Journal of Al-Qadisiyah for Computer Science and Mathematics VOL.17.(4) 2025,PP.COMP 27–44

Available online at www.qu.edu.iq/journalcm
JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS
ISSN:2521-3504(online) ISSN:2074-0204(print)

Energy Conserving Communication in WSN based on static data prediction by using SEKF

Furqan Muhammad Abbas[a], Hadeel Noori Saad[b]

[a]*Department of Compute, College of Education, University of Kufa, Iraq. Email: furqanm.aldaraji@student.uokufa.eu.iq*

[b]*Department of Compute, College of Education for Women, University of Kufa, Iraq. Email: hadeel.jrew@uokufa.edu.iq*

ARTICLE INFO

Article history:
Received: 11/07/2025
Rrevised form: 03/08/2025
Accepted : 04/08/2025
Available online: 30/12/2025

Keywords:

Energy, Wireless Sensors Network (WSNs), Static ,Prediction, Extended Kalman Filter (EKF) .

ABSTRAC

Wireless Sensors Networks (WSNs) recently, have drawn a lot of attention. Despite its potential applications in a wide range of fields, wireless sensor nodes' restricted communication bandwidth, inadequate processing units, little memory, and power limitations severely limit their capabilities. One of the main challenges in this area is extending the life of battery-powered sensors in WSNs by reducing energy usage. This problem is addressed using a variety of techniques, such as deep learning, machine learning techniques, statistical techniques, and time series forecasting. One strategy is to utilize data prediction to reduce the volume of transmitted data without sacrificing its quality. This paper presents a model for wireless sensor networks energy saving using the Static Extended Kalman Filter (SEKF). The technique is used to accurately dual predict. The plan consists of two stages. In the first stage, the transmission from the sensor node to the sink node is reduced based on four steps (data equality, data deviation computation, faulty data detection, data reduction based on prediction). In the second stage, the data is reconstructed at the sink node to maintain system reliability. The proposed model demonstrated superior performance compared to other methods, reducing data throughput in the first phase by 60.72%. In the second phase, data was reconstructed with 97.86% accuracy at a data reduction rate of 62–60%, with an energy consumption of 3.928 J. These results were achieved by SEKE for single-node reconstruction. Furthermore, the proposed model performed well when applied to data containing negative values, achieving acceptable data reduction with accuracy ranging from (94-95%) in several experiments. The Intel Berkeley Research Lab (IBRL) dataset was used for all experiments.

MSC.

https://doi.org/10.29304/jqcsm.2025.17.42533

1. Introduction

Wireless sensor networks (WSNs) are low-cost systems composed of wireless sensor nodes and antennas for environmental monitoring [1]. Their small size and portability make them suitable for remote areas. Despite limited energy, processing power, and storage, WSNs have many applications in environmental sustainability and smart cities. These networks consist of sensors distributed in the environment that collect and wirelessly transmit data to a central sink node. Battery life is a major concern since radio transmissions consume the most energy due to large data volumes. Effective data management is crucial to extend network life and reduce energy use. Because sensory data show high temporal and spatial correlation, predicting data is a promising strategy to reduce transmitted data by using past data to forecast future values, thus eliminating redundant information before transmission[2]. Several

∗Corresponding author

Email addresses:

Communicated by ‘sub etitor’

advanced techniques such as data aggregation [3][4] data compression [5][6], adaptive sampling [7], and data prediction [8] have been proposed. Compared to others, data prediction achieves a higher data Suppression Ratio (SR), making it a more effective and preferred approach. data prediction builds a prediction model based on correlations in previously collected data to forecast future measurements. Data that can be estimated at cluster heads, sinks, or base stations are not transmitted. By comparing measured data with predicted values at sensor nodes, transmissions are stopped when predictions meet accuracy requirements. Only unexpected data are sent to the sink. Prediction accuracy depends on the sink's ability to reconstruct non-transmitted data. However, some DP strategies impose high computational loads on sensor nodes, which may not be suitable for resource-constrained environments. This paper proposes a new prediction-based data reduction approach aiming to improve transmission reduction, data reliability, and energy consumption. The main contributions are:

- **Developing a data reduction algorithm** to discard redundant, predictable, or faulty data at sensor nodes .
- **Developing a data prediction algorithm** at the sink node based on the Kalman Filter to predict missing data reliably

Figure (1) illustrates the structure of WSN

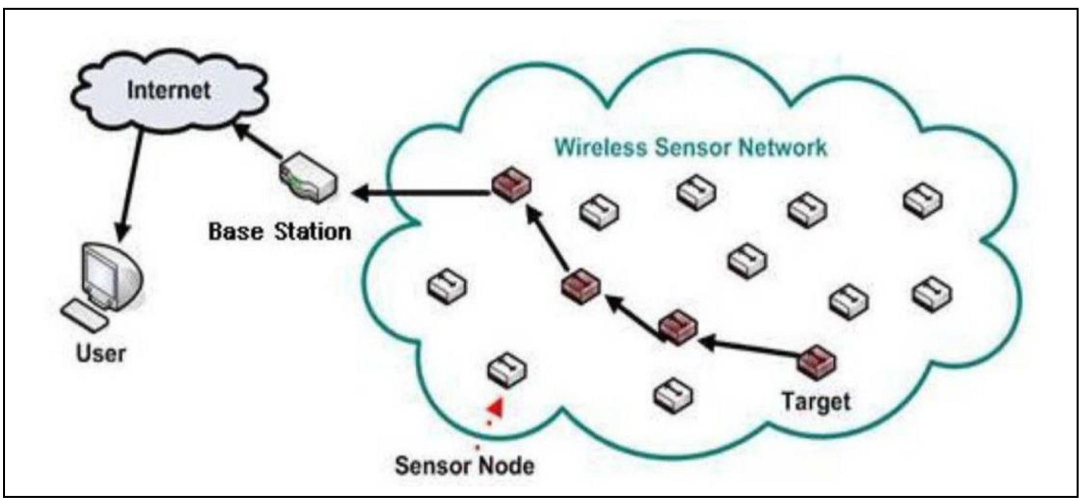

Figure (1): WSN Structure

2. Related Researches

Prediction-based data transmission reduction in wireless sensor networks uses for past data analysis to forecast future values, allowing just critical information to be transmitted rather than all data. According to some research, employing prediction can also increase sensor device battery life and enhance network performance as a whole. Predictive models have been created by researchers using a variety of methods, including:

Bashar Chreim et al. (2021) [9] introduce (RADAR), a simultaneous prediction model that utilizes linear correlation among all data variables. The models include time series prediction and linear regression. These models produce satisfactory and accurate results while still being straightforward. A time series model will predict the value of the first applications' variables, which in turn input to the next SLR prediction model to predict the second variable. Then, successive predictions of the value of the next corresponding variable take place by multiple linear regression (MLR) simultaneous models. The proposed dual prediction scheme (DPS) implemented on both the source and the destination nodes. RADAR outperforms LMS_MOD in terms of RMSE, data reduction percentage, and energy consumption. The results fill in between 5 and 14% and 23 and 34% for humidity and black photons data reduction respectively.

Marcin Lewandowski and Bartłomiej Płaczek. (2021) [10] proposed a novel dual prediction model that can be applied to neural nets, decision trees, random forests, and other frameworks. The suggested method determines whether or not to transmit the sensed data without using prediction error. This approach, on the other hand, checks to see if the anticipated data are accurate enough to identify the relevant events. The parent node's event detection task can be characterized as a binary function with the formula

$$E_{(t)} = event(spt,sct,E) \quad (1)$$

Where E(t) = 0 in the absence of an event and 1 in the presence of one and sensor readings taken at time step t by the parent (Spt) and the child nodes (Sct) respectively. the experimental results achieves 94% accuracy, 79% data reduction with 0.1 % accuracy depletion in comparison with ANN, and Naïve models.

