ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
https://doi.org/10.17559/TV-20240920001999
Original scientific paper

Enhanced Data Prediction and Compression in Wireless Sensor Networks Using Bidirectional LSTM with Offset Gaussian Modified Least Mean Square and Renyi-Entropy PCA

VINI ANTONY GRACE N.*, NADA ALZABEN, FAIZ ABDULLAH ALOTAIBI, AHMAD A. ALZAHRANI

Abstract: This paper addresses the critical challenge of energy consumption in Wireless Sensor Networks (WSN), which are often deployed in remote areas with limited battery replacement options. To enhance the network's lifetime, a novel approach that combines two key techniques: the Offset Gaussian Modified Least Mean Square (OGMLMS) filter for data prediction and the Renyi-Entropy Principal Component Analysis (PCA) for data compression is proposed. The OGMLMS filter predicts future data values based on historical data, reducing the amount of data that needs to be transmitted. Subsequently, the Renyi-Entropy PCA compresses the predicted data, minimizing the energy required for transmission. The proposed method is implemented using MATLAB and validated with three univariate datasets. The findings highlight substantial enhancements in performance metrics, including Mean Squared Error (MSE), energy efficiency, transmission costs, and compression ratio. These improvements surpass those achieved by traditional algorithms such as Principal Component Analysis (PCA), Least Mean Square (LMS), Auto-Regressive Integrated Moving Average (ARIMA), and hybrid approaches like LMS combined with Renyi PCA, which often struggle with high energy consumption and inadequate data management, leading to reduced network lifetimes. The proposed method effectively integrates data prediction and compression techniques, enhancing energy efficiency and data transmission while maintaining high data quality. The results clearly indicate a notable reduction in energy consumption by up to 14.457%, along with an impressive 99% compression ratio. This enables sensor nodes to operate longer on limited battery resources, making it highly beneficial for remote monitoring applications such as environmental tracking and disaster management. Additionally, the reduced data volume leads to lower transmission costs.

Keywords: data compression; data prediction; wireless sensor networks (WSNs); energy consumption

1 INTRODUCTION

WSN is a network responsible for sensing, processing and communicating wireless data to the intended base station. Since these sensors are usually deployed at remote sites, despite the recent advances in the WSN technology, its applications still face significant challenges. In a typical WSN, each sensor node is equipped with a sensing unit for acquiring data from the environment, a microcontroller for data processing and storage, a transceiver unit for transmission/reception of data of/from other connected devices and a power unit or a battery that empowers all components of a sensor node. The different components that are inbuilt in a sensor node are shown in Fig. 1 given below. The sensor node may vary in size according to specific applications.

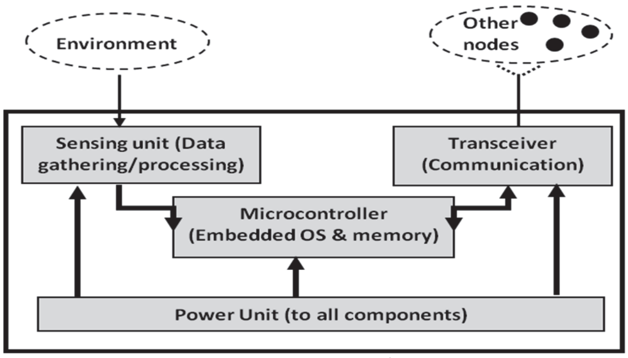

Figure 1 Hardware components of a sensor node

Energy consumption is considered to be the major constraint in WSN. The energy is consumed in a network during three phases, namely sensing, processing of data, and data transmission. Usually, most of the energy is depleted during the transmission of data. The networks lifetime can be increased by reducing the energy consumption. Also, the scope of WSNs is restricted to features like communication range, bandwidth and storage. However, the lifetime of the network can be increased, if the WSN is designed in such a way that it is power efficient and reduced in size

The nodes in the WSN have a limited power source as they are powered by batteries of a very small size. Major applications of WSN require battery to be operated for a long period of time after initial deployment. Each battery supplying energy to the sensor for acquiring data at remote location will discharge at a high rate for transmitting all the sampled data to the monitoring station [1]. The battery of a node will be completely drained in two days if no energy efficient techniques are being employed. Thus, it becomes a big challenge to operate, maintain and manage the data acquisition and storage systems in application scenarios.

In this paper, the data prediction and data compression techniques are combined for reducing the energy consumption. It aims to develop an energy efficient method that not only maintains a good balance between compression ratio and mean square error, but also reduces energy consumption further in comparison with some other standard algorithms available in literature. This is done by combining Offset Gaussian Modified Least Mean Square (OGMLMS) filter for prediction phase and the Renyi-entropy Principal Component Analysis (PCA) algorithm for data compression.

1.1 Organization of the Paper

This paper is sectioned as follows. Section 2 reports the review of literature, in which the pros and cons of the related works are analyzed. Section 3 expounds some of the popular existing data prediction and data compression algorithms available in literature. Section 4 elaborates the techniques employed in the proposed method for efficient energy consumption. In section 4 the results obtained with the proposed method are compared with the standard baseline algorithms. Section 5 gives conclusive remarks about the efficacy of the proposed algorithm.

2 LITERATURE REVIEW

Data reduction techniques aim to reduce the amount of sampled data, thereby the volume of data transmitted by the source nodes is reduced. This way of decreasing the transmission of unneeded samples, in turn reduces the consumption of energy and increases the networks life time.

