



OPEN Energy consumption minimisation at edge node using C_cBPS approach in predicting sensor parameters in WSNs

Vipin Maurya¹, Sumit Kumar², Sonali Raj¹ & Ruchir Gupta¹✉

Owing to limited storage and battery power, wireless sensor nodes often face challenges in maintaining long-term energy sustainability. To address this, only a subset of sensors remains active to monitor different sensor parameters while others get predicted to minimize sensor node energy consumption. In prediction, not all active parameters are equally important, as low-correlated parameters increase computational complexity and decrease accuracy. Researchers use highly correlated active parameters, though existing solutions often use polynomial time and don't ensure optimal parameter set. This paper proposes a cross-correlation-based parameter selection (C_cBPS) approach, ensuring the selected parameter set is stable and Pareto-optimal. Simulations are performed on nine publicly available datasets of environmental data collected from different places and at different sampling intervals to validate the effectiveness of the C_cBPS approach. It has been observed that C_cBPS approach selects a subset of active parameters faster than existing approaches and reduces energy consumption at the edge node ranges from 6.5% - 34.2% in the prediction of sleep sensor parameters on various datasets.

Keywords Cross-correlation, Edge computing, Feature selection, Game theory, Gaussian process regression (GPR)

The rapidly growing Internet of Things (IoT) technology aims to connect millions of devices that are becoming an integral part of daily life. IoT sensor nodes, equipped with multiple sensors to monitor various parameters, are commonly deployed in diverse applications and are expanding rapidly. These sensor nodes are wirelessly connected to a central entity, forming wireless sensor networks (WSNs). WSNs play a crucial role in industrial monitoring, smart city management, agriculture, hospital monitoring, and border surveillance. However, the sensor nodes in WSNs typically have limited computing power and energy resources, which makes it challenging to process and analyze the collected data locally. As a result, this task is often offloaded to the cloud¹. Unfortunately, relying solely on cloud services can introduce unpredictable delays due to the involvement of the core network, which also increases energy consumption and bandwidth usage². Edge computing addresses these limitations by providing computing resources closer to the WSN, thereby reducing latency, energy consumption, and bandwidth requirements.

Different sensors of each sensor node of the WSN measure their respective parameters at predefined intervals known as measurement cycles {Measurement Cycle: A measurement cycle is a recurring process in a Wireless Sensor Network (WSN) that involves data sensing and transmission for monitoring the physical environment}. These parameters exhibit correlations among them^{3,4}. Consequently, some parameters are directly measured, while others are predicted at the edge node using these correlations. An edge intelligence-based sensing strategy is employed to reduce the energy consumption of sensor nodes in WSNs^{3,5}. In this strategy, the edge node selects an optimal subset of sensors to remain active for sensing environmental parameters in the measurement cycle while the rest of the parameters are in sleep mode. A Gaussian process regression (GPR)-based machine learning prediction model predicts sleep sensor parameters at edge node using correlated active sensor parameter⁵⁻⁷. Using this strategy, the researcher minimized the energy consumption of the sensor node but did not consider the computational complexity and energy consumption of edge nodes during the prediction of sleep sensor parameters. The edge node's computational complexity and energy consumption are also important factors.

¹Department of Computer Science and Engineering, Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh, India. ²Department of Computer Science and Engineering, Dr. B R Ambedkar National Institute of Technology, Jalandhar, Punjab, India. ✉email: rgupta.cse@iitbhu.ac.in

Therefore, for the efficient performance of the edge node, it is evident that computational complexity and energy consumption should be minimized while predicting sleep sensor parameters.

In the prediction of sleep sensor parameters, it is observed that not all active parameters contribute equally to the prediction process, as they do not exhibit the same level of correlation. Using active parameters with low correlation to the sleep sensor parameters adds unnecessary complexity to the prediction process and degrades the accuracy of predicted sleep sensor parameters^{8,9}. Bhuyan et al.¹⁰ used the correlation coefficient among parameters and the gradient descent technique to select the parameters from the active parameter set. In their approach, first, they assign the weight value $w \in [0, 1]$ to each active parameter and then use the gradient descent technique to select the parameters. Tripathi et al.¹¹ used the chi-square parameter selection technique to select the parameter from the active parameter set. They considered positive and negative chi-square values for parameters and then used the probabilistic chi-square value to select the parameter from the active parameter set.

In the above parameter selection approach, Bhuyan et al.¹⁰ used the gradient descent technique to optimize the parameter selection process, which takes many iterations to optimize the weights for each parameter. Tripathi et al.¹¹ select the static number of parameters that limit their ability to ensure the best parameters for each sleep sensor parameter. All these approaches take polynomial time to select parameters from the active parameter set, and these approaches do not ensure whether the selected parameter set is an optimal parameter set {Optimal Parameter Set: A parameter set is optimal among all possible parameter sets if it has a higher correlation with the predicted parameter and the prediction model consumes less energy in prediction using this set.} of active sensor parameters.

In this paper, we propose a cross-correlation-based parameter selection (C_cBPS) approach that selects optimal parameters from active parameter set to predict each sleep parameter. The C_cBPS approach selects those active parameters that exhibit either high correlation or correlation greater than or equal to the average correlation of all active parameters with the sleep sensor parameter. By selecting the optimal parameters from the active sensor parameter set, the proposed approach reduces the computational complexity and energy consumption of the edge nodes in the prediction of sleep sensor parameters. Furthermore, by eliminating low-correlated parameters from the active parameter set, our method improves the accuracy of predicted sleep sensor parameters. Simulations are conducted to validate the effectiveness of the C_cBPS approach using nine environmental datasets. These datasets were collected from different geographical locations and time intervals and are publicly available. These datasets consist of nine sensors designed to monitor specific environmental parameters, including PM2.5, PM10, NO, CO, NO₂, NH₃, SO₂, Ozone, and Benzene.

The main contributions of this paper are as follows:

1. An algorithm based on C_cBPS is proposed to select the optimal parameter set $\mathbb{K}$ from the active sensor parameter set to predict each sleep parameter at the edge node.
2. The C_cBPS approach is theoretically analyzed and observed that the selected parameter set is both stable and Pareto optimal.
3. The time complexity of the proposed parameter selection algorithm is calculated and compared with the existing state-of-the-art parameter selection approaches.
4. Simulations are performed on various datasets to validate the effectiveness of the proposed approach. Simulation results show that the proposed approach selects the smallest parameter set from active sensor parameters and reduces energy consumption at the edge node ranges from 6.5% - 34.2% in the prediction of sleep sensor parameters on various datasets.

The paper is structured as follows: The literature review is presented in Section 2, followed by a description of the system model in Section 3. Section 4 discusses the cross-correlation-based optimal parameter selection approach, followed by Section 5, which presents the experimental setup for the simulation. Section 6 combines the results and discussion, while Section 7 provides the conclusion.

Literature Review

Parameter selection is an important approach in prediction because it reduces the dimensionality of the parameter, simplifying the predictive model while retaining the most relevant information. By eliminating less significant parameters, the parameter selection approach minimizes the prediction's computational load and enhances the accuracy of predicted parameters. The authors utilized parameter selection strategies to improve the model's efficiency and minimize energy consumption and computational costs. Bhuyan et al.¹⁰ used the correlation coefficient among data and the gradient descent technique to select the parameter from the parameter set. In their approach, first, they assign the weight value $w \in [0, 1]$ to each active parameter and then use the gradient descent technique to select the parameters. Tripathi et al.¹¹ used the chi-square parameter selection technique to select the parameters from the dataset. They considered positive and negative chi-square values for parameters and then used the probabilistic chi-square value to select the parameter from the dataset. Tu et al.¹² proposed an enhanced approach for Cross-Validated Recursive Feature Elimination (RFECV) by randomly sampling different data, building multiple models, and comparing their scores to improve the robustness of the optimal parameter subset. Belhaoui et al.¹³ proposed the bird's eye view parameter selection technique, which incorporated elements of evolutionary algorithms, genetic algorithms, dynamic Markov chain, and reinforcement learning to improve classification performance and reduced the number of parameters compared to conventional methods. Alhussan et al.¹⁴ used the binary waterwheel plant optimization technique for the selection of parameters in classification problems. Zhan et al.¹⁵ proposed a PD-based parameter selection technique that uses Possibility Degrees (PDs) to indicate parameter significance and mitigate data uncertainty. They used Gini entropy minimization to prioritize informative parameters, maximum correlation parameter

selection to assure high relevance and minimal redundancy, and multisource information fusion to improve parameter informativeness and confidence. Fan et al.¹⁶ introduced an LCIFS technique that improves multi-label learning by capturing label correlations and managing parameter redundancy using information entropy and Pearson correlation analysis. It uses a manifold-based regression model and adaptive spectral graphs to discover accurate structural label correlations. In the above studies, either authors^{12–14} are utilizing machine learning or optimization techniques to get the optimal parameters or authors^{10,11,15,16} do not ensure that the selected parameter set in an optimal parameter set.

To the best of our knowledge, the existing studies either did not ensure that the selected parameter set is optimal parameter set or may take a long computational time to select the optimal parameter set from the active parameters. Therefore, we propose a C_cBPS approach that relies on cross-correlations among sensor parameters to select the optimal parameter set from the active sensor parameter set to predict sleep sensor parameters faster than the existing parameter selection approaches. As a result, the proposed approach reduces computational complexity and energy consumption at the edge nodes in the prediction of sleep sensor parameters while improving the accuracy of predicted sleep sensor parameters.