Haibin Wang and colleagues. (2021) [11] presented a data reduction method based on Dual Prediction. The model is divided into two stages. Data reduction is the focus of the first phase, which comprises: fault and equal data detection, and data deviation computation. Kalman filter in the second phase aims to predict unsent data as an expectation of previously seen data. The experiments were evaluated with the Intel Berkeley Research Lab (IBRL). The obtained results demonstrated that the suggested method could preserve data reliability while reducing data transmission by up to 75.75%. The suggested method not only reduces data but also finds and removes inaccurate data.

A Combinational Data Prediction Model (CDPM) was developed by KHUSHBOO JAIN et al. (2022) in [12]; it can predict future data to minimize data transmission and build previous data to manage latency. The training and prediction phases serve as the foundation for the model's construction. The training phase delay is reduced by adjusting the training data size in accordance with the data interrelations. Comparing the CDPM model to the HLMS, ELR, and P-PDA algorithms, respectively, experiments conducted on temperature and humidity in the Intel Berkeley Research Laboratory show significant transmission reduction about (16.49%, 19.51%, and 20.57%), enhanced energy saving (29.56%, 50.14%, 61.12%), and enhanced accuracy (15.38%, 21.42%, 31.25%).

AROUNA NDAM NJOYA et al. (2022) [13] proposed a model that uses a sequence-to-sequence (Seq2Seq) encoder-decoder neural network with LSTM units to predict spatial features from sensed data in WSNs. By producing additional information, the previously mentioned method may reduce network traffic and energy consumption for data transmission on WSNs. The experimental results on Berkeley Research Laboratory data from Intel, demonstrate that the proposed model can save twice as much energy and accurately forecast data with little error (measured by Root Mean Squared Error) when the appropriate nodes are used.

In [23] El-Sayed, Walaa M. et al [14]. (2023) proposed the Distributed Data Predictive Model (DDPM), which operates in three phases: Dissemination/Mobility, Classification, and Data Generating. Sensors transmit data to clusterhead nodes that categorize faults, identify missing data, and assess sensor status. Data prediction is performed using Recursive Least Squares (RLS) and Finite Impulse Response (FIR) adaptive filters. RLS minimizes a weighted least squares cost function, while FIR adapts without feedback, producing output via convolution. The model recovered nearly 99% of lost data, reduced energy consumption, enhanced network performance, and decreased transmitted signals, achieving 19% reliability in WSNs.

M. Revanesh et al. (2023) proposed the ANN-ILMNN model in [15], an enhancement of LEACH and ESR protocols, incorporating the Levenberg-Marquardt Neural Network (LMNN) for improved anomaly detection and energy efficiency. Simulations showed ANN-ILMNN outperformed other models, achieving 97.85% accuracy with 600 data points compared to LEACH (84.89%), EESR (87.94%), LEACH-LMNN (90.69%), and EESP-LMNN (94.59%). Energy consumption was significantly lower with ANN-ILMNN, requiring 29.12 J for 20 nodes, compared to LEACH (41.24 J) and others. The model demonstrated superior accuracy and energy efficiency across varying network sizes.

A prediction-error-based method (PEM) was proposed by Umut Yildirim et al. (2024) in [16] to optimize transmissions and detect damages in wireless sensor networks, the methodology combines prediction, FFT, and bandpass filters for noise reduction. Decisions are made based on the deviation of predicted data from the actual sensor reading. The term "excitation data," which refers to the set of data read from a trustworthy set of sensors as reference data for prediction, is attractively used by the model. The results show that the transmissions for each sensor can be minimized to 10% where 90% of the readings are unsent with 45% predefined error tolerance.

Li Wu et al. (2024) in [17] proposed Fourier transform for a new data collection strategy in the energy control system, which exhibits power in periodic characteristics detection of the temperature data. The study demonstrates that these data can be linearly fitted using Fourier transforms, where data parameters can be optimized with least squares. This model reduces equal-value data and prolongs the lifespan of sensor nodes. Although Compared to the complex model, the Fourier transform needs more computation and high parameterization to fit with, it shows much better performance in data with periodic characteristics, where it exhibits fitness up to 93.0285.

Ting Hu. (2024) in [18] suggested a data fusion mechanism for WSNs to optimize clustering design. The proposal enhances the conventional cluster-based routing protocol and develops a deep-learning (DL) data fusion algorithm. In this manner, the cluster head plays the main role in extracting, classifying, and fusing similar data features, whereas the cluster's node members fit the gathered raw data using the DL model. The experiments performed on the KDDCup99 dataset, where results showed reduced data transmission, improved energy utilization, and prolonged network life. The proposed algorithm out performs LEACH and DFA-IACOBP by 32.2% and 15.9% in energy saving respectively.

3. Proposed Approach

The proposed model begins by loading datasets. Three sensor nodes are used to read the data. Data is predicted at the sensor node to reduce transmission and then transmitted to the sink node. Un transmitted data is then predicted at the last mentioned node. Finally, a model that conserves sensor battery power is created by combining data reduction at the sensor node and reconstruction it at the sink node using a Kalman filter algorithm. Figure (2) illustrates the basic steps of the proposed method.