A major scheme in the development of lossy data compression relies on the transformation of raw data into other data domains. Examples of these methods include Discrete Fourier Transform (DFT), Fast Fourier Transform (FFT) Mohsen, K. K. et al. (2022)[2] and the different types of Discrete Cosine Transforms (DCT) Bennani, M. T. et al. (2023)[3]. Naturally, these transformation methods exploit the knowledge of the application to choose the appropriate data domain that discards the least data content. However, such algorithms suffer from low performance when used to compress data spatially or when noises are present in the collected readings.

El-Bably, D. L. et al. (2022) [4] employed Adaptive-PCA for data gathering in WSN. In this work, PCA is dynamically implemented on the data sensed by the nodes. In the conventional PCA based data reduction process the PCA is implemented during normal states. During the emergency states, the sensed data dynamically alters and they are transmitted directly to the sink node without compression. In Surenthar, L. et al. (2023) [5] PCA is used to detect outliers and faulty nodes by performing offline and real time analysis.

PCA, also known as the Karhunen-Loeve transform, Oluwadara J. et al. (2020)[6], William P et al. (2022) [7] and Liu, Y. et al. (2022)[8] has been widely used to extract (linear) correlations among sensor nodes. El-Sayed et al. (2023) [9] have used the idea of time series analysis for predicting the present data from the past history. They used the Auto Correlation Function (ACF) plot for choosing the correct order which is a major concern in model selection. The experimental evaluations on performance measures like mean error, mean square error and root mean square error prove the efficiency of this model.

A transmission technique was presented by Kavitha, A. et al. (2023) [10] using Hybrid LMS (H-LMS) filter to predict the values at the sensor nodes and the sink node. The transmission was done by removing redundancy depending on the error estimate. Thus, the technique could conserve the nodal energy, offering better convergence rate. However, the selection of error threshold is risky, as higher value degrades the performance.

Jain K. et al. (2023) [11] had designed a framework by combining data prediction, recovery, and compression, for the simultaneous achievement of accuracy and efficiency of the processed data in clustered WSNs such that the communication cost was reduced. A Dual Prediction algorithm called Least Mean Square (LMS) was implemented for data prediction based on the optimally selected step size by reducing the Mean-Square Derivation (MSD).

This extensive literature review reveals that the different energy efficient methods designed in the recent past, even now face limitations in energy consumption. Efficient algorithms with less energy consumption have to be developed in order to increase the lifetime of the WSN network.

3 BASELINE ALGORITHMS

Before arriving at the perspectives that have to be considered before adopting a data reduction scheme in WSN, it becomes essential to have a detailed study about the currently available algorithms in literature.

3.1 Principal Component Analysis (PCA)

Principal component analysis is a multivariate statistical model and dimensionality reduction data analysis technique. It aims to extract important information from data using orthonormal transformations and present it as a set of new orthogonal variables called Principal Components (PCs) [12]. This algorithm helps in reducing the amount of data being transmitted among sensor nodes by selecting only significant principal components and discarding other lower order insignificant components from the model.

3.2 Least Mean Square (LMS)

LMS algorithm is a simple adaptive filter-based time series forecasting scheme [13]. It is used to predict the future values of a data series and the filter parameters are adapted in conjunction with reducing the mean square error (MSE) of the prediction process. This algorithm exploits more energy as it has slow convergence rate.

3.3 Auto Regressive Integrated Moving Average (ARIMA)

ARIMA is one among the various time series models, that are used for data prediction. ARIMA forecasting model is a stationary process which is a broadening class of ARMA model [14]. It is denoted by ARIMA (p, d, q). It estimates the forthcoming data values from the auto correlated time series. This method has higher prediction accuracy compared to AR, MA and ARMA methods, but the complexity is more as it combines both AR and MA models.

3.4 Least Absolute Shrinkage and Selection Operator (Lasso) Regression

Lasso is a regression analysis method that performs simultaneous variable selection and parameter estimation (shrinkage) to improve the prediction accuracy of the statistical model [15]. This is done by minimizing the Residual Sum of Squares (RSS). Lasso outperforms by shrinking the insignificant coefficients to zero, at the same time reducing the magnitude of significant variables [16].

These existing works target either on minimizing the expected number of transmissions between the source and sink by employing prediction model or try to reduce the size of data to be transmitted to sink by data compression methods.

This analysis of existing techniques has paved an idea of integrating both data prediction and data compression techniques so that very less energy is consumed. Thus, apart from comparing the performances of the above explained individual algorithms, this paper also compares the performance of the proposed technique with the integrated ARIMA_Lasso & LMS_RenyiPCA to prove its efficacy.

4 PROPOSED METHODOLOGY

The deployed source nodes of the network sense the data and sends to the sink node at the very first time slot. OGMLMS based data prediction model is developed and employed in both source and the sink nodes. The measured data and the predicted data at the consecutive time slots are compared. The data deviation is compared with the predefined error threshold (δ) value and the data are transmitted to the sink node if the deviation is above the threshold value. While transmitting data to the sink node, the dimension of the data is reduced by using the compression algorithm. The work flow of the proposed scheme is given Fig. 2.