System model

As shown in Figure 1, consider a sensor node with S sensors where every sensor monitors one environmental parameter such that the parameter set $\mathbb{S} = \{i; 1 \leq i \leq S\}$ contains all the parameters measured by the sensor node, where every i is an integer number and it represents a parameter of the sensor node. Assume a set $\mathbb{Q} = \{q_1, q_2, \dots, q_S\}$ such that $q_i = 1$ for every $i \in S$ if i^{th} parameter is active, otherwise $q_i = 0$. Set $\mathbb{G} = \{i \mid \forall q_i = 1, q_i \in \mathbb{Q}\}$ contains the active sensor parameters, and set $\mathbb{H} = \mathbb{Q} - \mathbb{G} = \{i \mid \forall q_i = 0, q_i \in \mathbb{Q}\}$ contains the sleep sensor parameters. The cardinalities of $\mathbb{G}$ and $\mathbb{H}$ are denoted by G and H , respectively. If set $\mathbb{Z} = \{\mathbb{G}_1, \mathbb{G}_2, \dots, \mathbb{G}_X\}$, where $X = 2^S - 1$, contains all subsets of active sensor parameter set excluding null set, then set $\mathbb{U} = \{(\mathbb{G}_k, \mathbb{H}_k) : 1 \leq k \leq X\}$ contains all pairs of active-sleep sensor parameters, where $\mathbb{U}_i$ represents the i^{th} pair of active-sleep sensor parameters and it is defined as $\mathbb{U}_i = (\mathbb{G}_i, \mathbb{H}_i)$. The matrix $\mathbb{CC} \in \mathbb{R}^{S \times S}$ represents the cross-correlation coefficient between all parameters of the parameter set $\mathbb{S}$. The matrix is defined as follows:

$$\mathbb{CC} = \begin{bmatrix} C_{S_1 S_1} & C_{S_1 S_2} & \dots & C_{S_1 S_S} \\ C_{S_2 S_1} & C_{S_2 S_2} & \dots & C_{S_2 S_S} \\ \vdots & \ddots & \ddots & \vdots \\ C_{S_{S-1} S_1} & C_{S_{S-1} S_2} & \dots & C_{S_{S-1} S_S} \\ C_{S_S S_1} & C_{S_S S_2} & \dots & C_{S_S S_S} \end{bmatrix}$$

where $C_{S_i S_j} = \mathbb{CC}[S_i][S_j]$ represents the value of the cross-correlation coefficient between any pairs $(S_i, S_j) \in \mathbb{S}$.

Problem Definition

In environmental data collection, some sensors become active to sense environmental parameters, while the rest parameters are in sleep mode in the measurement cycle. The sleep parameters in that measurement cycle are predicted at the edge node by utilizing the cross-correlation among parameters⁵. However, this prediction process also includes those active parameters that exhibit weak correlations with the sleep parameters, directly impacting the edge node's computational complexity and energy consumption. The sleep sensor parameters are predicted by using the GPR-based prediction model^{3,5}. In the GPR-based prediction model, the covariance matrix computation notably influences the model's computation and energy consumption. Therefore, reducing the energy consumption of the edge nodes is directly proportional to reducing the energy consumption of

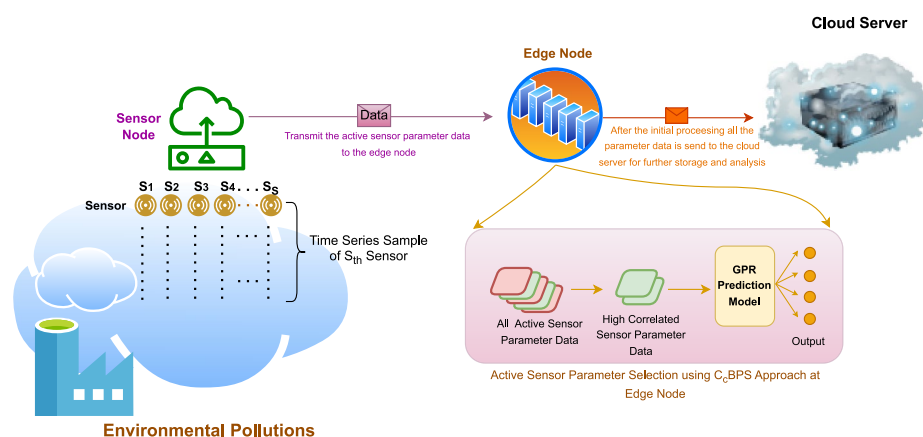


Figure 1. Environmental Data Collection and Transmission in Wireless Sensor Network via Edge Node. {Created using draw.io, Version: 24.7.17, <https://www.drawio.com/>}.

covariance matrix computation in the GPR prediction model. The covariance matrix computation is defined as follows:

$$k(x_i, x_j) = \exp\left(\frac{-1}{2l'^2} \|x_i - x_j\|^2\right) + \beta^2 \times \rho_{ij} \quad (1)$$

where β^2 represents the variance of noise vector and ρ_{ij} is the kronecker delta function. Here $\|x_i - x_j\|$ quantifies the pairwise euclidean distance in the original G - dimensional space, computed as:

$$\|x_i - x_j\| = \sqrt{\sum_{g=1}^G (x_{i,g} - x_{j,g})^2}. \quad (2)$$

where G represents the number of active parameters.

From Equation (1), to reduce the energy consumption in covariance matrix computation, it is necessary to minimize the number of active parameters. This paper aims to select the optimal parameter set $\mathbb{K}$ from the active parameter set $\mathbb{G}$ that will be used in the prediction of sleep sensor parameter h . The optimal set $\mathbb{K}$ from the active parameter set for predicting the sleep parameter h ($h \in \mathbb{H}$) is defined as follows:

$$\mathbb{K}_h = \{k \mid \forall k \in K[h] \text{ and } K[h][k] = 1\} \quad (3)$$

where K represents the matrix that contains all the optimal parameters used in the prediction of each sleep sensor parameter and $K[h][k]$ represents the value of matrix K at the position h and k , where h and k represents the sleep and active parameters, respectively. By finding this optimal parameter set $\mathbb{K}_h$ from the active parameters, the proposed approach effectively minimizes the computational complexity and energy consumption of the edge node in the prediction of sleep sensor parameter h .

GPR-Based Cross-Correlated Parameter Prediction Model

The sleep sensor parameters are predicted using Gaussian Process Regression (GPR)-based prediction models at the edge node. By harnessing the predictive capabilities of GPR, the edge node adeptly fills in data gaps, thereby ensuring a comprehensive dataset for subsequent analysis¹⁷. The GPR model is comprised of a total of U ($U = |\mathbb{U}|$) sub-models, corresponding to U active-sleep pairs of sensor parameters. Each sub-model is designed to include H_i where $(1 \leq i \leq U)$ number of the regressor, aimed at predicting each parameter of $\mathbb{H}_i$. When the subset $\mathbb{G}_j$ is active during the measurement cycle, the selection of the j^{th} sub-model becomes pertinent, as it is tailored to estimate each parameter of $\mathbb{H}_j$. Notably, in the GPR model, the prediction of each parameter h within $\mathbb{H}_j$ is done by strategically choosing an optimal parameter set $\mathbb{K}_h$ from the active sensor parameters. If p^{th} parameter collected N_p number of samples, then $\mathcal{Z}_p = \mathbb{R}^{N_p \times 1}$ denotes the temporal measurement values of p^{th} parameter and can be formulated as follows:

$$\mathcal{Z}_p = \mathcal{Y}_p + \eta_p \quad (4)$$

where $\mathcal{Y}_p$ is the actual signal vector and η_p represents the measurement noise vector associated with the p^{th} sensor parameter. The noise vector components are independent and identically distributed, where $\eta = N(0, \rho)$ for all sensor parameters.

In GPR prediction model the input matrix of the h^{th} regressor of the i^{th} submodel is $\mathbf{I}_{\mathbb{K}_h}^i \in \mathbb{R}^{m \times K_h}$ and the output vector is $\mathbf{I}_{\mathbb{H}_h}^i \in \mathbb{R}^{m \times 1}$, where m represents the number of collected samples by the sensor parameters. Each row of the matrix $\mathbf{I}_{\mathbb{K}_h}^i$ contains one parameter vector. The j^{th} row of the matrix $\mathbf{I}_{\mathbb{K}_h}^i$ is represented as $\mathbf{I}_{\mathbb{K}_h}^{i,j} = \{I_{\mathbb{K}_h}(j, 1), I_{\mathbb{K}_h}(j, 2), \dots, I_{\mathbb{K}_h}(j, K)\}$. The Equation (4) for the active parameter $\mathbf{I}_{\mathbb{K}_h}^i[j]$ and sleep sensor parameter $\mathbf{I}_{\mathbb{H}_h}^i(j)$ can be written as, $\mathcal{Z}_{\mathbf{I}_{\mathbb{K}_h}^i}^i(j) = \mathcal{Y}_{\mathbf{I}_{\mathbb{K}_h}^i}^i(j) + \eta_{\mathbf{I}_{\mathbb{K}_h}^i}^i(j)$ and $\mathcal{Z}_{\mathbf{I}_{\mathbb{H}_h}^i}^i(j) = \mathcal{Y}_{\mathbf{I}_{\mathbb{H}_h}^i}^i(j) + \eta_{\mathbf{I}_{\mathbb{H}_h}^i}^i(j)$, respectively. For the test vector $\mathbf{I}_{\mathbb{K}_h}^i(*)$, GPR finds the underlying function $\mathcal{F}_{h,*}^i$ for the h^{th} regressor as:

$$\mathcal{Y}_{\mathbb{H}_h}^i(*) = \mathcal{F}_h(\mathbf{I}_{\mathbb{K}_h}^i(*)) = \mathcal{F}_{h,*}^i \in \mathbb{R} \quad (5)$$

where $\mathcal{F}_h^i \sim N(0, K_{m \times m})$. $K_{m \times m}$ is the covariance matrix whose elements are derived from the Radial Basis Function (RBF) kernel. $K_{m \times m} = [k(x_i, x_j)]_{m \times m}$, where $k(x_i, x_j)$ represents the i^{th} and j^{th} sampling instance of the input matrix $\mathbf{I}_{\mathbb{K}_h}^i$. The RBF kernel for the h^{th} regressor is derived as follows:

$$k(x_i, x_j) = \exp\left(\frac{-1}{2l'^2} \sum_{k=1}^{K_h} (x_{i,k} - x_{j,k})^2\right) + \beta^2 \times \rho_{ij} \quad (6)$$