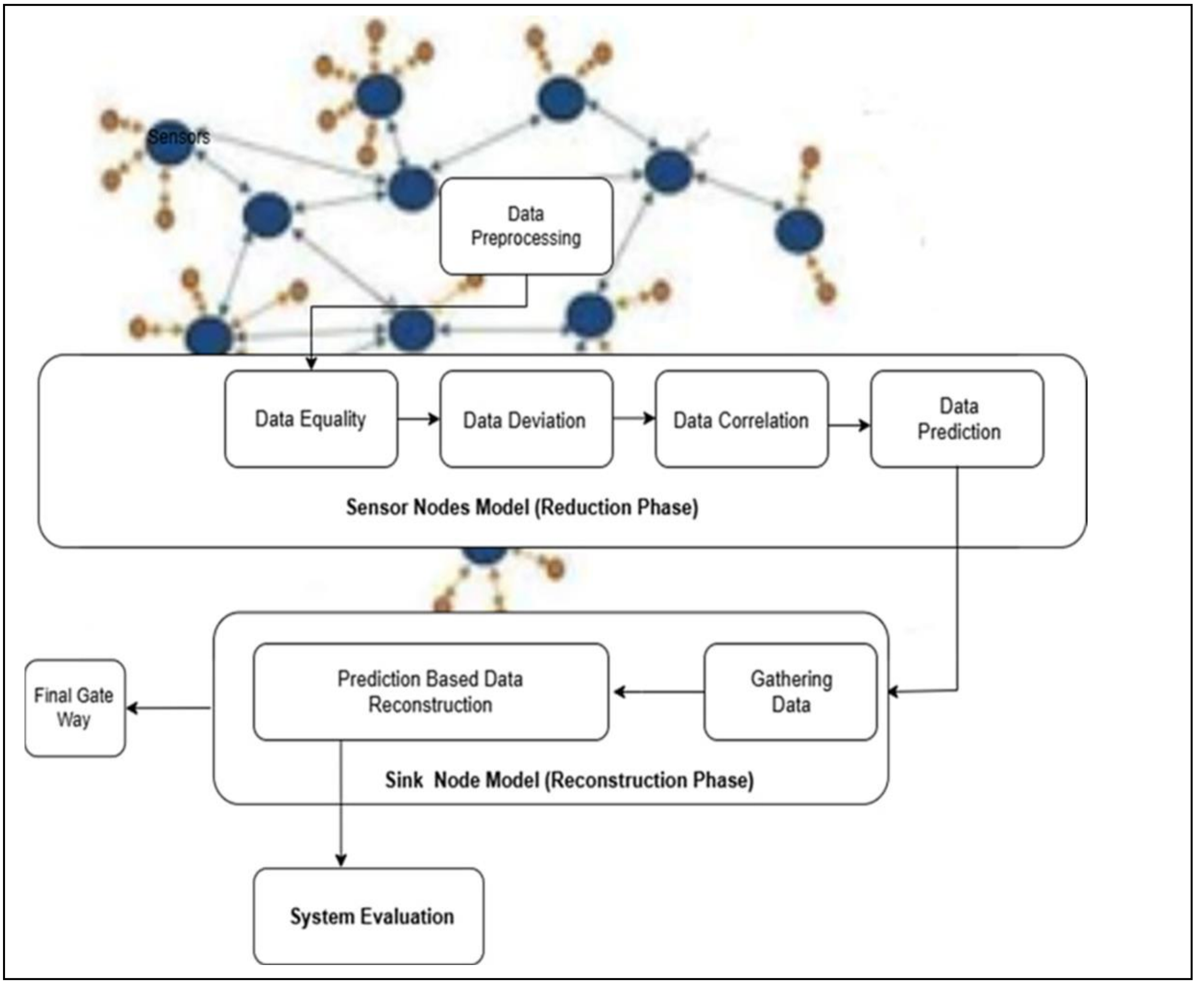

Figure (2): Proposed System Model

3.1 Experiments Environment

It is important to mentioned that all the experiments were conducted within MATLAB Integrated Environment version (2020) for model simulation. The computational setup consisted of a laptop equipped an Intel Core i5 processor, 8 GB RAM, and running Windows 10 operating system.

3.2 Data Set

This work focused on working with WSN-environmental data so, Intel Berkeley research lab (IBRL) Dataset is a consistent to be utilized in this study. This dataset consists of real sensor nodes readings collected between February and April, 2004. A total of 54 Mica2Dot sensor nodes were deployed within IBRL facilities to record environmental data, including temperature, humidity, light, and voltage within sampling interval of about 31 seconds. For the purpose of, a subset of 10,000 humidity readings was selected. Sensors 1, 2, and 3 were utilized to assess the proposed method. For additional details on the datasets employed, Further information regarding the dataset can be found in [19].

Figure (3) shows an architectural diagram of the dataset.

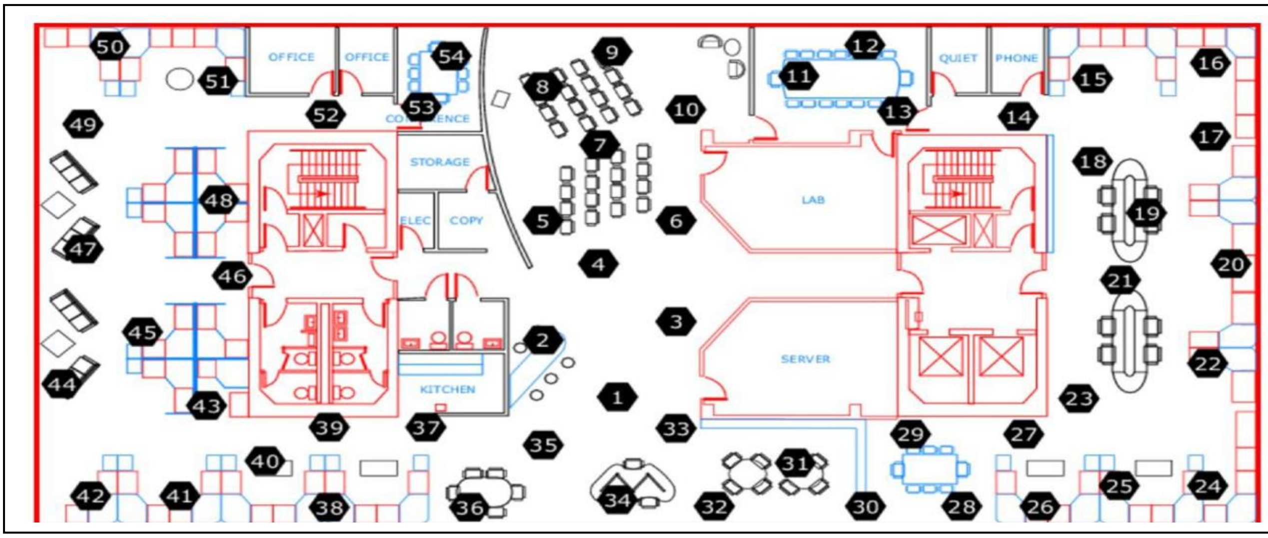

Figure (3): Intel Berkeley Research Lab Dataset

3.3 **Data Preprocessing**

Collecting high-quality datasets is essential to building an excellent model, as the model's effectiveness depends on the data it inputs. When building our system, we used humidity data to implement the algorithm. The data was cleaned by removing negative values.

3.4 Sensor Nodes Model (Data Reduction Phase)

The data reduction phase aims to reduce the number of data transfers between sensor nodes and sink nodes. In the presented work, links between each sensor node and its corresponding sink node are used to achieve the targeted reduction. The data reduction phase relies on four algorithm steps implemented in each sensor node start with data equality, data deviation computation, faulty data detection, and data reduction based on prediction. The following subsections detailed the processing operations within each step.

3.4.1 Data Equality (DE)

The data reduction phase algorithm's initial phase, data equality, determines whether or not the newly sensed information is equivalent to the prior reading, as specified by Equation (2):

$$z_t - l_{x-1=0} \qquad (2)$$

where l_{x-1} is the previous value and z_t is the current reading. A certain amount of values from each sensor node in the network were first cached by the data reduction phase before being sent to the sink node. The same sensor node then compares each newly sensed reading z_t at time t of sensor s_i with the previously acquired reading l_{x-1} .Therefore, if no change is found, the current reading z_t , is disregarded. And will begin the second phase of the suggested method.

3.4.2 Data Deviation Computation (DDC)

In order to determine the value of this deviation and transmit or delete the reading appropriately, data deviation computation makes sure that the current sensed reading z_t , differs in some way from the previous reading l_{x-1} . In fact, two distinct methods are suggested in the proposed data reduction phase to determine the data deviation computation. The initial procedure seeks to determine, using Equation (2), the difference between the present sensed value, z_t , and the earlier readings l_{x-1} . If the difference between z_t and l_{x-1} is smaller than the predetermined e_{max}, the data transmission will be discarded, and the cache updated. If not, the second data deviation computation procedure begins. To determine how far the current sensed reading deviates from the expected values, a data deviation computation is presented to calculate the deviation of the current sensed reading from their predicted values. Comparing the current reading z_t with the Kalman Filter based estimated value est, which is nearly identical to the prior reading, is the principle behind this procedure because Kalman Filter estimated values are quite accurate. Equation (4) computes the difference between z_t and l_{x-1} , If deviation E_{dev} is bigger than the predefined threshold e_{max}, then z_t is transmitted to the sink node; else, the z_t data transmission is discarded, and the cache is updated:

$$\boldsymbol{Vdev_{t\,=}\, z_{t\,-\,l_{x-1}}} \tag{3}$$

$$\boldsymbol{Edev_t\text{ - }|z_{t-\,est_t}|} \tag{4}$$

3.4.3 Faulty Data Detection (FDD)

Faulty data detection is used to stop inaccurate sensed readings from being transmitted. Indeed, because of their limited resources, wireless sensor nodes are prone to malfunction. Therefore, ensuring that the data collected is error free is crucial for data reliability and accuracy. Equations (5) and (7) serve as the foundation for the suggested faulty data detection approach in this stage. Fault detection is a crucial procedure since WSNs are prone to malfunctions. The suggested method takes defect detection procedures into account, in contrast to several cutting edge data reduction techniques:

$$dis = \sum_{i=0}^{n} |z_{t\,-\,l_i}| \tag{5}$$

$$corr = |\, dis - (|\, l_{max}\text{ - } l_{min}\, |)\, | \tag{6}$$

$$\boldsymbol{z_t = \{\, transmitted\ if\ dis < \ \theta\ , discarded\ otherwise} \tag{7}$$

where $\boldsymbol{z_t}$ denotes the current sensed reading, $\boldsymbol{l_{max}}$ and $\boldsymbol{l_{min}}$ are the maximum and minimum cached readings, respectively, and $\boldsymbol{\theta}$ is a predetermined value determined by the application requirements. Dis indicates the distance between the values that are cached and the current reading. When comparing the current sensed reading to the pre-cached readings, the $\boldsymbol{corr}$ is the difference between the maximum and minimum cached values. Equations (5) and (7) determine how to discard defective data transfer and update the cache with the estimated value [11].