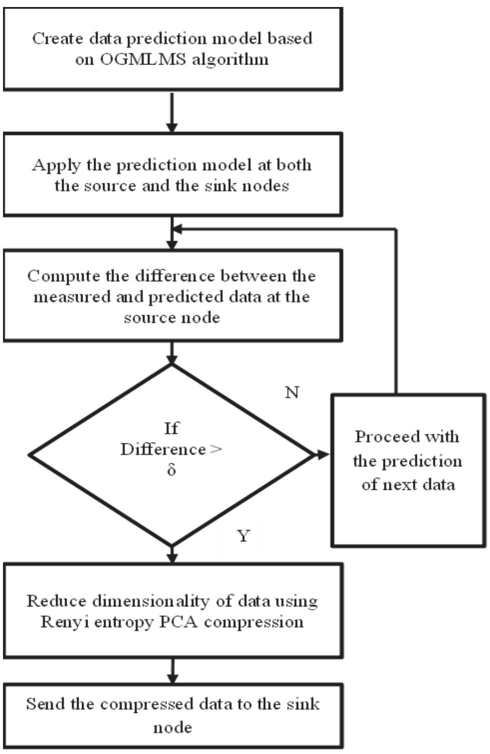

Figure 2 Proposed work strategy involved in OGMLMS prediction combined with RenyiPCA compression

4.1 OGMLMS Filter for Prediction

The Offset Gaussian Modified Least Mean Square (OGMLMS) filter is designed to enhance the efficiency of data prediction in Wireless Sensor Networks (WSNs). Its primary role is to forecast future sensor readings based on historical data, which is crucial for minimizing energy consumption and optimizing data transmission. In WSNs, where sensor nodes are often deployed in remote locations with limited battery life, the OGMLMS filter helps to reduce the volume of data that needs to be sent over the network. By accurately predicting data points, the filter allows nodes to transmit only significant changes or differences from predicted values, thereby conserving energy and extending the operational lifespan of the network.

At first the operation of conventional LMS filter is discussed here [17]. If the given data series is denoted as $d(t)$ at an instant t, then the series predicted by the LMS filter is given as in Eq. (1) where $p.d(t)$ represents the predicted data series at instant t. The length of the series is equal to the number of samples which is denoted by L. $W^T(t)$ denotes the weight vector. Thus, the predicted data series can be obtained as the linear combination of data series $d(t)$ of L samples weighted by its corresponding vector $W^T(t)$.

$$p.d(t) = W^T(t)d(t) \tag{1}$$

Then predicted data sequence $p.d(t)$ is compared with the data series which is considered for adaptation by the filter and the error $e(t)$ is obtained as in Eq. (2).

$$e(t) = d(t) - p.d(t) \tag{2}$$

For arriving at small MSE in prediction process, the filter weights associated with the data series $d(t)$ must be adjusted. Thus, the adaptive weight filter can be obtained. The adaptation of weights can be accomplished using a suitable adaptation algorithm. The minimum MSE ensures the accuracy of the system. Fig. 3 shows the basic design of LMS filter when using in prediction.

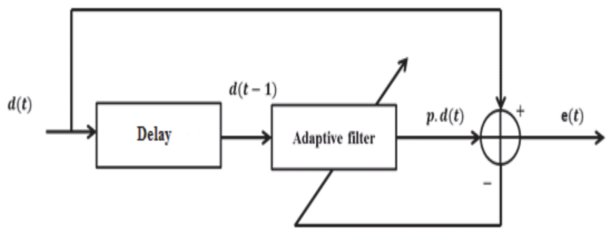

Figure 3 Structure of LMS filter for data prediction application

Another parameter that needs to be altered in the LMS filter is filter step size, which helps in obtaining accurate prediction and energy efficiency by providing faster convergence [18]. The condition for the value of step size is that the higher step size needs to be assigned if the deviation is more.

Therefore, the convergence can be arrived at quickly. When the convergence point is very close, smaller step size is assigned so that the balanced prediction can be obtained. Thus, the step size is optimally chosen based on the state of prediction for obtaining faster convergence as well as less MSE. Eq. (3) shows the weight adaptation in LMS filter.

$$W(t+1) = W(t) + s \times p.d(t) \times (t) \tag{3}$$

where W and $p.d$ are obtained from Eq. (4) and Eq. (5).

$$W(l) = \left(W_1(l).W_2(l)\ldots W_n(l)\right)^T \tag{4}$$

$$p.d(L) = \left(p.d(l-1) \times p.d(l-2)\ldots \times p.d(l-)\right)^T \tag{5}$$

The data can be continuously predicted by applying a delay unit with the measured data $d(t)$ of the present instant. The delay unit makes the data to be lagged by a single time period. Then the filter computes the predicted value $p.d(t)$ for the data $d(t)$. For obtaining this, a linear combination of L past readings is obtained. The error value is calculated by subtracting the predicted data from the desired data. This error value is given as feedback for changing the filter weight.

The two essential parameters that are required for updating of filter weights are the filter length and the step size parameter s. The step-size parameter s is significant for the convergence of the algorithm. The criteria for selecting the value for step-size s, are given below.

- A larger value for step size of s is chosen until certain numbers of good predictions are obtained. The order of s is chosen as two for obtaining robust system.
- The stable value for step size can be determined when the type of received data is adequately learned.

The step size parameter s can be made to converge faster by using the Offset Gaussian function ($OG(x)$) as in Eq. (6).

$$s_{\text{new}} = \frac{OG(s_{\text{old}})}{L} \tag{6}$$

$$OG(s_{\text{old}}) = G_\sigma(\sin s_{\text{old}}) - G_\sigma(\cos s_{\text{old}}) \tag{7}$$

The general Gaussian function $G_\sigma(x)$ is given as in Eq. (8)

$$G_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-m}{\sigma}\right)^2} \tag{8}$$

$$s_{\text{old}} = 2E_x^{-1}10^{-2} \tag{9}$$

where E_x is the mean power as given in Eq. (10).

$$E_x = \frac{1}{N}\sum_{n=1}^{N}\left|d[n]\right|^2 \tag{10}$$

where N is the number of iterations required in training phase. In real-time applications, WSN uses identical filters for prediction at source node and sink node.