Now, the GPR prediction model is given by $\mathcal{F}_{h,*}^i \mid \mathbf{I}_{\mathbb{K}_h}^i, \mathbf{I}_{\mathbb{H}_h}^i, \mathbf{I}_{\mathbb{K}_h}^i(*) \sim N(\overline{\mathcal{F}_{h,*}^i}, \text{Cov}(\mathcal{F}_{h,*}^i))$, where $\overline{\mathcal{F}_{h,*}^i}$ and $\text{Cov}(\mathcal{F}_{h,*}^i)$ represents the mean function and covariance function and are derived from (A.1) and (A.2) of Appendix. The mean ($\overline{\mathcal{F}_{h,*}^i}$) and covariance ($\text{Cov}(\mathcal{F}_{h,*}^i)$) are expressed^{3,5,18,19} as:

$$\overline{\mathcal{F}_{h,*}^i} = K(\mathbf{I}_{\mathbb{K}_h}^i, \mathbf{I}_{\mathbb{K}_h}^i(*)) [K(\mathbf{I}_{\mathbb{K}_h}^i, \mathbf{I}_{\mathbb{K}_h}^i) + \beta^2 I]^{-1} (\mathbf{I}_{\mathbb{H}_h}^i) \quad (7)$$

$$Cov(\mathcal{F}_{h,*}^i) = k(I_{\mathbb{K}_h}^i(*), I_{\mathbb{K}_h}^i(*)) + K(I_{\mathbb{K}_h}^i(*), I_{\mathbb{K}_h}^i(*)) [K(I_{\mathbb{K}_h}^i, I_{\mathbb{K}_h}^i) + \beta^2 I]^{-1} K(I_{\mathbb{K}_h}^i, I_{\mathbb{K}_h}^i(*)) \quad (8)$$

This GRP prediction model will be used in Section 4 to predict the sleep sensor parameter values at the edge node using the cross-correlated optimal parameters selected from active sensor parameters.

PROPOSED APPROACH

This section presents a C_cBPS approach for selecting the optimal parameter set from active sensor parameters to predict sleep sensor parameter at the edge node. The edge node selects the optimal parameter set $\mathbb{K} \subseteq \mathbb{G}$, which has a strong correlation with sleep parameter $h (h \in \mathbb{H})$. The edge node predicts the sleep parameter h using the set $\mathbb{K}$ and employing the GPR prediction model.

Assume two sensing parameters a and b . If $|\text{CC}[a][b]| \geq 0.5^3$, the parameters a and b are considered as correlated. The cross-correlation factor $\delta(\mathbb{G}, h)$ determines the degree of cross-correlation of the active parameter set $\mathbb{G}$ with sleep parameter $h (h \in \mathbb{H})$. The cross-correlation factor $\delta(\mathbb{G}, h)$ is defined as follows:

$$\delta(\mathbb{G}, h) = \frac{1}{G} \sum_{i=1}^G |\text{CC}[i][h]|; \forall i \in \mathbb{G}. \quad (9)$$

The performance of C_cBPS approach is evaluated using the reconstruction error of the sensing parameters at the edge node. The reconstruction error for parameter s in the measurement cycle x is Se_s^x , then the reconstruction error of all parameters is as follows:

$$SE^x = \frac{1}{S \times Z} \sum_{s=1}^S \sum_{z=1}^Z Se_{s,z}^x \quad \forall s \in S \quad (10)$$

where Z represents the number of the measurement cycle.

Optimal subset of active parameters set

This section presents the proposed approach for selecting the optimal parameter set from active sensor parameters for each sleep sensor parameter as given in Algorithm 1. This algorithm iteratively selects an optimal parameter set for every sleep sensor parameter $i (i \in \mathbb{H})$ (Lines 2-27). In an iteration, first, we select two sets of active sensor parameters $\mathbb{X}$ and $\mathbb{W}$ that have a cross-correlation coefficient greater than or equal to 0.9 and 0.5 with sleep parameter i , respectively (Lines 5-12), as shown in Equations (11) and (12).

$$\mathbb{X} = \{x \mid \forall x \in \mathbb{G} \wedge |\text{CC}[i][x]| \geq 0.9\} \quad (11)$$

and

$$\mathbb{W} = \{w \mid \forall w \in \mathbb{G} \wedge |\text{CC}[i][w]| \geq 0.5\} \quad (12)$$

The reason for selecting the parameter set $\mathbb{W}$ is that only the active parameters with a correlation coefficient greater than or equal to 0.5 with i are considered in the prediction of the parameter i , and the parameters that exhibit a correlation less than 0.5 are discarded³ and parameter set $\mathbb{X}$ contains the active parameters that exhibit high correlation with the sleep sensor parameter i^{20} . In Equations (11) and (12), to get the correlation coefficient between parameters, we use the correlation matrix $\text{CC} \in \mathbb{R}^{S \times S}$ which is already available for the given data samples of the parameter set.

Input: $\mathbf{CC} \in \mathbb{R}^{S \times S}$, $\mathbb{G}$, $\mathbb{H}$
Output: The matrix $\mathbf{K}$

```

1 Initialization Matrix  $\mathbf{K} = \mathbf{0}_{S \times S}$ 
2 for each element  $i$  in  $\mathbb{H}$  do
3    $\mathbb{W} = \{\}$  ;
4    $\mathbb{X} = \{\}$  ;
5   for each element  $j$  in  $\mathbb{G}$  do
6     if  $|\mathbf{CC}[i][j]| \geq 0.9$  then
7        $\mathbb{X} = \mathbb{X} \cup j$  ;
8     end
9     if  $|\mathbf{CC}[i][j]| \geq 0.5$  then
10       $\mathbb{W} = \mathbb{W} \cup j$  ;
11    end
12  end
13  Compute  $\delta(\mathbb{W}, i)$  using set  $\mathbb{W}$ , sleep parameter  $i$ , and Equation (13)
14  if  $|\mathbb{X}| \neq 0$  then
15    for each element  $x$  in  $\mathbb{X}$  do
16       $\mathbf{K}[i][x] = 1$  ;
17    end
18  end
19  else
20    for each element  $w$  in  $\mathbb{W}$  do
21      if  $|\mathbf{CC}[i][w]| \geq \delta^x(\mathbb{W}, i)$  then
22         $\mathbf{K}[i][w] = 1$  ;
23      end
24    end
25  end
26 end
27 return ( $\mathbf{K}$ )

```

Algorithm 1. Selection of active parameter for prediction of sleep sensor parameter

The proposed approach first checks whether there exist parameters that are highly correlated with sleep sensor parameter i ; if yes, then it uses all the highly correlated active parameters to predict the sleep parameter i (Lines 14–18). Otherwise, it selects parameters from set $\mathbb{W}$, ensuring that the selected subset is small in size and captures the maximum information about the sleep sensor parameter (Lines 19–25). To select such parameters, we find the average of selected parameters using Equation (13) (Line 13).

$$\delta(\mathbb{W}, i) = \frac{1}{|\mathbb{W}|} \sum_{w=1}^{|\mathbb{W}|} |\mathbf{CC}[i][w]|; \forall w \in \mathbb{W} \quad (13)$$

where $i \in \mathbb{H}$. This average will be a new threshold, and we will identify those parameters whose correlation with sleep sensor parameter i is greater than or equal to this threshold (Lines 20–25).

For other sleep sensor parameters, this algorithm will select the optimal active parameter set in a similar way. In this way, the algorithm returns matrix $\mathbf{K}$, which contains optimal active parameters corresponding to each sleep sensor parameter.

Theoretical Analysis of Proposed Approach

In this section, the proposed approach is theoretically analyzed, and it is observed that the selected parameter set is both stable and Pareto optimal. The optimal parameter set $\mathbb{K}$ to predict the sleep sensor parameter h ($h \in \mathbb{H}$) is defined as follows:

$$\mathbb{K} = \{k \mid \forall k \in \mathbb{K}[h] \text{ and } \mathbf{K}[h][k] = 1\} \quad (14)$$

The C_cBPS approach is analyzed using the Shapley value from the cooperative game. Shapley value is the contribution of each player based on their marginal contributions to all possible coalitions. Therefore, the allocation of Shapley value in the coalition is stable because the contribution value of each player in the coalition is greater than or equal to their worth. The Shapley value of player i is defined as follows:

$$\phi_i = \sum_{k \subseteq \mathbb{K} \setminus \{i\}} \frac{|k|!(n - |k| - 1)!}{n!} (v(k \cup \{i\}) - v(k)) \quad (15)$$

where $\mathbb{K}$ represents the set of all players and n is the total number of players in the cooperative game. $v(k)$ represents the value/worth of the coalition of the set of players in the cooperative game.

Theorem 1 The allocation of sensors in sensor set $\mathbb{K} (\mathbb{K} \subseteq \mathbb{G})$ using C_cBPS approach is a stable allocation.

Proof The allocation of sensors in the sensor set is considered a stable allocation if each sensor k of sensor set $\mathbb{K}$ has a Shapley value greater than or equal to their individual worth.

All the sensors of the sensor set $\mathbb{K}$ are considered the players of the cooperative game. The action profile of each player is the decision that the player takes to participate (play) in the coalition or not to participate (pass). The action profile of each player is defined as:

$$\text{Action} = \{\text{Play}, \text{Pass}\}$$

The worth function is the value of the coalition of the game. The worth function of the coalition depends on the factors that determine how strong the coalition can get the degree of cross-correlation while expanding the energy consumption in the prediction. This also includes the number of active parameters that strongly correlate with the sleep parameter. The worth function $v(k)$ of any correlation k in the game is defined as follows:

$$v(k) = \rho(k, h) \times \Psi(k)^2 - e_h(k) \quad (16)$$

where $k \subseteq \mathbb{K}$. Terms $\rho(k, h) = \sum_{a \in k} CC[h][a]$ and $\Psi(k) = \|\{k \mid k \in \mathbb{K}\}\|$ represent the sum of the cross-correlation coefficient of all parameters and the total number of parameters in the coalition k , respectively. The term $e_h(k)$ represents the energy consumption in the prediction of sleep parameter h using k active parameter set.