3.4.4 Predictable Data Reduction

In Data Reduction based on Prediction step Extended Kalman Filter is used to recursive estimation are made to predict sensor readings, reducing data transmissions by sending the new data, only when prediction errors exceed a predefined threshold, this can further reduce the transmitted data and consequently enhancing energy harvesting. For efficient and accurate data prediction, manual tuning estimation mode were used.

Algorithm (1): Data Reduction Phase

Input: Sensor current Readings

Output: Reduced Transmitted Readings

Step 1: Initialize variables

- l_{cache} $\longrightarrow$ size for previously collected readings
- $z(t)$ $\longrightarrow$ current reading value

- l_{x-1} → last cache value
- est → Kalman Filter based estimated value
- e_{max} → maximum acceptable deviation for data transmission
- $corr$ → correlation factor of current sensed reading
- $theta$ → faulty data detection threshold
- dis → distance between the current reading and the cached values
- $threshold$ → threshold of prediction

Step 2: Read current sensor value

1. If it is the last *l* readings, then send to Sink
2. else go to next step

Step 3: Data Equality (DE)

1. If $zt = l_{x-1}$ then
   Discard transmission
   l.append(*zt*) (add new value to cache)
   EKF.update(*zt*)
   Continue to the next iteration
2. Else go to next step

Step 5: Data Deviation Computation (DDC)

2. Mean(l_{x-1}) = mean(*l*_cach)
3. if $|zt-l_{x-1}|$ < mean(*l*) then
4. Discard transmission
   l.append(*zt*) (Remove the oldest value from cache and append zt)
   EKF.update(*zt*)
   Continue to the next iteration
5. else go to next step

Step 6: Faulty Data Detection (FDD)

1. Compute dis based on equation (4)
2. Compute *Corr* based on equation (5)
3. If *Corr > theta*
4. Discard transmission
5. l.append(*zt*) (Remove the oldest value from cache and append zt)
   EKF.update(zt)
   Continue to the next iteration
6. Else continue to next step

Step 7: Data Prediction

1. *est=EKF.estimate (lx-1)*
2. *EstErr = abs (zt – est)*
3. if *EstErr>= threshold*
   Discard transmission
   l.append(zt) (Remove the oldest value from cache and append zt)
4. Else send to sink and Continue to the next iteration

End

3.5 Sink Node Model Data Reconstruction Phase (DRCP)

In WSN, the sensor node senses, generates, and transmits all of the data that the sink node receives. In general, sink nodes outperform sensor nodes in terms of performance, transmission capacity, and computing power. Thus, the primary objective of the suggested method is to strike a balance between data accuracy and dependability on the one hand, and data reduction and energy consumption on the other. As a result of the Sensor Nodes DRP, the accuracy and dependability of WSN data are impacted since the data gathered by sensor nodes is deemed full in comparison to the data received by the sink node. The data reconstruction phase, are suggested as a solution to these problems. The core function of this phase is to predict un transmitted data at the sink node at each time interval (t) using Extended Kalman Filter. Un transmitted readings are Reconstructed using previously cached readings from sensor nodes, where each reading is cached as (l) for a sensor node (si). For every sensor node (si) the reading obtained at time t is used to update the cached values. WSN nodes in many applications are characterized by geographically closed spaced (the proposal of this work), then at any time point t where there is no reading is received, two methods are fired to forecast the missing reading, self-prediction (SP) and neighbor prediction (NP) method.

3.5.1 Neighboring-based prediction (NP)

This method benefit from the close proximity of neighbor sensors which make them reported high similarity data, this the good news for sink node to compensate missing data with real existing neighbor similar data and increase reconstruction accuracy by avoiding further predictions. NP aims to check whether one of the targeted sensor neighboring nodes is transmitting readings at time t. In the case of one or more neighboring transmit readings, the sink node fill missing one with this data. The selection of the neighbor whose data to be candidate for compensation is strait forward process follow the sequential search within neighbor set, while it is a small set, that require negligible time consuming.

3.5.2 Self-based prediction (SP)

Self-based prediction (SP) method aims to predict un transmitted readings passed through the first step. In fact, the nature of the network plays a crucial role in determining the prediction technique. In some WSN, sensor nodes are spatially redundant, which increases reading similarity, and cause all neighbor nodes to behave the same upon data reduction (deleting data for the same data slots). However, the proposed SP data reconstruction used extended Kalman filters to predict missing data points. Every time slot t the sink node update Kalman Filter state to synchronize it with the last reading of the targeted Sensor.

Algorithm (2) Sink Node Model (Data Reconstruction phase).

Input: Reading Data (SN, T) Received time stamped Data

Sensor Nodes Pool (SN)

Output: Reconstructed Data (RD(T))

Step 1: for the received data pool (D) check missing data at time t

```
If ¬(∃dData(sn,t))≡ ∀d¬Data(sn,t) then
    Identify Missing Data (Time Slot t)
else
    update EKF state
    RD(t) = cache (data (d,t))
```

Step 3: Check Neighboring node's Data at (Time Slot t)

```
If ∃sn(sn(j)∈((negibor)∧SN)) ∧Data(sn(j), t)) then
    RD(t)= fill. Data(sn(j), t) where j=1,2,…,n neighbor sensors index
else
    RD(t)= EKF.Predict (Data(sn(i), t-1)) where i=1,2,…,n current sensor index
Update Kalman Filter
```

Step 4: send constructed data to gateway node

end

4. Results and analysis

This section examines the outcomes of data reduction and prediction algorithms using various e_{max} values (0.0002, 0.0003,0.0005,0.0007and 0.0009). The suggested performance for data reduction is assessed by contrasting the percentage sizes of the input and output data. Additionally, ten thousand humidity readings from sensors 1, 2, and 3 were used to assess the effectiveness of the suggested strategy. Additionally, to confirm the efficacy of the suggested data reduction technique, the outcomes of three well-known data reduction approaches are compared with the results of the proposed approach. The criteria chosen to assess the outcomes of the suggested strategy are data reduction percentage, data correctness, and energy conception.

4.1 Data Reduction Stage

The first stage of the proposed model begins by minimizing data transmission from the sensor node to the sink node using the proposed Kalman Filter based model to save the nodes' battery power, where data reduction encompasses removing equal data readings (data equality), data deviation computation, faulty data detection, and EKF-based predictable data. The experiments was accommodated using manual tuned EKF parameters and a certain threshold value e_{max} .

4.1.1 Static Extended Kalman Filter (SEKF)

The experiment involved applying an EKF with the process noise set to 0.03 and the measurement noise set to 0.16. Table (1) shows the Percentage of reduced transmitted data with selected threshold values e_{max} and 10000 data readings.

