not? An Energy-Efficient Framework for Clustering-Based Data Collection in Wireless Sensor Networks,” IEEE Transactions on Parallel and Distributed Systems, vol. 22, no. 6, pp. 1064–1071, Jun. 2011, doi: 10.1109/tpds.2010.174.

[10]. G. Mustafaraj, G. Lowry, and J. Chen, “Prediction of room temperature and relative humidity by autoregressive linear and nonlinear neural network models for an open office,” Energy and Buildings, vol. 43, no. 6, pp. 1452–1460, Jun. 2011, doi: 10.1016/j.enbuild.2011.02.007.

[11]. A. Agarwal and A. Dev, "A Data Prediction Model Based on Extended Cosine Distance for Maximizing Network Lifetime of WSN," WSEAS Transactions on Computer Research, Vol 7, pp. 23–28, 2019.

[12]. G. B. Tayeh, A. Makhoul, D. Laiymani, and J. Demerjian, “A distributed real-time data prediction and adaptive sensing approach for wireless sensor networks,” Pervasive and Mobile Computing, vol. 49, pp. 62–75, Sep. 2018, doi: 10.1016/j.pmcj.2018.06.007.

[13]. S. A. Soleymani et al., “A Hybrid Prediction Model for Energy-Efficient Data Collection in Wireless Sensor Networks,” Symmetry, vol. 12, no. 12, p. 2024, Dec. 2020, doi: 10.3390/sym12122024.

[14]. G. C. Jagan and P. J. Jayarin, “A Novel Machine Language-Driven Data Aggregation Approach to Predict Data Redundancy in IoT-Connected Wireless Sensor Networks,” Wireless Communications and Mobile Computing, vol. 2022, pp. 1–20, Oct. 2022, doi: 10.1155/2022/7096561.

[15]. H. Wang, Z. Yemeni, W. M. Ismael, A. Hawbani, and S. H. Alsamhi, “A reliable and energy efficient dual prediction data reduction approach for WSNs based on Kalman filter,” IET Communications, vol. 15, no. 18, pp. 2285–2299, Jul. 2021, doi: 10.1049/cmu2.12262.

[16]. S. Loganathan, J. Arumugam, and V. Chinnababu, “An energy-efficient clustering algorithm with self-diagnosis data fault detection and prediction for wireless sensor networks,” Concurrency and Computation: Practice and Experience, vol. 33, no. 17, Apr. 2021, doi: 10.1002/cpe.6288.

[17]. Uma Maheswari Arumugam, Suganthi Perumal, “Trust based Secure and Reliable Routing Protocol of Military Communication on MANETs”, Journal of Machine and Computing, pp. 047-057, January 2023. doi: 10.53759/7669/jmc202303006.
[18]. K. Jain and A. Kumar, “An energy-efficient prediction model for data aggregation in sensor network,” Journal of Ambient Intelligence and Humanized Computing, vol. 11, no. 11, pp. 5205–5216, Mar. 2020, doi: 10.1007/s12652-020-01833-2.
[19]. W. Zheng, “Current Technologies and Applications of Digital Image Processing,” Journal of Biomedical and Sustainable Healthcare Applications, pp. 13–23, Jan. 2023, doi: 10.53759/0088/jbsha202303002.
[20]. N. Wambui, “Medical Identification and Sensing Technology for Assisting and E-Health Monitoring Systems for Disabled and Elderly Persons,” Journal of Biomedical and Sustainable Healthcare Applications, pp. 9–17, Jan. 2022, doi: 10.53759/0088/jbsha202202002.
[21]. A. Agarwal, K. Jain, and A. Dev, “Modeling and Analysis of Data Prediction Technique Based on Linear Regression Model (DP-LRM) for Cluster-Based Sensor Networks,” International Journal of Ambient Computing and Intelligence, vol. 12, no. 4, pp. 98–117, Oct. 2021, doi: 10.4018/ijaci.2021100106.
[22]. S. Ninisha Nels and J. Amar Pratap Singh, “Hierarchical Fractional Quantized Kernel Least mean Square Filter in Wireless Sensor Network for Data Aggregation,” Wireless Personal Communications, vol. 120, no. 2, pp. 1171–1192, May 2021, doi: 10.1007/s11277-021-08509-w.
[23]. H. Wang, Z. Yemeni, W. M. Ismael, A. Hawbani, and S. H. Alsamhi, “A reliable and energy efficient dual prediction data reduction approach for WSNs based on Kalman filter,” IET Communications, vol. 15, no. 18, pp. 2285–2299, Jul. 2021, doi: 10.1049/cmu2.12262.
[24]. K. Jain and A. Kumar, “A lightweight data transmission reduction method based on a dual prediction technique for sensor networks,” Transactions on Emerging Telecommunications Technologies, vol. 32, no. 11, Aug. 2021, doi: 10.1002/ett.4345.
[25]. S. Famila, A. Jawahar, S. L. S. Vimalraj, and J. Lydia, “Integrated Energy and Trust-Based Semi-Markov Prediction for Lifetime Maximization in Wireless Sensor Networks,” Wireless Personal Communications, vol. 118, no. 1, pp. 505–522, Jan. 2021, doi: 10.1007/s11277-020-08028-0.
[26]. X. Liu, G. Li, and P. Shao, “A Multi-Mechanism Seagull Optimization Algorithm Incorporating Generalized Opposition-Based Nonlinear Boundary Processing,” Mathematics, vol. 10, no. 18, p. 3295, Sep. 2022, doi: 10.3390/math10183295.
[27]. X.-J. Mao, C. Shen, and Y.-B. Yang, "Image Restoration Using Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections," Neural Information Processing Systems (NIPS 2016).
[28]. T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," Conference Paper at ICLR 2017, doi: 10.48550/arXiv.1609.02907.
[29]. X. Fu, Q. Qi, Z.-J. Zha, Y. Zhu, and X. Ding, “Rain Streak Removal via Dual Graph Convolutional Network,” Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 2, pp. 1352–1360, May 2021, doi: 10.1609/aaai.v35i2.16224.
[30]. X.-Y. Liu et al., “<italic>CDC</italic>: Compressive Data Collection for Wireless Sensor Networks,” IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 8, pp. 2188–2197, Aug. 2015, doi: 10.1109/tpds.2014.2345257.
[31]. D. Fernandes, A. G. Ferreira, R. Abrishambaf, J. Mendes, and J. Cabral, “A machine learning-based dynamic link power control in wearable sensing devices,” Wireless Networks, vol. 27, no. 3, pp. 1835–1848, Jan. 2021, doi: 10.1007/s11276-020-02539-1.