The Shapley value of player i is determined as ;

$$\phi_i = \sum_{k \subseteq \mathbb{K} \setminus \{i\}} \frac{|k|!(n - |k| - 1)!}{n!} (v(k \cup \{i\}) - v(k)) \quad (17)$$

$$\begin{aligned} \phi_i = & \frac{|k_1|!(n - |k_1| - 1)!}{n!} (v(k_1 \cup \{i\}) - v(k_1)) + \frac{|k_2|!(n - |k_2| - 1)!}{n!} (v(k_2 \cup \{i\}) - v(k_2)) \\ & + \dots + \frac{|k_m|!(n - |k_m| - 1)!}{n!} (v(k_m \cup \{i\}) - v(k_m)). \end{aligned} \quad (18)$$

$$\begin{aligned} \phi_i = & \alpha_1 \left[\rho(k_1 \cup \{i\}, h) \times \Psi(k_1 \cup \{i\})^2 - e_h(k_1 \cup \{i\}) - (\rho(k_1, h) \times \Psi(k_1)^2 - e_h(k_1)) \right] \\ & + \alpha_2 \left[\rho(k_2 \cup \{i\}, h) \times \Psi(k_2 \cup \{i\})^2 - e_h(k_2 \cup \{i\}) - (\rho(k_2, h) \times \Psi(k_2)^2 - e_h(k_2)) \right] \\ & + \dots \dots \\ & + \alpha_m \left[\rho(k_m \cup \{i\}, h) \times \Psi(k_m \cup \{i\})^2 - e_h(k_m \cup \{i\}) - (\rho(k_m, h) \times \Psi(k_m)^2 - e_h(k_m)) \right] \end{aligned} \quad (19)$$

where $k_m \subseteq \mathbb{K} \setminus \{i\}$. The weight factor α_i is defined as:

$$\alpha_i = \frac{|k_i|!(n - |k_i| - 1)!}{n!} \quad (20)$$

from Equation (19);

Consider the m_{th} term as an empty set, then the Shapley value of player i is defined as:

$$\begin{aligned} \phi_i = & \alpha_1 \left[\rho(k_1 \cup \{i\}, h) \times \Psi(k_1 \cup \{i\})^2 - e_h(k_1 \cup \{i\}) - (\rho(k_1, h) \times \Psi(k_1)^2 - e_h(k_1)) \right] \\ & + \dots \dots + \frac{1}{n} (\rho(\{i\}, h) \times \Psi(\{i\})^2 - e_h(\{i\})) \end{aligned} \quad (21)$$

In Equation (16), since $v(k)$ is a convex function, therefore from Equations (16) and (21), $\forall k \subseteq \mathbb{K} \setminus \{i\}$;

$$(\rho(k \cup \{i\}, h) \times \Psi(k \cup \{i\})^2 - e_h(k \cup \{i\})) \geq (\rho(k, h) \times \Psi(k)^2 - e_h(k)) \quad (22)$$

and

$$v\{i\} = \rho(\{i\}, h) \times \Psi(\{i\})^2 - e_h(\{i\}) \quad (23)$$

therefore, from Equations (22) and (23);

$$\phi_i \geq v\{i\}; \forall i \in \mathbb{K}, (\mathbb{K} \subseteq \mathbb{G}). \quad (24)$$

Equation (24) shows that the active parameter set $\mathbb{K}(\mathbb{K} \subseteq \mathbb{G})$ is stable. Therefore, no sensor from the active parameter set $\mathbb{K}(\mathbb{K} \subseteq \mathbb{G})$ of the proposed C_cBPS approach leaves the coalition. $\square$

Theorem 2 The active parameter set $\mathbb{K}(\mathbb{K} \subseteq \mathbb{G})$ is a Pareto optimal set.

Proof An active parameter set $\mathbb{K}(\mathbb{K} \subseteq \mathbb{G})$ is Pareto optimal if it is impossible to add or remove any sensor parameter from $\mathbb{K}(\mathbb{K} \subseteq \mathbb{G})$ without either increasing the energy consumption in prediction or reducing the average cross correlation between the active parameter set and sleep sensor parameter. It also means that no modification to the set can improve one criterion without worsening the other.

CASE – 1 : A set $\mathbb{K}(\mathbb{K} \subseteq \mathbb{G})$ is Pareto optimal if there is no other set $\mathbb{K}'(\mathbb{K}' \subseteq \mathbb{G})$ ($\|\mathbb{K}'\| > \|\mathbb{K}\|$), where the average cross-correlation of the selected active parameter set is higher and the energy consumption is equal to or lower as compared to the set $\mathbb{K}(\mathbb{K} \subseteq \mathbb{G})$.

Suppose that the active parameter set $\mathbb{K}$ is not Pareto optimal. Then there exists another set $\mathbb{K}'$ with;

$$\delta(\mathbb{K}', h) > \delta(\mathbb{K}, h); h \in \mathbb{H} \quad (25)$$

and

$$e_h(\mathbb{K}') \leq e_h(\mathbb{K}) \quad (26)$$

since prediction energy is a monotonic increasing function. Therefore, increasing the number of active parameters increases the prediction energy consumption. But, from our assumption,

$$\|\mathbb{K}'\| > \|\mathbb{K}\| \quad (27)$$

from Equation (27):

$$e_h(\mathbb{K}') > e_h(\mathbb{K}) \quad (28)$$

Equation (28) contradicts our assumptions of Equation (26). Therefore, this concludes that there is no other set $\mathbb{K}'$ that has lower energy consumption compared to $\mathbb{K}$.

CASE – 2 : A set $\mathbb{K}$ is Pareto optimal if there is no other set $\mathbb{K}'$ ($\|\mathbb{K}'\| \leq \|\mathbb{K}\|$) where the degree of cross-correlation of the selected sensor is higher and the energy consumption is equal to or lower.

Assume that $\mathbb{K}$ is not Pareto optimal. Suppose that another set $\mathbb{K}'$ exists that achieves the same or lower energy consumption and a higher degree of cross-correlation as follows:

$$e_h(\mathbb{K}') \leq e_h(\mathbb{K}) \quad (29)$$

and

$$\delta(\mathbb{K}', h) > \delta(\mathbb{K}, h); h \in \mathbb{H} \quad (30)$$

since the sensor set is selected based on Equations (11)-(14), and the set $\mathbb{K}'$ and $\mathbb{K}$ both are the subsets of $\mathbb{G}$. from Equation (13);

$$\delta(\mathbb{K}', h) \leq \delta(\mathbb{K}, h) \quad (31)$$

The above Equation (31) contradicts our assumption of Equation (30). Therefore, based on the studies of CASE-1 and CASE-2, this concludes that there is no other set $\mathbb{K}'(\mathbb{K}' \subseteq \mathbb{G})$ that has the degree of cross-correlation of the selected sensor is higher, and the energy consumption is equal to or lower. Therefore, the parameter set $\mathbb{K}(\mathbb{K} \subseteq \mathbb{G})$ is a Pareto optimal. $\square$

Prediction of sleep parameter at edge node

In this section, we introduce an algorithm that uses an optimal parameter set to predict each sleep sensor parameter. As stated in the problem definition in Section III-B, the sensor node transmits active parameter data to the edge node at the end of each measurement cycle. After receiving data from the sensor node, the edge node applies Algorithm 1 to determine the matrix $\mathbb{K}$, which contains the optimal active parameters used in the prediction of each sleep parameter.

Following the computation of $\mathbb{K}$ using Algorithm 1, in Algorithm 2, the edge node predicts the sleep sensor parameters using optimal active parameters (Lines 1–24). To do this, first edge node sets the value of variable $u = 1$ to train the model (Line 1). Now, the edge node selects the optimal active parameters from matrix $\mathbb{K}$ for each sleep sensor parameter $h(h \in \mathbb{H})$ and trains the model using recently collected samples of optimal active

| <https://doi.org/10.1038/s41598-025-21171-7>

nature portfolio

Algorithm 2. Prediction of sleep parameter's data at Edge Node

In this section, the time complexity of the selection of an optimal parameter set from the active parameter set using proposed C_cBPS approach is calculated and compared with other parameter selection approaches. C_cBPS approach uses algorithm 1 to select the optimal parameter set to predict the sleep sensor parameter h ($h \in \mathbb{H}$).

Research Paper	Time Complexity	Ensure Optimal set
Bhuyan et al. ¹⁰	$O(G^2 + m.n_{iter})$	×
Tripathi et al. ¹¹	$O(GH)$	×
Tu et al. ¹²	$O(G.n_{iter}.n_{models})$	✓
Belhaouari et al. ¹³	$O(G^2.m)$	✓
Alhussan et al. ¹⁴	$O(n_{iter}.(w.(G + C(M))))$	✓
Zhan et al. ¹⁵	$O(G + G.\ S\)$	×
C_cBPS	$O(GH)$	✓

Table 1. Comparison of Computational Time Complexity of Different Parameter Selection Approaches.

approach with the state-of-the-art. In Table 1, n_{iter} denotes the number of algorithm iterations, m represents the number of samples of the parameter set, and $\|S\|$ represents the number of unselected parameters from the parameter set. n_{models} is the total number of machine learning models, $C(M)$ represents the computational complexity of a machine learning model, and w represents the number of waterwheels in the waterwheel plant optimization technique.

Experimental setup

This section presents a description of dataset usage, data preprocessing methods, performance measure criteria, and a comparative analysis of various parameter selection approaches with the proposed approach.

Dataset description and preprocessing

Table 2 shows a description of the datasets used in this experiment, which were taken from different locations and at different sampling intervals²¹. Encompassing the timeframe spanning {To increase the dataset size in IESD-24H for effective model analysis, we include data from the last three months [1 November 2024 - 31 January 2025].} from 1 January 2025 to 31 January 2025, this dataset intricately examines various environmental parameters, notably concerning PM2.5, PM10, NO, CO, NO₂, NH₃, SO₂, Benzene, and Ozone. It is acknowledged that the PM2.5 and PM10 sensors operate on a 60-minute interval. Therefore, to promise reliable data collection at 15 and 30-minute intervals, four and two identical PM2.5 and PM10 sensor values must be used. Initial preprocessing involves employing imputation techniques to address null values. The dataset is then partitioned into training and testing subsets. The training subset encompasses 24 days of data, while the testing subset includes 7 days of data. Ensuring dataset quality necessitates handling missing values, which is accomplished through data cleaning. To rectify missing attribute values, this process replaces them with the mean value of the attribute, ultimately contributing to a reliable dataset for analysis.