The OGMLMS filter is exceptionally well-suited for wireless sensor networks (WSNs) due to its adaptability, energy efficiency, simplicity, and compatibility with advanced techniques. Its ability to adjust weights in real-time enables accurate predictions in dynamic environments where sensor readings are influenced by noise or environmental changes. By reducing the volume of data transmitted, the filter significantly conserves energy, a crucial advantage for sensor nodes with limited battery power. Its straightforward implementation and low computational requirements make it ideal for resource-constrained hardware. Additionally, the filter enhances data quality by mitigating the effects of noise and outliers, ensuring reliable and valuable information reaches the sink node.

4.2 Renyi-Entropy PCA

The performance of PCA based compression is improved by involving Renyi entropy with the PCA technique [19].

Renyi PCA is utilized for dimensionality reduction and data compression in Wireless Sensor Networks (WSNs), which helps to minimize the transmission burden and conserve energy by effectively reducing the size of the data collected from sensor nodes, thereby enhancing the overall efficiency of data processing and storage within the network.

The data selection is a very important step of the algorithm because it influences the compression model accuracy. The selection criteria can use a random selection or an active selection. A popular active selection method, based on the density distribution of the sample, maximizes the quadratic Renyi entropy. Given a dataset of M data points with different probabilities $\{p_1, \dots p_M\}$, the Renyi entropy of order α $(\alpha > 0, \alpha \neq 1)$ is defined as:

$$H_\alpha(x) = \frac{1}{1-\alpha}\log\left(\sum_{m=1}^{M} p_m^\alpha\right) \tag{11}$$

It measures the degree of the set randomization. More random samples involve more large entropy. A case of particular interest is $\alpha = 2$, noted as quadratic Renyi entropy. If the probability density function is a Gaussian function, the quadratic Renyi's entropy can be calculated as follows.

$$H_2(x) - \log\left(\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} K(x_i, x_j)\right) \tag{12}$$

The quadratic Renyi entropy can be used as criteria to select a subsample of dataset based on the high value of the entropy.

4.2.1 Principal Component Analysis

PCA is a statistical technique used to reduce the size of the original data series by removing the redundant information from the analysis of correlation presenting among the values of the series [20]. In PCA technique the imperative task is to study the most significant variations of the values. The covariance method is one of such approaches used to implement PCA. A dataset $\mathcal{G}$ with dimensions $U \times V$ is considered, where U denotes the number of measurements of V variables and V denotes the number of variables. The dataset D can be reduced into a new subset with A variables where $A < U$. A matrix B representing the new subset of dimension $V \times U$ can be provided as in Eq. (13).

$$B = \mathcal{G} - hu^T \tag{13}$$

where $u[j]=\frac{1}{M}\sum_{i=1}^{M}\mathcal{G}[i,j]$, $h[i]=1$, $j=1,\dots,V$ and $i=1,\dots,U$. The covariance matrix C of matrix B with dimension $V\times V$ can be given as in Eq. (14).

$$C=\frac{1}{U-1}B^{T}B \tag{14}$$

The D diagonal matrix $V \times V$ of eigenvalues of C is given as below.

$$D[k,1]=\begin{cases}\lambda_k \text{ for } k=1\\ 0 \text{ for } k\neq 1\end{cases} \tag{15}$$

where λ_k denotes the k-th eigenvalue of the covariance matrix C. The $V\times V$ eigenvector matrix E that diagonalizes the covariance matrix C can be as in Eq. (16).

$$E^{-1}CE=D \tag{16}$$

The eigenvector matrix E has V column vectors that denote V eigenvectors of the covariance matrix C. Then the columns of the matrix V are arranged in the order of decreasing eigenvalue and the new matrix F can be obtained with the columns elements as the first A columns of E as:

$$F[k,1]=E[k,1] \tag{17}$$

with $k=1,..V$ and $a=1,\dots A$ where $A<V$. Hence the dataset $\mathcal{G}$ can be transformed into a new U-by-A space by:

$$\bar{\mathcal{G}}=BF \tag{18}$$

In the PCA the suitable number of the components is chosen to ensure the accuracy of the compression algorithm.

4.2.2 Application of PCA in Data Compression

The dimensionality reduction compression method, as explained previously, can be applied to the dataset available at the sensor nodes. The original dataset is a N-by-P matrix, where N rows represent the observations of P monitored parameters along with the predicted data. The dimensionality reduction of the training set consists to select M rows ($M\times N$) and R columns ($R\times P$) starting from the original data matrix without loss of the information. The value M that represents the number of the data is freely chosen by the user. The maximum value of M that can be used depends on the computational resources and the features of the dataset [21]. Fixed M, the quadratic Renyi entropy criterion is used to select a subsample of dataset. So, a new data matrix is obtained of dimension $M\times P$, whose rows represent the Perfect Vector (PVT).

Subsequently in order to reduce further the size of each data, the PCA has been performed, eliminating the principal components that contribute less than 2% to the total variation in the data set. The identification procedure of the PVT in combination with the decomposition through the principal components produces a new dataset of reduced dimension $M\times R$ that is sent to the sink node.