Table (1) Data Reduction Percentage in Sensors (1-3) with (SEKF)

e_{max}	Process Noise (Q)	Measurement Noise (R)	Sensor 1	Sensor 2	Sensor 3
0.0002	0.03	0.16	23.29	23.52	23.60
0.0003	0.03	0.16	30.36	30.75	30.71
0.0005	0.03	0.16	42.61	43.42	43.48
0.0007	0.03	0.16	52.72	53.01	53.18
0.0009	0.03	0.16	60.25	60.51	60.72

The results show the impact of change e_{max} parameter on the data reduction Percentage achieved across three sensor nodes, The percentage increases with increasing e_{max} from 0.0002 to 0.0009, Sensor 1 shows an increase from 23.29% to 60.25%, Sensor 2 from 23.52% to 60.51%, and Sensor 3 from 23.60% to 60.72%. On other hand, system parameters were carefully tuned with 0.03 and 0.16 for Q and R, respectively, ensuring stable and reliable prediction performance across all scenarios.

Figure (4) illustrates the relationship between e_{max} and transmitted data reduction for the three sensor nodes.

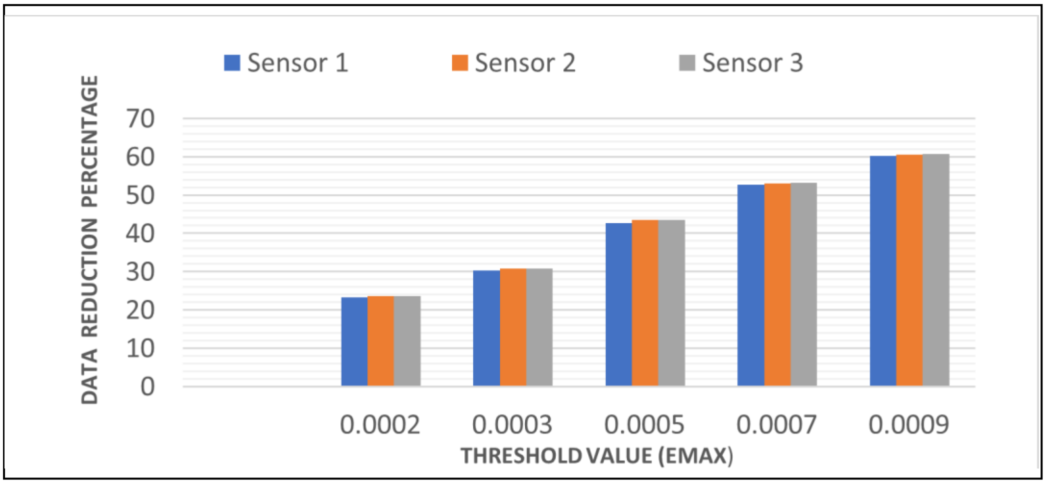

Figure (4) Threshold variation effect on data reduction percentage

5. Data Reconstruction Stage

After reducing data transmission at the sensor node, the sink node is responsible for reconstructing the missing readings by applying the same Kalman Filter used during the data reduction stage. Since both the sensor and sink nodes share the same prediction logic, the sink can accurately estimate the skipped values based on previously received data. This stage is essential to restore the original data sequence as closely as possible. It ensures that the system maintains reliable data quality while significantly lowering energy consumption.

5.1 Single Node-Based Reconstruction with Static EKF

The experiment involved reconstructing sensed data from each sensor node independently, without relying on communication or data exchange with neighboring nodes. Static Extended Kalman Filter (SEKF) was applied with process noise set to 0.03, and the measurement noise set to 0.16. The results of the Kalman Filter algorithm are presented in Tables (2), (3), and (4), respectively below. These tables show the results in terms of the percentage of reduced transmitted data and prediction accuracy after data reconstruction. The proposed method was compared with several other methods, and according to the results, the proposed method outperforms the other methods.

Table (2) Accuracy comparison of proposed approach, Reliable KF, DP_LSTM, DDR-IoT, and Least-Mean-Square LMS for sensor 1

Data Reduction% / **Methods**	23-25%	30-32%	42-44%	52-54%	60-62%
Proposed Method	98.37	**98.26**	**98.07**	**97.8**	**97.55**
Reliable KF	**98.71**	96.23	95.16	94.5	94.56
DP_LSTM	74.12	63.45	56.81	48.26	45.75
DDR-IoT	86.83	75.22	77.12	69.41	68.56
LMS	88.42	75.67	74.51	71.92	72.49

Table (3) Accuracy comparison of proposed approach, Reliable KF, DP_LSTM, DDR-IoT, and Least-Mean-Square LMS for sensor 2

Data Reduction% / Methods	23-25%	30-32%	42-44%	52-54%	60-62%
Proposed Method	98.48	98.38	98.17	98.004	97.86
Reliable KF	99.06	98.52	97.85	97.18	97.18
DP_LSTM	76.79	67.31	55.09	45.75	45.75
DDR-IoT	86.13	82.36	77.43	68.56	68.56
LMS	88.37	77.09	75.72	72.49	72.49

Data Reduction% / Methods	23-25%	30-32%	42-44%	52-54%	60-62%
Proposed Method	**98.44**	**98.28**	**98.002**	**97.82**	**97.67**
Reliable KF	98.8	97.66	96.07	94.56	94.56
DP_LSTM	74.12	64.23	56.81	48.26	45.75
DDR-IoT	86.83	82.77	77.12	69.41	68.56
LMS	88.42	76.39	74.51	71.92	72.49

Table (4) Accuracy comparison of proposed approach, Reliable KF, DP_LSTM, DDR-IoT, and Least-Mean-Square LMS for sensor 3

The obtained results for sensor 1 showed that the proposed approach achieved the highest accuracy (97.55%) with highest data reduction ratio (60-62%), demonstrating its robustness. In comparison, (Reliable Kalman Filter) Reliable KF maintained an accuracy of 94.56%, while Differential Privacy _Long Short Term Memory (DP_LSTM), Double Data Rate _Internet Of Things (DDR-IoT), and Least Mean Squares Algorithm (LMS) achieved only 45.75%, 68.56%, and 72.49%, respectively. For Sensor 2, the results show the superior performance of the proposed method under various levels of data reduction. At the highest reduction level (60%_62%), the proposed approach achieved an accuracy of 97.67%, by comparison, Reliable KF maintained an accuracy of 94.56%, while DP_LSTM, DDR-IoT, and LMS systems achieved much lower accuracies of 45.75%, 68.56%, and 72.49%, respectively. The obtained results for sensor 3 showed, the proposed method achieved high prediction accuracy for all data reduction percentage. At the highest reduction percentage (60-62%), the method maintained an accuracy of 97.86%, significantly outperforming competing methods. A reliable KF approach achieved 97.18%, while DP_LSTM dropped to 45.75, DDR-IoT and LMS both achieved 68.56% and 72.49%, respectively. Across the other reduction ranges from (23-25%) to (52-54%) the proposed method maintained accuracy values above 97% for all sensors, outperforming all other methods.

Figures (5) _ (7) illustrates the prediction accuracy of the sensors 1-3 across different data reduction percentage, comparing the Proposed Method with (LMS, DDR-IoT, DP-LSTM, and Reliable KF), respectively.