Performance measure

This study analyzed the performance of each approach using well-known assessment metrics including Mean Square Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error ($rMSE$), and R^2 Score, which are summarized as follows:

Let M_p and $\bar{M}_p$ represent the measurement vectors of the actual and predicted parameter sequences in connection to the parameter p .

MSE

It's a commonly used metric to measure the average squared difference between actual and model-predicted data. The effectiveness of predictive models is frequently evaluated using it.

$$MSE = \frac{1}{N} \sum (M_p^x - \bar{M}_p^x)^2 \quad (32)$$

$rMSE$

The accuracy of a prediction model is frequently assessed using the Root Mean Square Error ($rMSE$) statistic. It has the advantage of being expressed in the same unit as the original data, making it simpler to understand and compare to the actual values.

$$rMSE = \sqrt{\frac{1}{N} \sum (M_p^x - \bar{M}_p^x)^2} \quad (33)$$

MAE

Mean Absolute Error (MAE) indicates the average absolute difference between predicted and actual values.

$$MAE = \frac{1}{N} \sum |M_p^x - \bar{M}_p^x| \quad (34)$$

Dataset	Location	Interval	No. of Samples
IESD-15	IESD-BHU, Varanasi	15 Minute	2962
IESD-30	IESD-BHU, Varanasi	30 Minute	1481
IESD-1H	IESD-BHU, Varanasi	1 Hour	741
IESD-4H	IESD-BHU, Varanasi	4 Hour	186
IESD-8H	IESD-BHU, Varanasi	8 Hour	93
IESD-24H	IESD-BHU, Varanasi	24 Hour	92
BTM-L-B	BTM Layout, Bengaluru	15 Minute	2947
IITM-D	IITM-Delhi	15 Minute	2943
AN-RSPCB	Adarsh Nagar, Jaipur	15 Minute	2944

Table 2. Description of Datasets used in the Experiments.

R^2 Score

It demonstrates how well the data adhere to the regression model.

$$R^2 \text{ Score} = 1 - \frac{\sum (M_p^x - \bar{M}_p^x)^2}{\sum (M_p^x - \hat{M}_p^x)^2} \quad (35)$$

where $\hat{M}_p^x$ represents the mean of the actual data.

Comparison of Parameter Selection Approaches

The proposed parameter selection approach is compared to the traditional approach where all parameters are used to predict the sleep sensor parameters and other parameter selection approaches, including gradient descent (G.Descent)¹⁰, chi-squared (CHI-S)¹¹, binary water-wheel plant optimization (bWWPA)¹⁴, and cross-validated recursive feature elimination (RFECV)¹². In all the above parameter selection approaches, the chi-squared (CHI-S)¹¹ approach selects the static (fixed) number of active sensor parameters for each sleep sensor parameter. To ensure a fair and consistent comparative analysis, we set this fixed number equal to the average number of active parameters.

RESULTS AND DISCUSSION

This section discussed the comparative analysis of proposed C_eBPS approach with the other state-of-the-art parameter selection approaches. Simulation results show that C_eBPS approach performs superiorly in all comparison metrics.

Parameter setting

This part discusses the simulation results of proposed C_eBPS approach on different datasets, as outlined in Part IV. To simulate the proposed approach, air pollution monitoring datasets²¹ composed of nine sensors measuring the unique parameters (PM2.5, PM10, NO, CO, NO2, NH3, SO2, Benzene, Ozone) from the environment is considered. Figure 2 shows the heat map of the cross-correlation coefficient between all sensor parameters that show how sensor parameters are correlated in the parameter set. The simulations are performed on a Dell Precision Workstation T7910, which is equipped with two CPU sockets containing eight Dual Intel Xeon Processors E5-2667 (8C 16HT, 20 MB Cache, 3.2 GHz) apiece and 256 GB of RAM. This workstation's operating system, Ubuntu 20.4, was chosen for its dependability and performance. To validate the proposed strategy, simulations are conducted over 100 measurement cycles to mitigate the impact of extreme scenarios.

Performance analysis

This section analyses the performance of the proposed C_eBPS parameter selection approach with state-of-the-art parameter selection approaches. The comparison is focused on three essential criteria: (1) the number of active sensor parameters selected by each approach, (2) the total energy consumption at the edge node, and (3) the accuracy in predicting sleep sensor parameter values using C_eBPS approach. These metrics highlight the efficiency of the proposed C_eBPS approach over other state-of-the-art parameter selection approaches.

Comparison of parameter selection approach

The performance of C_eBPS approach is compared with the existing state-of-the-art parameter selection approach. Figure 3 compares the number of selected active parameters and Table 3 shows the maximum, average,

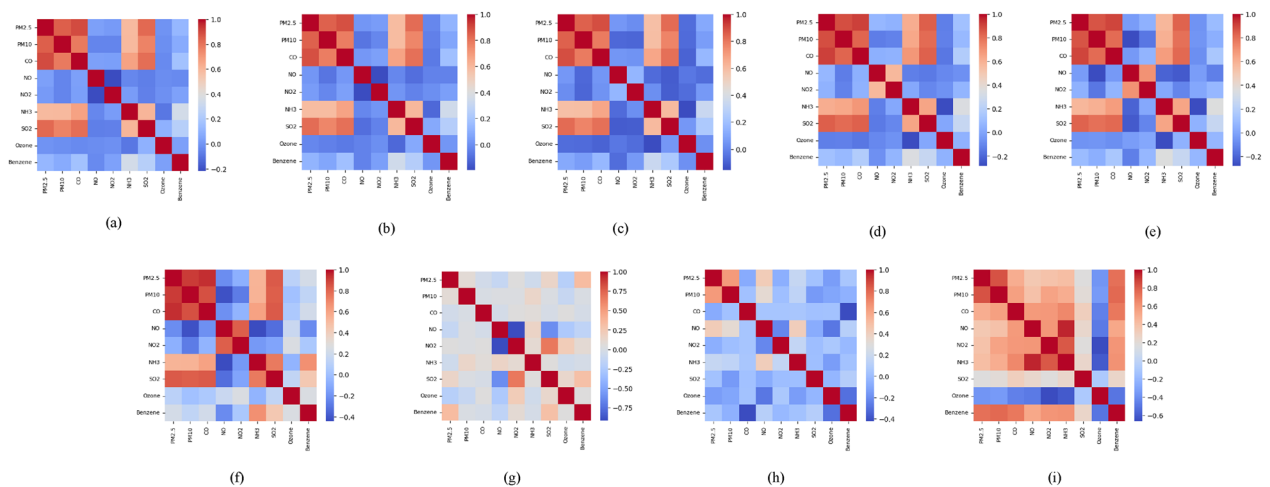


Figure 2. Heatmap of cross-correlation coefficient between all sensor parameters of all Datasets: (a) IESD-15, (b) IESD-30, (c) IESD-1H, (d) IESD-4H, (e) IESD-8H, (f) IESD-24H, (g) BTM-L-B, (h) IITM-D, (i) AN-RSPCB. {Created using Google Colab, Version: Python 3.10, <https://colab.research.google.com/>}.

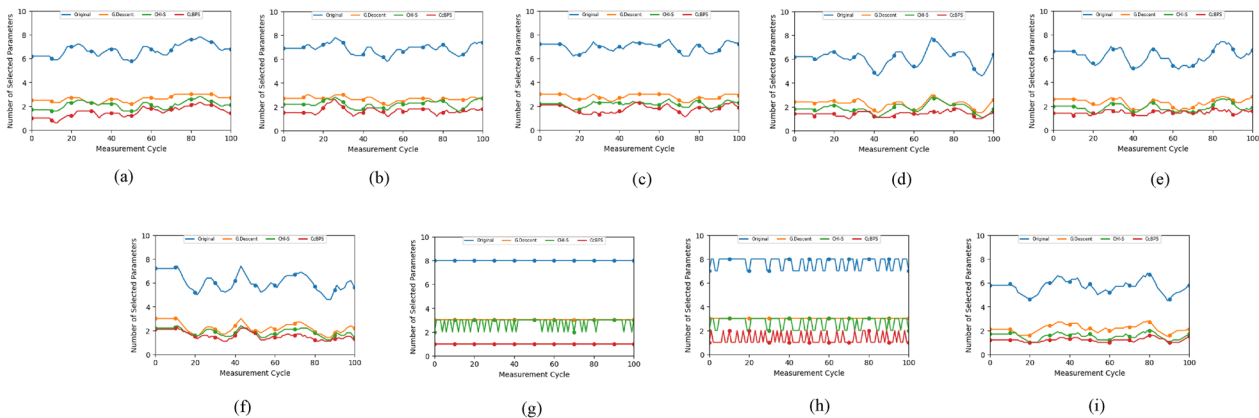


Figure 3. Selected parameters from active sensor parameter by different non-optimal parameter selection approaches among all datasets: (a) IESD-15, (b) IESD-30, (c) IESD-1H, (d) IESD-4H, (e) IESD-8H, (f) IESD-24H, (g) BTM-L-B, (h) IITM-D, (i) AN-RSPCB. {Created using Google Colab, Version: Python 3.10, <https://colab.research.google.com/>}.