Algorithm-Renyi PCA

Given a training set $\mathcal{D}_N=\{x_k,y_k\}_{k=1}^{N}$ randomly choose a working set $S_M=\{x_i,y_i\}_{i=1}^{M}$, $S_M\subset\mathcal{D}_N$

Randomly select a sample point $\vec{x}\in S_M$ and $\hat{x}\in\mathcal{D}_N$ swap $(\vec{x},\hat{x})$

If $H_2(x_1\dots x_{M-1}\dots\hat{x})>H_2(x_1\dots\hat{x}\dots x_M)$ then $\hat{x}\in S_M$ and $\vec{x}\notin S_M$, $\vec{x}\in\mathcal{D}_N$
Else $\hat{x}\notin S_M$ and $\vec{x}\in S_M$, $\vec{x}\in\mathcal{D}_N$
End if

Calculate H_2 for the present S_M by Eq. (12)

Stop if the change in entropy value is small

Figure 4 RenyiPCA algorithm

Its ability to capture complex patterns efficiently aligns well with the resource-constrained and data-intensive nature of WSNs, ensuring more meaningful data compression, enhancing data quality, and maintaining robustness against noise and outliers, which are common in sensor data, making Renyi PCA an ideal choice for optimizing data management in these networks.

The OGMLMS filter and Renyi PCA complement each other effectively in the context of Wireless Sensor Networks (WSNs) by working in tandem to enhance data quality and optimize data management. The OGMLMS filter preprocesses the raw data by improving its quality and reducing noise, which is crucial for ensuring that the data is reliable and representative of the actual conditions being monitored. This preprocessing step enables the Renyi PCA algorithm to perform more effective compression and prediction, as it can focus on the most informative components of the cleaned data without being adversely affected by noise or outliers. Consequently, the combination of the OGMLMS filter's noise reduction capabilities and Renyi PCA's advanced data compression techniques lead to a more efficient and effective approach for managing the data in resource-constrained WSN environments, ultimately conserving energy and enhancing overall network performance.

5 PERFORMANCE ANALYSIS

5.1 Compression Ratio (CR)

The compression ratio [22] is described as the ratio between size of the compressed signal Y to the size of the actual uncompressed signal from sensor X.

$$CR=100\,X\left(1-\frac{Compressed\ Size(Y)}{Uncompressed\ size(X)}\right) \tag{19}$$

5.2 Mean Square Error (MSE)

Mean Squared Error (MSE) is a critical metric for evaluating the accuracy of predictive models in Wireless Sensor Networks (WSN). In the context of WSN, MSE measures the average squared difference between the predicted sensor readings and the actual observed values. A lower MSE indicates that the model's predictions are closer to the actual data, reflecting higher accuracy.

If the observed reading from the sensor is x and the predicted reading is y then the mean square error is the difference between the observed and predicted reading [23]. It is given by:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - y_i\right)^2 \tag{20}$$

5.3 Energy Consumption

The total energy consumed by a sensor node is contributed by the energy utilized for data compression and the energy required for transmitting the data expressed in joules. The energy requirements for data compression are mainly dependent on the energy required for one cycle $\left(E_{\text{opt}}\right)$ and number of CPU cycles (M) required for compression and the requirements by its peripherals are considered negligible. Energy required for transmission and receiving the data depends on the energy consumed by equipment to transmit a bit of data $\left(E_{\text{bit}}\right)$ and the number of bits (N). The metric for estimating the energy consumption is as follows:

$$E_{\text{total}} = M \times E_{\text{opt}} + N \times E_{\text{bit}} \tag{21}$$

5.4 Transmission Cost

The transmission cost in WSN is the calculation of total number of messages transmitted from source to sink. If the number of messages transmitted is more, then the transmission cost is more and vice versa.

6 EXPERIMENTAL RESULTS

Assessment metrics that are used to quantify the effectiveness of the proposed algorithm are mean square error, compression ratio, energy consumption and transmission cost. The results of the proposed algorithm are compared with standard conventional algorithms available in literature like Principal Component Analysis (PCA), Least Mean Square (LMS), Auto Regressive Integrated Moving Average (ARIMA), ARIMA _ Lasso, LMS_ Renyi PCA combined together. The proposed work and the algorithms involved are implemented in MATLAB simulation software.

Three univariate datasets were considered to validate the proposed system. The univariate datasets referred in the study include time series data related to Temperature, Relative Humidity, and Microphone Data, which are commonly used in environmental monitoring and sensor applications. The first univariate dataset, published by Sensor Scope HES-SO Fish Net Deployment, includes temperature and relative humidity readings. The temperature values exhibit a relatively narrow range compared to the humidity values, and the time series data for both temperature and humidity demonstrate smooth trends. In contrast, another univariate dataset, the Microphone data, was collected using sensor nodes deployed at the Strata Clara Convention Center. This dataset is characterized by numerous peaks and noise, making it a discontinuous (non-smooth) signal.

The datasets used for analysing the performance of the proposed algorithm are given in Tab. 1.

Table 1 Dataset used for analysis

Data	Samples	Sampling Interval	Dimensions
Temperature	14721	2 minutes	1
Relative Humidity	14721	2 minutes	1
Microphone Data	2887	4-9 second	1

To validate the proposed method, we conducted MATLAB simulations using R2013 software under a wireless sensor network (WSN) environment. The simulation setup includes the following components:

The simulation setup involves low-power microcontrollers, such as the MSP430, to manage operations and data processing. Temperature and relative humidity are measured using SHT15 sensors (or equivalent), with a resolution of ±0.3 °C and ±2% RH, respectively. A low-power, high-sensitivity microphone sensor is used to capture sound data. A ZigBee-based transceiver is used for low-power wireless communication. Sensor nodes are deployed at fixed intervals to form a mesh network, ensuring spatial coverage to capture variations in environmental conditions. These nodes are battery-powered, with solar panels for energy harvesting in remote areas. Data is collected every 2 minutes for temperature and humidity, and every 4-9 seconds for microphone readings, with temporary storage on an onboard SD card for redundancy. Data is transmitted wirelessly to a base station at predefined intervals to minimize energy consumption, and basic data aggregation is performed onboard to reduce transmission overhead.