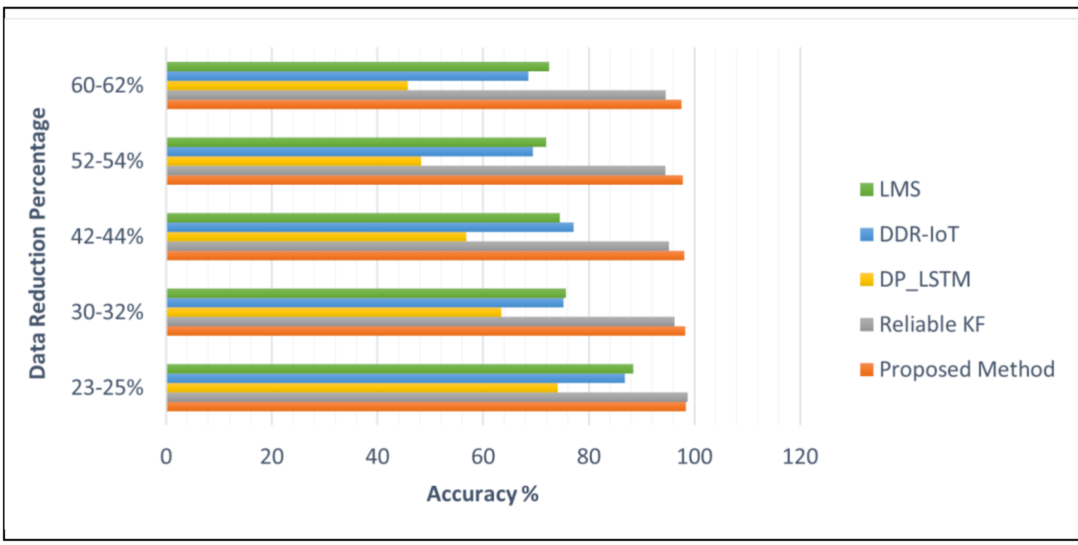

Figure 5: Accuracy _Reduction comparison for sensor 1

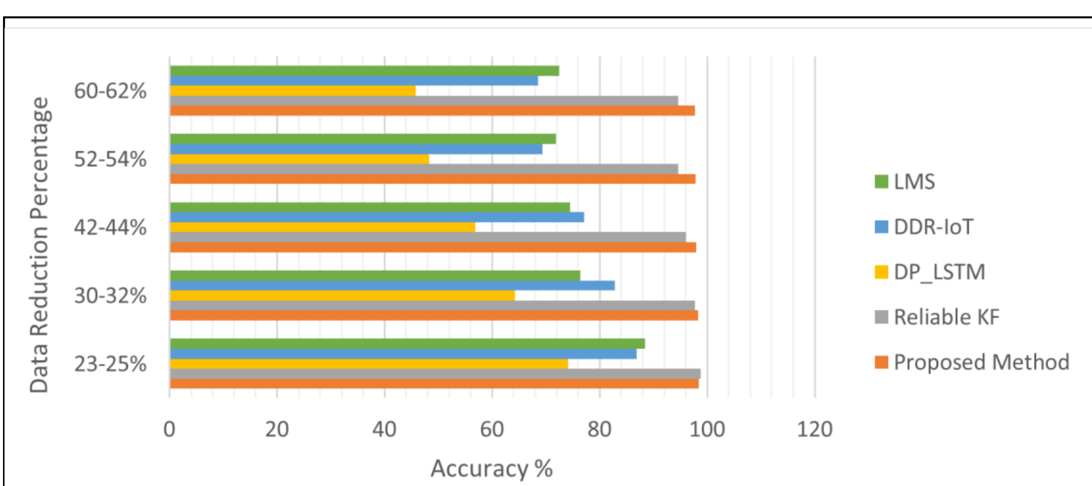

Figure 6: Accuracy _Reduction comparison for sensor 2

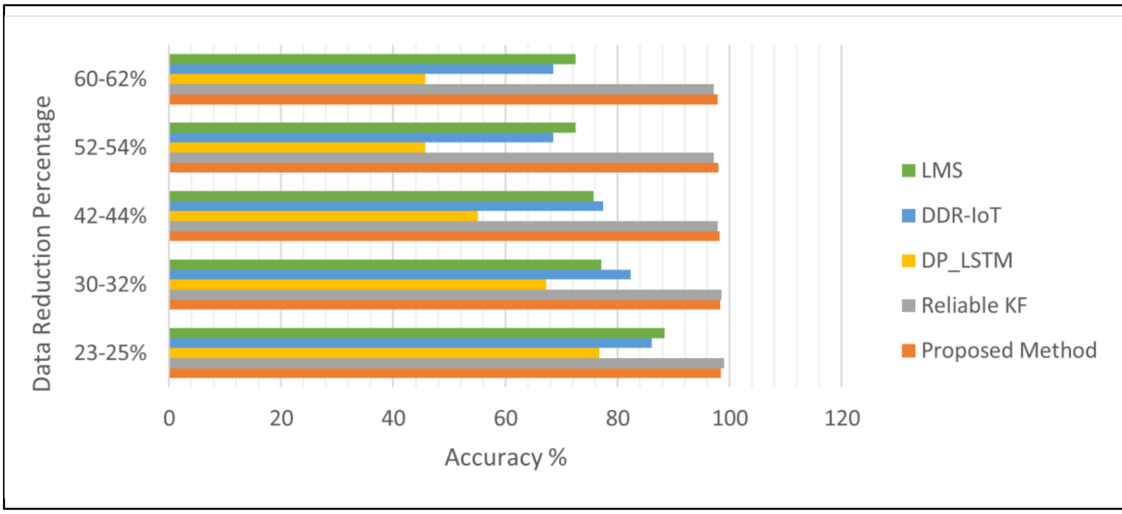

Figure 7: Accuracy _Reduction comparison for sensor 3

Figures (8) _ (10) show the model prediction accuracy , for sensor (1), sensor (2), and sensor (3)