Number of Selected Active Parameters												
Dataset	Maximum				Average				Minimum			
	Original	G.Descent	CHI-S	CcBPS	Original	G.Descent	CHI-S	CcBPS	Original	G.Descent	CHI-S	CcBPS
IESD-15	8.00	3.00	3.00	3.00	6.76	2.61	2.14	1.57	5.00	2.00	2.00	1.00
IESD-30	8.00	3.00	3.00	3.00	6.87	2.64	2.22	1.71	5.00	2.00	2.00	1.00
IESD-1H	8.00	3.00	3.00	3.00	6.93	2.70	2.22	1.71	5.00	2.00	2.00	1.00
IESD-4H	8.00	3.00	3.00	3.00	6.13	2.29	1.83	1.48	4.00	1.00	1.00	1.00
IESD-8H	8.00	3.00	3.00	3.00	6.26	2.36	1.89	1.57	4.00	1.00	1.00	1.00
IESD-24H	8.00	3.00	3.00	3.00	6.19	2.32	1.87	1.58	4.00	1.00	1.00	1.00
BTM-L-B	8.00	3.00	3.00	1.00	8.00	3.00	2.64	1.00	8.00	3.00	2.00	1.00
IITM-D	8.00	3.00	3.00	2.00	7.64	3.00	2.80	1.27	7.00	3.00	2.00	1.00
AN-RSPCB	8.00	3.00	3.00	2.00	5.68	2.16	1.52	1.24	4.00	1.00	1.00	1.00

Table 3. Maximum, minimum, and average number of selected parameters from active sensor parameters using different non-optimal parameter selection approaches among all datasets.

and minimum number of selected active parameters used in the prediction of sleep sensor parameters using the proposed and other state-of-the-art parameter selection approaches on all datasets. From Figure 3 and Table 3, our observations are summarized as follows:

1. The motivation behind the *CcBPS* parameter selection approach is to select the highly correlated parameters for predicting sleep sensor parameters. When highly correlated parameters exist in predicting sleep sensor parameters, *CcBPS* selects all the highly correlated parameters for the prediction as shown in Equation 11. For example, in the IESD-4H dataset, Figure 3d, if NO, NH3, SO2, CO, Ozone, and Benzene are active to sense the environmental parameters in a measurement cycle, and PM2.5, PM10, and NO2 are in sleep mode. The proposed approach selects only CO to predict sleep parameter PM2.5 because among all active parameters, CO exhibits a high correlation with PM2.5. (Corr(PM2.5, CO) = 0.91). On the other hand, CHI-S¹¹ uses a static number of parameters, which limits their ability to capture dependencies. The G.Descent¹⁰ technique focuses on discarding disparaging parameters from the active parameter set. Therefore, to preserve the dependencies among parameters, the G.Descent technique requires more number of active parameters to predict the sleep sensor parameter.
2. In Figure 3a–f,i where parameters in the active parameter set exhibit a medium²⁰ correlation with the sleep sensor parameters, the proposed approach selects only those that exhibit a correlation greater than the average correlation with all active parameters to the sleep sensor parameters as shown in Equation 12. The number of selected parameters by the *CcBPS* approach is always less than or equal to the CHI-S and G.Descent approach. For example, in the IESD-1H dataset, Figure 3(c), if NO, NO2, NH3, SO2, Ozone, and Benzene are active to sense the environmental parameters in a measurement cycle, and PM2.5, PM10, and CO are in sleep mode. The proposed approach selects only SO2 to predict sleep parameter CO because, among all active parameters, SO2 exhibits a correlation greater than or equal to the average correlation of all active sensor parameters. Whereas in the CHI-S parameter selection approach, NH3, SO2, and Benzene have CHI-S

Dataset	MSE				rMSE				MAE			R^2 Score			
	Original	G.Descent	CHI-S	CcBPS	Original	G.Descent	CHI-S	CcBPS	Original	G.Descent	CHI-S	Original	G.Descent	CHI-S	CcBPS
IESD-15	1.00	1.84	1.29	0.88	0.86	1.10	0.99	0.83	0.64	0.98	1.30	0.96	0.94	0.85	0.96
IESD-30	1.13	2.76	1.86	0.95	0.95	1.48	1.21	0.87	0.72	1.04	0.88	0.93	0.83	0.89	0.95
IESD-1H	2.43	3.47	2.47	2.29	1.31	1.56	1.31	1.26	1.09	1.77	1.10	0.83	0.76	0.82	0.85
IESD-4H	3.45	3.70	3.53	3.33	1.34	1.40	1.37	1.33	1.03	1.03	1.01	0.76	0.72	0.75	0.76
IESD-8H	2.21	3.87	2.83	1.93	1.09	1.43	1.23	1.02	0.85	0.89	0.92	0.76	0.60	0.72	0.79
IESD-24H	1.49	2.81	1.42	1.32	0.74	0.94	0.72	0.69	0.59	0.78	0.58	0.84	0.74	0.84	0.85
BTM-L-B	0.18	1.59	0.33	0.08	0.33	1.16	0.47	0.23	0.29	0.91	0.37	0.89	0.73	0.78	0.92
ITM-D	0.17	1.50	1.60	0.38	0.38	0.79	0.76	0.36	0.24	0.41	0.28	0.77	0.56	0.68	0.86
AN-RSPCB	0.85	1.01	0.70	0.60	0.77	0.88	0.76	0.67	0.44	0.71	0.72	0.87	0.72	0.71	0.90

Table 4. Comparison of performance metrics for different non-optimal parameter selection approaches with C_eBPS approach among all datasets.

values greater than the average CHI-S value of all active parameters, and similarly gradient descent approach, NH3, SO2, Benzene, and Ozone are selected to predict the sleep sensor parameter CO.

Table 4 shows the comparison of performance metrics of all parameter selection approaches on different datasets. The result shows that the C_cBPS approach surpasses all error metrics, such as MSE, rMSE, MAE, and R^2 Score because among all the parameter selection approaches, the C_cBPS approach selects the parameters that are highly correlated with the sleep sensor parameters, thus it reduces the errors in the prediction of sleep sensor parameters.

Comparison of total energy consumption at edge node

This section discusses the total energy consumption of all parameter selection approaches over all datasets. The energy consumption is defined as the sum of the amount of energy used to select the parameter set from the active sensor parameters and use those parameters to predict the sleep sensor parameter at the edge node. Let's assume that processing speed or hardware performance remains constant or negligible in their variation to study the impact of the number of parameters in the prediction. Let us assume that ' Δ ' is the energy consumed in the computation in unit time. The prediction energy (e_{pred}) is defined as follows:

$$e_{pred} = \Delta \times N_{\mathbb{K}}(\mathbb{K} \times T_{dist} + T_{kernel} + T_{mv}) \quad (36)$$

where $N_{\mathbb{K}}$ represents the number of samples collected by active sensors, $\mathbb{K}$ represents the number of parameters, and T_{dist} , T_{kernel} , and T_{mv} represent the time to compute the distance of data points, kernel matrix, and mean and variance of new data points, respectively. Now, the total energy consumption at the edge node is defined as follows:

$$ENERGY_{Edge} = \Delta \times T_{selection} + e_{pred} \quad (37)$$

where $T_{selection}$ is the time consumed to select the parameter set from the active sensor parameters. We assume the value of $\Delta = 1mJ$, which will be used as the basis for our calculations.

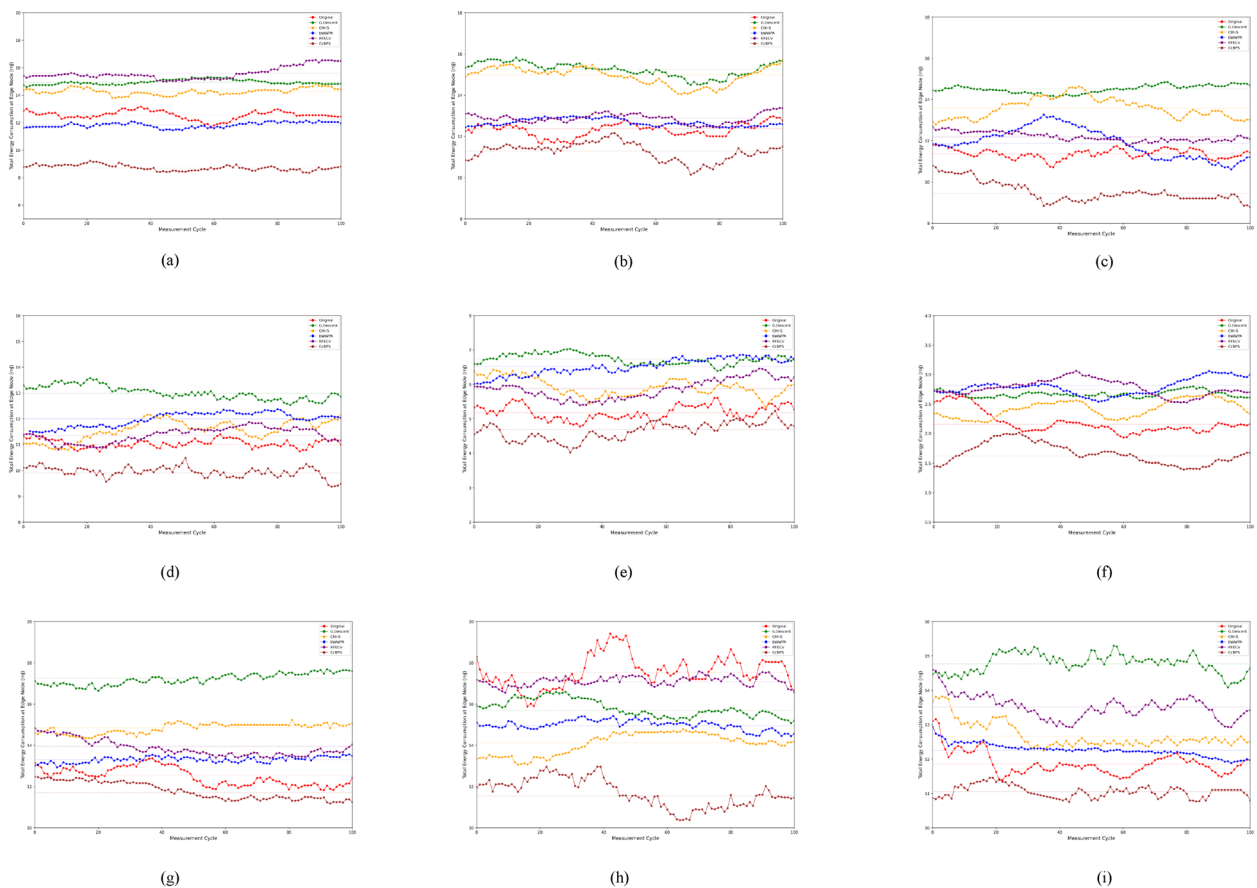


Figure 4. Total energy consumption comparison using different parameter selection approaches at edge node among all datasets: (a) IESD-15, (b) IESD-30, (c) IESD-1H, (d) IESD-4H, (e) IESD-8H, (f) IESD-24H, (g) BTM-L-B, (h) IITM-D, (i) AN-RSPCB. {Created using Google Colab, Version: Python 3.10, <https://colab.research.google.com/>}.