The sensor nodes are configured with an initial energy of 2 Joules. The transmission energy is calculated based on the data size and the distance between the sensor node and the base station. Each sensor node transmits data packets of size 500 bytes, which contain readings for environmental parameters like temperature and relative humidity. These sensor nodes are strategically placed in the fishnet to collect accurate environmental data while optimizing energy usage. The simulation runs for 500 rounds to evaluate energy consumption and network lifetime.

Fig. 5 represents the lifetime of the nodes of the WSN for the different error threshold values for temperature dataset. When error threshold is less and equal to 0.25, all nodes vanish in 2300 rounds whereas when it is large and equal to 1.25, the nodes vanish in 3500 rounds. Thus, it is inferred that the lifetime of the nodes is high when the error threshold is high. The high value of error threshold makes only few data to involve in transmission and therefore less energy is consumed.

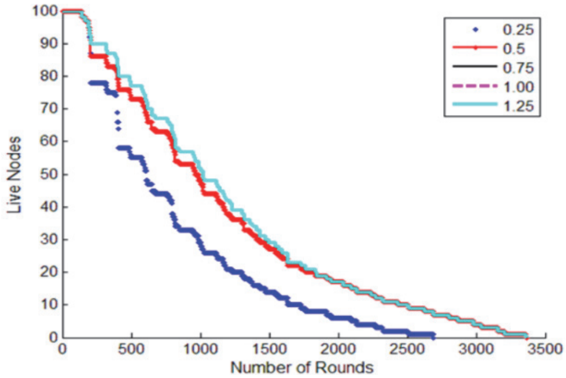

Figure 5 Lifetime of network for different error threshold for temperature dataset

Fig. 6 depicts comparative analysis chart for the MSE parameter observed for different methods by setting different error threshold values for all univariate dataset respectively. From the graph given below, it is clear that error in conventional PCA method is high which can be reduced with the utilization of Renyi PCA. Comparing with the conventional PCA compression work, the Renyi PCA reduces error by 34.24% in temperature dataset when error threshold is assigned as 0.25.

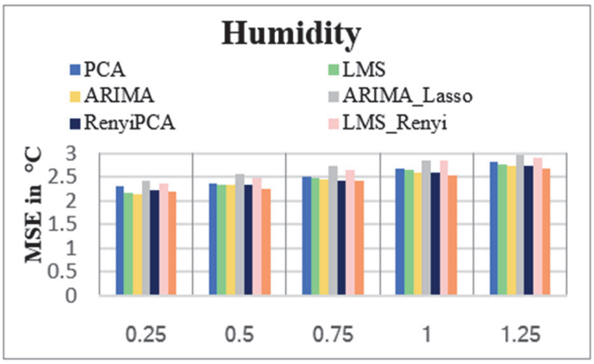

Figure 6 Comparative analysis of MSE with univariate dataset for different prediction and compression algorithms results of energy consumption

Similarly, the error in LMS prediction method can be reduced by introducing the OGMLMS filter which can be inferred from the result analysis. Comparing with the conventional LMS prediction work, the OGMLMS with Renyi PCA reduces error by 44.31% in temperature dataset when error threshold is assigned as 0.25. It is also inferred from the result table that for all datasets, the MSE increases when the prescribed error threshold is high.

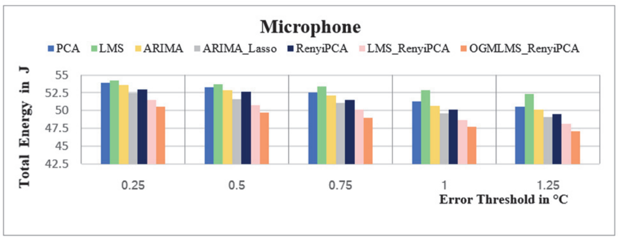

Figure 7 Comparative analysis of total energy consumption with univariate dataset for different prediction and compression algorithms

Fig. 7 depicts comparative analysis chart for the energy consumption parameter observed for different methods by setting different error threshold values for temperature dataset, humidity dataset and microphone dataset respectively.

The above graph shows that the energy consumption in conventional PCA method is high which can be reduced with the utilization of Renyi PCA. Comparing with the conventional PCA compression work, the Renyi PCA reduces consumed energy by 5.28% in temperature dataset when error threshold is assigned as 0.25. Similarly, the energy consumption in LMS prediction method can be reduced by introducing the OGMLMS filter which can be inferred from the result analysis. Comparing with the conventional LMS prediction work, the OGMLMS with Renyi PCA reduces consumed energy by 19.09% in temperature dataset when error threshold is assigned as 0.25. When the prescribed error threshold increases, the percentage of energy savings will also increase. A maximum energy saving of 14.457 % is achieved in the proposed method when compared with ARIMA_Lasso for temperature dataset.

The compression ratio and MSE of the proposed prediction-based compression algorithm on univariate data set achieved for error threshold of 1.25 is compared with standard algorithms and presented in Tab. 2.