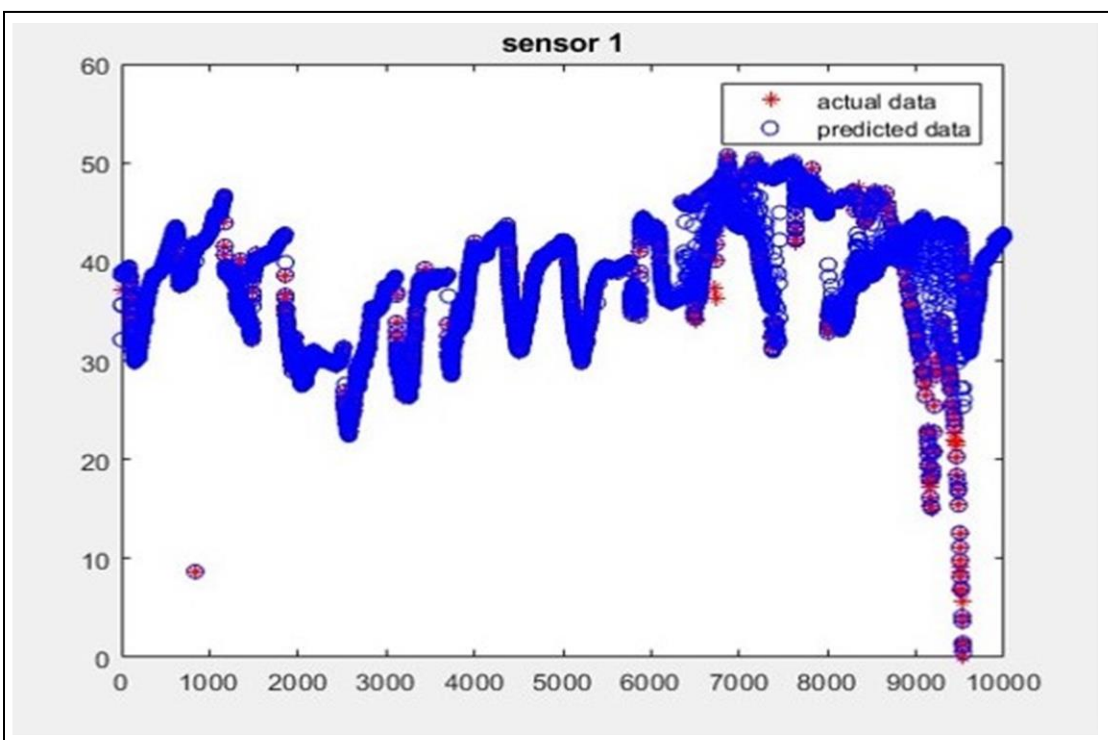

Figure 8: Proposed model data predictions for sensor 1

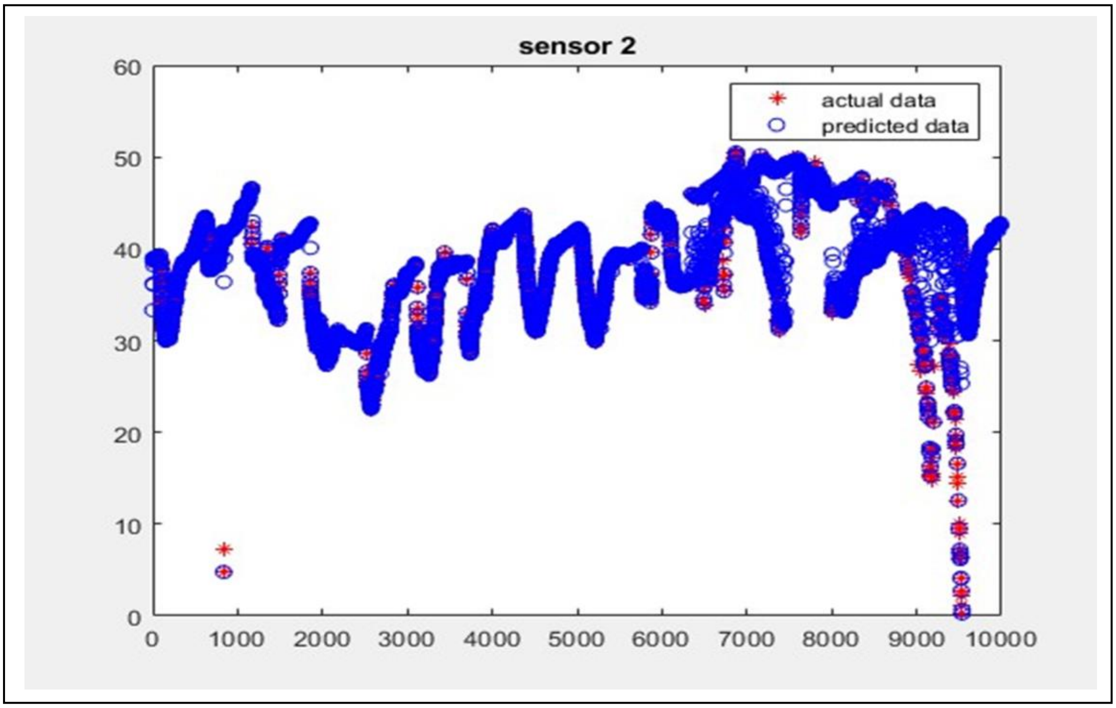

Figure 8: Proposed model data predictions for sensor 2

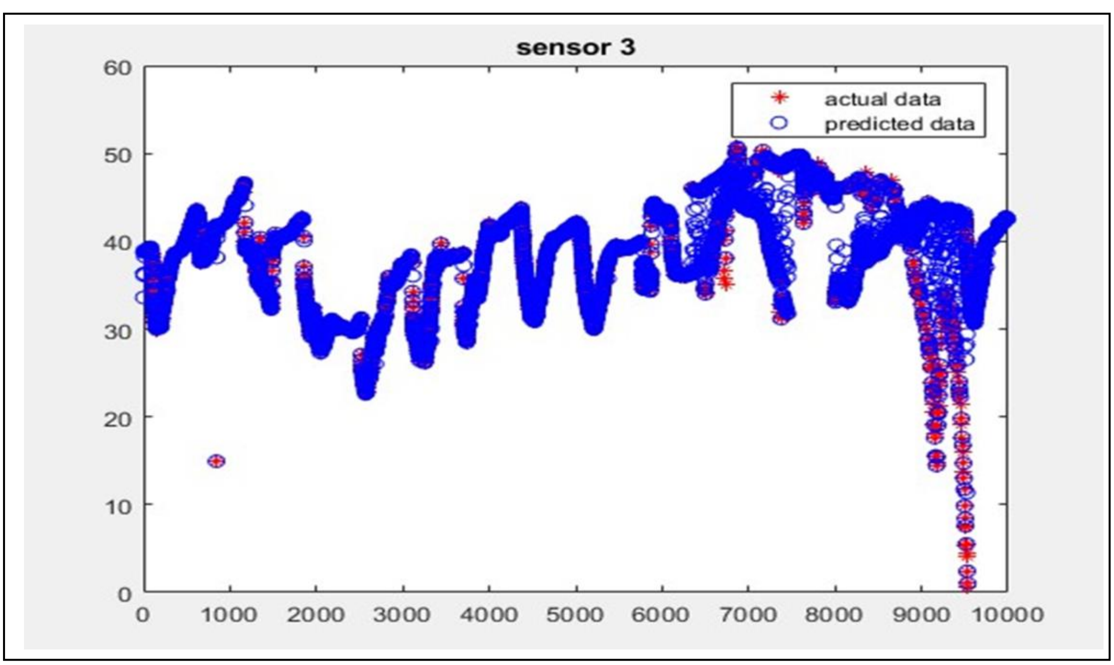

Figure 9: Proposed model data predictions for sensor 3

6.Wireless Sensor Network (WSN) Energy Consumption Analysis

Due to the limited resources of sensor nodes, energy consumption is critical in these devices the proposed model focuses on reducing waste and enhancing communication and processing efficiency. This section provide system evaluation in energy point of view, with comparison to other existing models.

6.1 Power Consumption with Single Node

The comparison is based on unified data reduction percentage ranges. Power consumption is calculated from the number of transmitted readings. Tables (5-7) presents a power consumption comparison with existing work, for sensors 1-3 across various data reduction levels.

TABLE (5) Power Consumption with Single Node (Sensor 1)

Data Reduction %	Proposed Method (Power J)	Reliable KF (Power J)	DP_LSTM (Power J)	DDR-IoT (Power J)	LMS (Power J)
23-25%	7,671J	7,658J	8,160J	8,439J	9,508J
30-32%	6,925J	6,964J	6,502J	6,729J	7,969J
42-44%	5,658J	5,739J	5,161J	5,346J	6,351J
52-54%	4,682J	4,699J	3,805J	4,682J	5,293J
60-62%	3,928J	3,949J	3,085J	3,809J	4,466J

TABLE (6) Power Consumption with Single Node (Sensor 2)

Data Reduction %	Proposed Method (Power J)	Reliable KF (Power J)	DP_LSTM (Power J)	DDR-IoT (Power J)	LMS (Power J)
23-25%	7,648J	7,508J	8,160J	8,439J	9,508J
30-32%	6,925J	6,964J	6,502J	6,729J	7,969J
42-44%	5,658J	5,739J	5,161J	5,346J	6,351J
52-54%	4,682J	4,699J	3,805J	4,682J	5,293J
60-62%	3,928J	3,949J	3,085J	3,809J	4,466J

TABLE (7) Power Consumption with Single Node (Sensor 3)

Data Reduction %	Proposed Method (Power J)	Reliable KF (Power J)	DP_LSTM (Power J)	DDR-IoT (Power J)	LMS (Power J)
23-25%	7,640J	7,501J	8,328J	8,649J	9,473J
30-32%	6,929J	4,926J	6,222J	6,452J	7,369J
42-44%	5,652J	3,305J	4,874J	5,152J	6,452J
52-54%	4,682J	2,895J	3,805J	4,682J	5,293J
60-62%	3,928J	2,425J	3,085J	3,809J	4,466J