Energy Consumption at Edge Node (mJ)															
Dataset	Maximum					Average					Minimum				
	Original	G.Descent	CHI-S	bWWPA	RFECV	CcBPS	Original	G.Descent	CHI-S	bWWPA	RFECV	CcBPS	Original	G.Descent	CHI-S
IESD-15	16.97	17.12	20.52	15.85	27.19	10.93	12.60	14.87	14.25	11.85	15.67	8.70	8.33	10.23	7.98
IESD-30	15.16	22.18	21.15	19.11	17.65	17.77	12.40	15.33	15.07	12.63	12.98	11.26	8.33	8.54	10.23
IESD-1H	14.20	17.59	19.14	15.90	15.82	14.83	11.39	14.98	13.99	11.96	12.24	9.53	7.64	10.65	7.70
IESD-4H	14.10	17.94	15.41	15.64	13.98	13.68	11.02	12.94	11.71	11.95	11.28	9.90	7.52	7.82	7.00
IESD-8H	7.99	9.37	8.94	8.95	9.47	7.84	5.14	6.75	5.96	6.39	5.84	4.72	3.19	3.66	3.68
IESD-24H	3.31	4.10	3.76	4.21	4.13	2.75	2.18	2.69	2.41	2.79	2.71	1.58	1.48	1.63	1.61
BTM-L-B	15.88	23.69	21.22	18.43	22.16	14.69	12.59	17.36	14.84	13.26	14.25	11.77	7.56	12.72	10.71
IITM-D	26.00	22.88	20.31	25.51	26.43	23.20	17.60	15.67	14.08	14.96	17.08	11.57	8.42	10.84	8.30
AN-RSPCB	20.87	20.26	19.41	19.67	20.77	14.55	11.96	14.70	12.71	12.35	13.63	11.03	8.06	9.53	9.14

Table 5. Comparison of maximum, minimum, and average energy consumption at edge node using different parameter selection approaches among All Datasets.

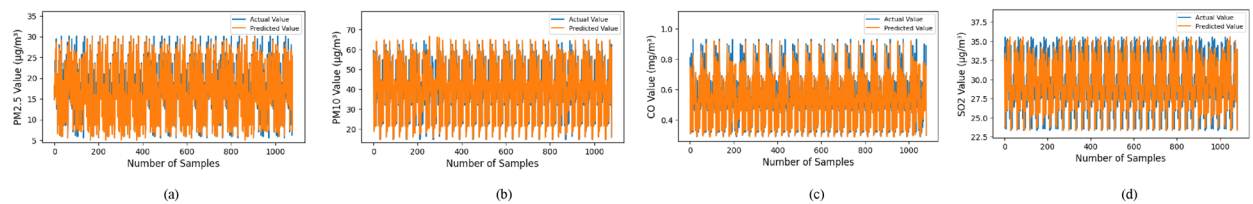


Figure 5. Graph of actual vs predicted sleep sensor parameters value using C_cBPS approach on IESD-15 Dataset: (a) PM2.5, (b) PM10, (c) CO, (d) SO2. {Created using Google Colab, Version: Python 3.10, <https://colab.research.google.com/>}.

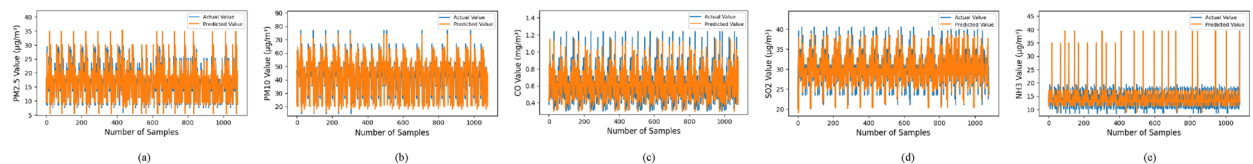


Figure 6. Graph of actual vs predicted sleep sensor parameters value using C_cBPS approach on IESD-30 Dataset: (a) PM2.5, (b) PM10, (c) CO, (d) SO2, (e) NH3. {Created using Google Colab, Version: Python 3.10, <https://colab.research.google.com/>}.

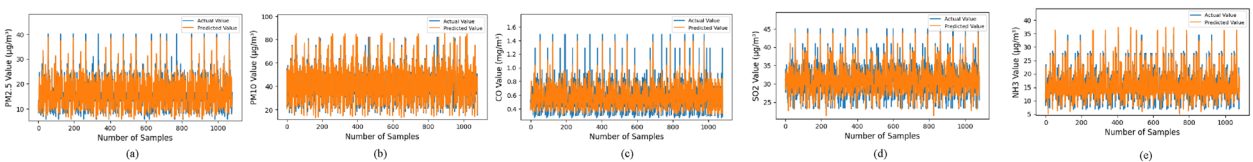


Figure 7. Graph of actual vs predicted sleep sensor parameters value using C_cBPS approach on IESD-1H Dataset: (a) PM2.5, (b) PM10, (c) CO, (d) SO2, (e) NH3. {Created using Google Colab, Version: Python 3.10, <https://colab.research.google.com/>}.

Figure 4 and Table 5 demonstrate the total energy consumption at the edge node by using different approaches on various datasets. The result is depicted by moving average with a window size of 10 to reduce the variations in the graph and make it more readable. The observations are summarized as follows:

1. The C_cBPS approach takes minimum energy in the prediction of sleep sensor parameters compared to state-of-the-art parameter selection approaches.
2. The CHI-S approach selects the static parameters for the prediction of sleep sensor parameters, limiting their performance in the parameter selection. The G.Descent method eliminates only the parameters that do not exploit the dependencies within the input parameters and also takes more computational time to select the parameters from the active parameters. The RFECV method takes every pair of input data to select the best parameter set. Therefore, the RFECV method requires more computation in different machine learning model training, which increases its computation time and energy in the parameter selection process. The bWWPA algorithm optimizes the parameter selection problem using a binary waterwheel plant optimization algorithm. This algorithm itself can take more computation and energy in the optimization process. As a result, RFECV, and bWWPA approaches select parameters by increasing the energy at the edge node.
3. Considering Figure 3 and utilizing Equation (37), C_cBPS the approach achieves approximately {30.90, 9.23, 16.17, 9, 70, 8.10, 27.40, 6.50, 34.20, 7.50} substantial energy reductions compared to the original approach, where all the active parameters are used in the prediction of the sleep sensor parameters on {IESD-15, IESD-30, IESD-1H, IESD-4H, IESD-8H, IESD-24H, BTM-L-B, IITM-D, AN-RSPCB} datasets, respectively.

Accuracy of predicted sleep sensor parameters using C_cBPS approach

This section assesses the accuracy of C_cBPS parameter selection approach to predict sleep sensor parameters over all datasets. To evaluate prediction accuracy, we compare the actual (A) values of sleep sensor parameters to the predicted (P) values acquired by the proposed C_cBPS approach. Figures 5, 6, 7, 8, 9, 10, 11, 12 and 13 visually depict the model's accuracy of the actual vs. predicted parameters value graph in reconstructing the original parameter sequence using C_cBPS approach on all the datasets. In these figures, we consider only those parameters that are in sleep mode in at least one measurement cycle. The parameters that are active in each measurement cycle are not considered in the graph. Table 6 shows the statistical measure of each predicted

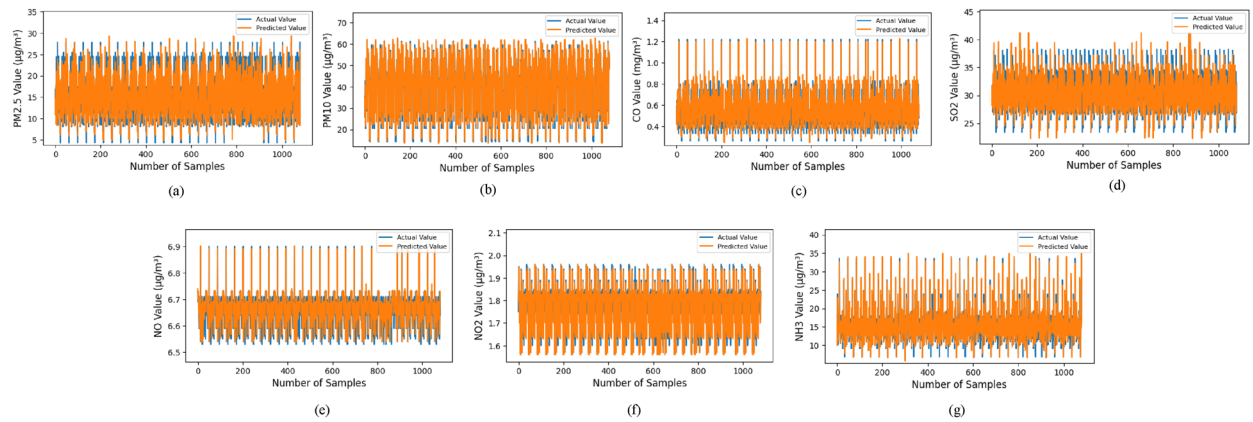


Figure 8. Graph of actual vs predicted sleep sensor parameters value using C_cBPS approach on IESD-4H Dataset: (a) PM2.5, (b) PM10, (c) CO, (d) SO₂, (e) NO, (f) NO₂, (g) NH₃. {Created using Google Colab, Version: Python 3.10, <https://colab.research.google.com/>}.