Table 2 Comparison of MSE for univariate datasets for different prediction and compression algorithms for error threshold of 1.25

Methods	MSE		
	Temperature	Humidity	Microphone
PCA	0.785	2.828	4.778
LMS	0.723	2.772	4.765
ARIMA	0.686	2.728	4.736
ARIMA_Lasso	0.926	2.957	4.935
RenyiPCA	0.731	2.731	4.674
LMS_RenyiPCA	0.813	2.913	4.833
OGMLMS_Renyi PCA	0.572		

The compression rate when handling data with disturbances is influenced by several key factors, including the nature and severity of the disturbances, the characteristics of the data being compressed, and the chosen compression algorithm. Disturbances, such as noise or outliers, can complicate the data structure, making it more challenging to achieve high compression rates without losing significant information. The inherent redundancy within the data also plays a crucial role; data with high redundancy can be compressed more effectively than data with less redundancy. Additionally, the efficiency of the compression algorithm itself, including its ability to adapt to varying data patterns and its computational complexity, directly affects the compression rate. Finally, the desired quality of the reconstructed data, often measured by metrics such as Mean Square Error (MSE), can impose constraints on the compression process, as higher quality typically requires lower compression rates to preserve essential information.

The model balances compression and distortion by employing advanced algorithms that optimize the trade-off between reducing data size and maintaining acceptable quality levels, thereby minimizing energy and transmission costs. By utilizing techniques such as adaptive data compression and predictive modeling, the system intelligently determines the optimal compression ratio that can be achieved without exceeding predefined distortion thresholds, often measured by metrics like Mean Square Error (MSE). This approach allows the model to effectively reduce the volume of data transmitted over the network, which directly lowers energy consumption associated with data transmission and processing. Additionally, the model incorporates mechanisms to prioritize critical data, ensuring that essential information is preserved while less important data can be compressed more aggressively. By continuously monitoring and adjusting the compression parameters based on real-time data characteristics and network conditions, the model achieves an efficient balance that enhances overall system performance while conserving energy and reducing costs.

6 CONCLUSION

In this paper, OGMLMS filter-based prediction model is designed and it is combined with Renyi PCA compression in order to increase the energy conservation in WSN. The performance of this proposed joint scheme is compared with other conventional algorithms on assessment metrics like MSE, energy consumption, transmission cost and compression ratio. All the above metrics are analysed for different values of error threshold on univariate datasets. From the results it is clear that the MSE observed for univariate data sets are less in the proposed method. MSE is very much reduced with a maximum reduction of 38% in univariate temperature dataset. It is also inferred that a maximum of 14.457% of energy saving is achieved in temperature univariate dataset when compared with ARIMA_Lasso combination. A maximum of 99% compression ratio is observed in all univariate datasets. The transmission cost incurred is also found to be very less in the OGMLMS_Renyi PCA method when the prescribed error threshold is high. Since the PCA method removes the unremarkable data, the size of compressed data is more reduced. Consequently, the energy consumption and transmission costs are reduced to a better extent. In addition, the PCA preserves the data quality and thereby low MSE is obtained. Thus, it can be concluded that the proposed method is well suited for energy conservation in all kinds of datasets with different statistical behaviour.

Acknowledgments

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R733), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. Ongoing Research Funding program, (ORF-2025-838), King Saud University, Riyadh, Saudi Arabia.

7 REFERENCES

[1] Engmann, F., Katsriku, A., Abdulai, J., & Adu-Manu, K. (2020). Reducing the energy budget in WSN using time series models. *Wireless Communications and Mobile Computing*, 1-15. https://doi.org/10.1155/2020/8893064

[2] Mohsen, K. K. & Asedkhan, M. K. (2022). Structural health monitoring of oil pipelines using wireless sensor networks. *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 1-5. https://doi.org/10.1109/HORA55278.2022.9799948

[3] Bennani, M. T. & Yaden, M. F. (2023). An optimized discrete wavelet transform compression technique for image transferring over wireless multimedia sensor networks. *International Journal of Electrical & Computer Engineering*, *13*(3), 2769-2777. https://doi.org/10.11591/ijece.v13i3.pp2769-2777

[4] El-Bably, D. L. & Fouad, K. M. (2022). A hybrid approach for improving data classification based on PCA and enhanced ELM. *International Journal of Advanced Intelligence Paradigms*, *23*(3-4), 217-247. https://doi.org/10.1504/IJAIP.2022.126688

[5] Surenthar, I., Sridhar, K. P., & Roberts, M. K. (2023). Maximizing energy efficiency in wireless sensor networks for data transmission: A deep learning-based grouping model approach. *Alexandria Engineering Journal*, *83*, 53-65. https://doi.org/10.1016/j.aej.2023.10.016

[6] Oluwadara, J., Odeyinka, O. A., Michael, C., Ndinechi, O. C., & Nosiri, N. C. O. (2020). A review on conservation of energy in wireless sensor networks. *International Journal of Smart Sensor Technologies and Applications*, *1*(2). https://doi.org/10.4018/IJSSTA.20200401.oa1

[7] William, P., Badholia, A., Verma, V., Sharma, A., & Verma, A. (2022). Analysis of data aggregation and clustering protocol in wireless sensor networks using machine learning. *Evolutionary Computing and Mobile Sustainable Networks: Proceedings of ICECMSN 2021*, 925-939. https://doi.org/10.1007/978-981-16-9605-3_65