The obtained results indicate that the proposed method achieves lower energy consumption compared to the LMS algorithm across all levels of data reduction. Moreover, the proposed approach outperforms both the DP-LSTM and DDR-IoT methods specifically at data reduction level (23% -25%). For sensor 1 and sensor 2, the proposed method exhibits performance that is nearly equivalent to that of the Reliable KF method across all reduction levels. Except for these cases, the Reliable KF, DP_LSTM, and DDR-IoT methods show slightly better energy savings than the proposed method. However, this reduction in energy consumption comes at the cost of estimation accuracy and result reliability. As shown in the accuracy evaluation tables in Section 5.1, the lower energy usage of those methods correlates with a noticeable decline in performance metrics, highlighting their limited capability to preserve result reliability during aggressive data reduction. In contrast, the proposed method demonstrates a more effective trade-off between energy efficiency and estimation fidelity. It maintains high data quality and prediction accuracy while still offering considerable energy savings. This balance substantially enhances the overall efficiency of the system, ensuring optimized resource utilization without compromising the required level of performance.

7.Conclusion and Future Work

A dual prediction data reduction strategy for WSNs is put out in this study. There are two stages to the suggested data reduction strategy. Data reduction is the focus of the first stage. founded on four methods: data equality, data deviation computation, faulty data detection, and data reduction based on prediction. The second stage estimates the filtered-out data from the sensor nodes using the Kalman filter, for reconstructed it which increases data reliability. Reducing transmissions while maintaining data dependability and correctness is the primary goal of the suggested strategy. The findings obtained demonstrated effective performance for model with a 60.72% data throughput reduction achieved by the SEKF model based on a single sensor node during the first stage. In the second stage, the data was reconstructed with an accuracy of 97.86% and a data reduction rate of 60.62% with an energy conception of 3,928 J . The suggested method finds and removes inaccurate data. Using 10,000 humidity real-world data points, the suggested method is contrasted with three distinct data prediction-based data reduction techniques, DP_LSTM, DDR-IoT, and LMS. Based on the results, the suggested method has the maximum efficiency in terms of energy usage, data accuracy, and data reduction. Rebuilding the missing data that might arise from a network failure will be the focus of future development.

REFERENCES:

[1] D. Puccinelli and M. Haenggi, "Wireless sensor networks: Applications and challenges of ubiquitous sensing," *IEEE Circuits Syst. Mag.*, vol. 5,

no. 3, pp. 19–29, 2005, doi: 10.1109/MCAS.2005.1507522.

[2] G. M. Dias, B. Bellalta, and S. Oechsner, "A survey about prediction-based data reduction in wireless sensor networks," *ACM Comput. Surv.*, vol. 49, no. 3, 2017, doi: 10.1145/2996356.

[3] J. M. Bahi, A. Makhoul, and M. Medlej, "A two tiers data aggregation scheme for periodic sensor networks," *Ad-Hoc Sens. Wirel. Networks*, vol. 21, no. 1–2, pp. 77–100, 2014.

[4] T. Du, Z. Qu, Q. Guo, and S. Qu, "A High Efficient and Real Time Data Aggregation Scheme for WSNs," *Int. J. Distrib. Sens. Networks*, vol. 2015, 2015, doi: 10.1155/2015/261381.

[5] H. Wu, M. Suo, J. Wang, P. Mohapatra, and J. Cao, "A Holistic Approach to Reconstruct Data in Ocean Sensor Network Using Compression Sensing," *IEEE Access*, vol. 6, pp. 280–286, 2017, doi: 10.1109/ACCESS.2017.2753240.

[6] A. Basheer and K. Sha, "Cluster-based quality-aware adaptive data compression for streaming data," *J. Data Inf. Qual.*, vol. 9, no. 1, 2017, doi: 10.1145/3122863.

[7] J. M. C. Silva, K. A. Bispo, P. Carvalho, and S. R. Lima, "LiteSense: An adaptive sensing scheme for WSNs," *Proc. - IEEE Symp. Comput. Commun.*, pp. 1209–1212, 2017, doi: 10.1109/ISCC.2017.8024689.

[8] U. Raza, A. Camerra, A. L. Murphy, T. Palpanas, and G. Pietro Picco, "Practical data prediction for real-world wireless sensor networks," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 8, pp. 2231–2244, 2015.

[9] B. Chreim, J. Nassar, and C. Habib, "RADAR - Regression Based Energy-Aware DAta Reduction in WSN: Application to Smart Grids," *Lect. Notes Networks Syst.*, vol. 226 LNNS, pp. 1–14, 2021, doi: 10.1007/978-3-030-75075-6_1.

[10] M. Lewandowski and B. Płaczek, "Data transmission reduction in wireless sensor network for spatial event detection," *Sensors*, vol. 21, no. 21, pp. 1–21, 2021, doi: 10.3390/s21217256.

[11] 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 Commun.*, vol. 15, no. 18, pp. 2285–2299, 2021, doi: 10.1049/cmu2.12262.

[12] K. Jain, A. Agarwal, and A. Abraham, "A Combinational Data Prediction Model for Data Transmission Reduction in Wireless Sensor Networks," *IEEE Access*, vol. 10, pp. 53468–53480, 2022, doi: 10.1109/ACCESS.2022.3175522.

[13] A. N. Njoya *et al.*, "Data Prediction Based Encoder-Decoder Learning in Wireless Sensor Networks," *IEEE Access*, vol. 10, no. September, pp. 109340–109356, 2022, doi: 10.1109/ACCESS.2022.3213671.

[14] W. M. El-Sayed, H. M. El-Bakry, and S. M. El-Sayed, "Integrated data reduction model in wireless sensor networks," *Appl. Comput. Informatics*, vol. 19, no. 1–2, pp. 41–63, 2023, doi: 10.1016/j.aci.2019.03.003.

[15] M. Revanesh, S. S. Gundal, J. R. Arunkumar, P. J. Josephson, S. Suhasini, and T. K. Devi, "Artificial neural networks-based improved Levenberg–Marquardt neural network for energy efficiency and anomaly detection in WSN," *Wirel. Networks*, vol. 30, no. 6, pp. 5613–5628, 2024, doi: 10.1007/s11276-023-03297-6.

[16] A. A. Hammad, S. R. Ahmed, M. K. Abdul-Hussein, M. R. Ahmed, D. A. Majeed, and S. Algburi, *Deep Reinforcement Learning for Adaptive Cyber Defense in Network Security*, vol. 1, no. 1. Association for Computing Machinery, 2024. doi: 10.1145/3660853.3660930.

[17] G. H. A. Alshmeel, A. S. B. Al-Doori, S. R. Ahmed, Z. A. Abrahim, A. J. Ghaffoori, and A.-S. T. Hussain, "Self-sustaining buoy system: harnessing water wave energy for smart, wireless sensing and data transmission," in *Proceedings of the Cognitive Models and Artificial Intelligence Conference*, 2024, pp. 349–356.

[18] D. A. Majeed, M. Yaseen, A. J. Ghaffori, S. R. Ahmed, B. A. Hakim, and A.-S. T. Hussain, "Advanced optimization techniques for CNC machining: a comparative study of genetic algorithms, simulated annealing, and particle swarm optimization," in *2024 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, IEEE, 2024, pp. 1–6.

[19] P. Bodik, W. Hong, C. Guestrin, S. Madden, M. Paskin, and R. Thibaux, "Intel lab data," *Online dataset*, vol. 90, 2004.