[8] Liu, Y., Li, C., Xiao, J., Li, Z., Chen, W., Qu, X., & Zhou, J. (2022). QEGWO: Energy-efficient clustering approach for industrial wireless sensor networks using quantum-related bioinspired optimization. *IEEE Internet of Things Journal*, *9*(23), 23691-23704. https://doi.org/10.1109/JIOT.2022.3189807

[9] El-Sayed, W. M., El-Bakry, H. M., & El-Sayed, S. M. (2023). Integrated data reduction model in wireless sensor networks. *Applied Computing and Informatics*, *19*(1-2), 41-63. https://doi.org/10.1016/j.aci.2019.03.003

[10] Kavitha, A., Guravaiah, K., Velusamy, L., Suseela, S., & Kumar, D. (2023). DR-NAP: Data reduction strategy using neural adaptation phenomenon in wireless sensor networks. *International Journal of Communication Systems*, *36*(8), e5467. https://doi.org/10.1002/dac.5467

[11] Jain, K. & Singh, A. (2023). Data-prediction model based on stepwise data regression method in wireless sensor networks. *Wireless Personal Communications*, *128*(3), 2085-2111. https://doi.org/10.1007/s11277-022-10034-3

[12] Shahina, K. & Pradeep Kumar, T. S. (2022). Similarity-based clustering and data aggregation with independent component analysis in wireless sensor networks. *Transactions on Emerging Telecommunications Technologies*, *33*(7), e4462. https://doi.org/10.1002/ett.4462

[13] Azarnia, G. (2023). Steady-state analysis of diffusion least-mean squares with deficient length over wireless sensor networks. *Computing*, *105*, 2443-2458. https://doi.org/10.1007/s00607-023-01187-5

[14] Song, H., Sui, S., Han, Q., Zhang, H., & Yang, Z. (2020). Autoregressive integrated moving average model-based secure data aggregation for wireless sensor networks. *International Journal of Distributed Sensor Networks*, *16*(3). https://doi.org/10.1177/1550147720912958

[15] Al Shahrani, A. M., Alomar, M. A., Alqahtani, K. N., Basingab, M. S., Sharma, B., & Rizwan, A. (2022). Machine learning-enabled smart industrial automation systems using Internet of Things. *Sensors*, *23*(1), 324. https://doi.org/10.3390/s23010324

[16] Nasser, A. & AL-Khazraji, H. (2021). A hybrid of convolutional neural network and long short-term memory network approach to predictive maintenance. *International Journal of Electrical and Computer Engineering*, *12*(1), 721-730. https://doi.org/10.11591/ijece.v12i1.pp721-730

[17] Al-Fayoumi, M. & Al-Haija, Q. A. (2023). Capturing low-rate DDoS attack based on MQTT protocol in software defined-IoT environment. *Array*, *19*, 100316. https://doi.org/10.1016/j.array.2023.100316

[18] Haoxiang, D. W. & Smys, D. S. (2020). Soft computing strategies for optimized route selection in wireless sensor networks. *Journal of Soft Computing Paradigm*, *2*(1), 1-12. https://doi.org/10.36548/jscp.2020.1.001
[19] Diwakaran, S., Perumal, B., & Vimala Devi, K. (2019). A cluster prediction model-based data collection for energy-efficient wireless sensor networks. *The Journal of Supercomputing*, *75*, 3302-3316. https://doi.org/10.1007/s11227-018-2437-z
[20] Ketshabetswe, K. L., Zungeru, A. M., Mtengi, B., Lebekwe, C. K., & Prabaharan, S. R. S. (2021). Data compression algorithms for wireless sensor networks: A review and comparison. *IEEE Access*, *9*, 136872-136891. https://doi.org/10.1109/ACCESS.2021.3116311
[21] Sah, D. K., Hazra, A., Kumar, R., & Amgoth, T. (2022). Harvested energy prediction technique for solar-powered wireless sensor networks. *IEEE Sensors Journal*, *23*(8), 8932-8940. https://doi.org/10.1109/JSEN.2022.3208730
[22] Nurgaliyev, M., Saymbetov, A., Yashchyshyn, Y., Kuttybay, N., & Tukymbekov, D. (2020). Prediction of energy consumption for LoRa-based wireless sensor networks. *Wireless Networks*, *26*, 3507-3520. https://doi.org/10.1007/s11276-020-02276-5
[23] Haque, M. E. & Baroudi, U. (2020). Dynamic energy-efficient routing protocol in wireless sensor networks. *Wireless Networks*, *26*, 3715-3733. https://doi.org/10.1007/s11276-020-02290-7

Contact information:

VINI ANTONY GRACE N.
(Corresponding Author)
Department of Electronics and Communication Engineering,
R.M.D. Engineering College, Kavaraipettai, Chennai, Tamilnadu, India
E-mail: viniantonygrace@protonmail.com

NADA ALZABEN
Department of Computer Sciences,
College of Computer and Information Sciences,
Princess Nourah bint Abdulrahman University,
P. O. Box 84428, Riyadh 11671, Saudi Arabia
E-mail: Ni.zaben@pnu.edu.sa

FAIZ ABDULLAH ALOTAIBI
Department of Information Science,
College of Humanities and Social Sciences,
King Saud University,
P. O. Box 28095, Riyadh 11437, Saudi Arabia
E-mail: fa.otaibi@ksu.edu.sa

AHMAD A. ALZAHRANI
Department of Computer Science and Artificial Intelligence,
College of Computing, Umm Al-Qura University, Saudi Arabia
E-mail: ah.zhrani@uqu.edu.sa