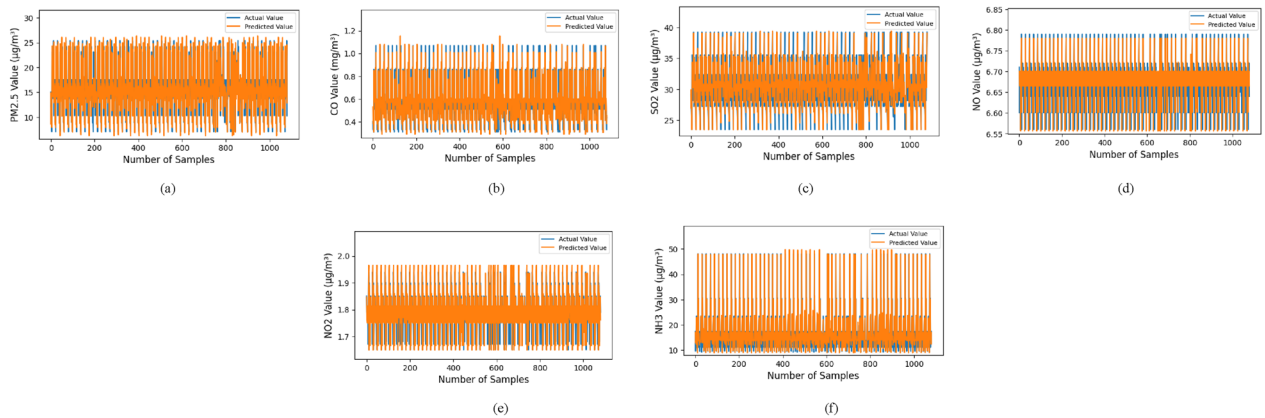


Figure 9. Graph of actual vs predicted sleep sensor parameters value using C_cBPS approach on IESD-8H Dataset: (a) PM2.5, (b) CO, (c) SO₂, (d) NO, (e) NO₂, (f) NH₃. {Created using Google Colab, Version: Python 3.10, <https://colab.research.google.com/>}.

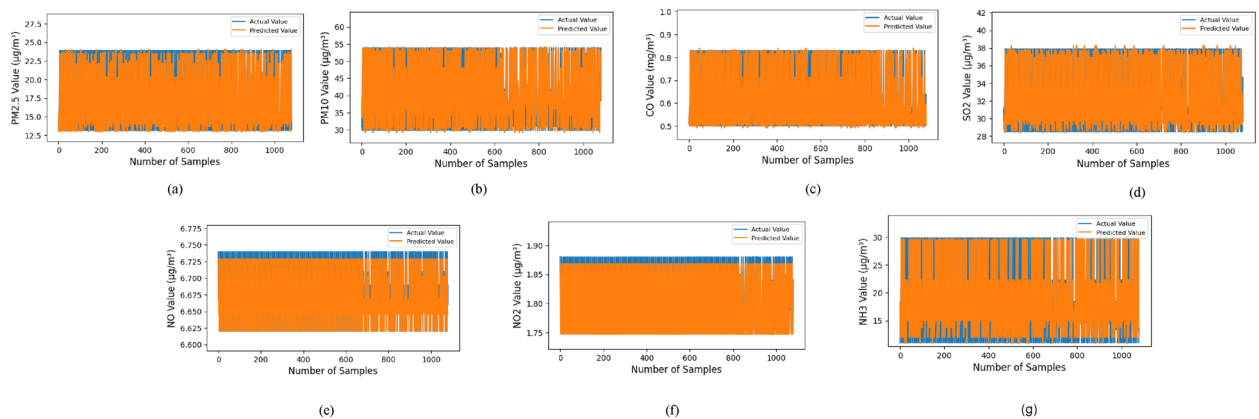


Figure 10. Graph of actual vs predicted sleep sensor parameters value using C_cBPS approach on IESD-24H Dataset (a) PM2.5, (b) PM10, (c) CO, (d) SO₂, (e) NO, (f) NO₂, (g) NH₃. {Created using Google Colab, Version: Python 3.10, <https://colab.research.google.com/>}.

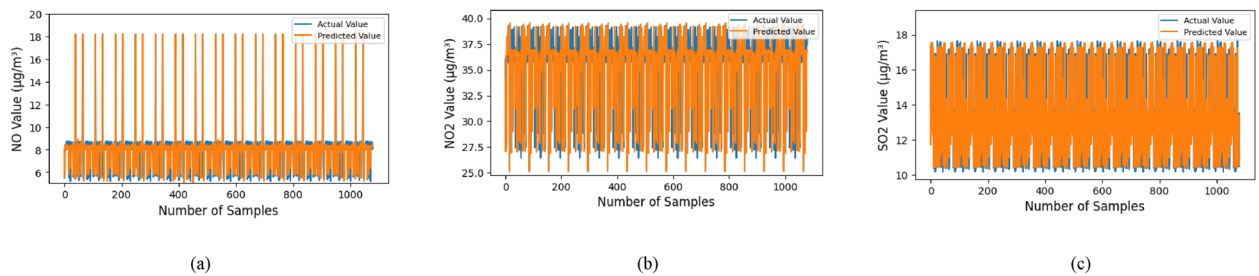


Figure 11. Graph of actual vs predicted sleep sensor parameters value using C_cBPS approach on BTM-L-B Dataset: (a) NO (b) NO₂, (c) SO₂. {Created using Google Colab, Version: Python 3.10, <https://colab.research.google.com/>}.

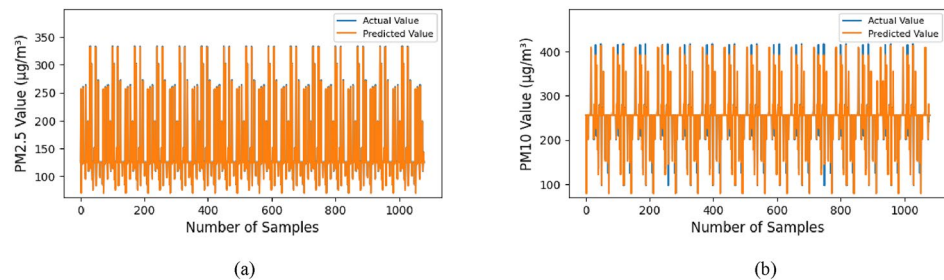


Figure 12. Graph of actual vs predicted sleep sensor parameters value using C_cBPS approach on IITM-D Dataset: (a) PM_{2.5}, (b) PM₁₀. {Created using Google Colab, Version: Python 3.10, <https://colab.research.google.com/>}.

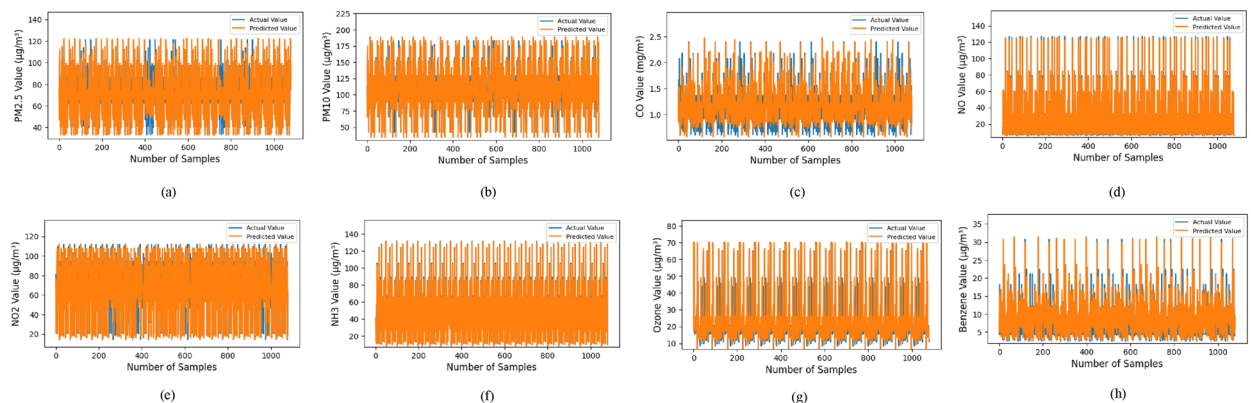


Figure 13. Graph of actual vs predicted sleep sensor parameters value using C_cBPS approach on AN-RSPCB Dataset: (a) PM_{2.5}, (b) PM₁₀, (c) CO, (d) NO, (e) NO₂, (f) NH₃ (g) Ozone (h) Benzene. {Created using Google Colab, Version: Python 3.10, <https://colab.research.google.com/>}.

parameter compared to the actual parameter over all the datasets. This measure includes the minimum value (Min), maximum value (Max), mean value (Mean), median value (Median), standard deviation (std), and variance (Variance) of the distribution of the actual and predicted sleep sensor parameters. We study the performance of the C_cBPS approach by dividing the graphs into the following categories which are summarized as follows:

1. The first category shows a high degree of alignment, with a significant match between the actual and the predicted sleep sensor parameter and having the average error between the actual and predicted parameters is less than 0.05%, showing an efficient model approximation.
2. The second category has a moderate fit, suggesting a good approximation with slight deviations that might be due to built-in difficulties in the dataset or modeling assumptions, and the average error between the actual and predicted parameters ranges from 0.05% – 0.5%.

Dataset	Actual(A) vs Predicted(P) Sensor Parameters																NH3(A)				Benzene(A)				Benzene(P)			
	PM2.5(A)	PM2.5(P)	PM10(A)	PM10(P)	CO(A)	CO(P)	SO2(A)	SO2(P)	NO(A)	NO(P)	NO2(A)	NO2(P)	NH3(A)	NH3(P)	Ozone(A)	Ozone(P)	Benzene(A)	Benzene(P)										
IESD-15	Min	5.56	5.00	14.64	14.71	0.29	0.28	23.34	23.36	-	-	-	-	-	-	-	-	-										
	Max	30.13	29.98	64.49	66.67	0.93	0.91	35.53	35.57	-	-	-	-	-	-	-	-	-										
	Mean	16.90	16.08	40.95	40.86	0.55	0.54	29.48	29.59	-	-	-	-	-	-	-	-	-										
	Median	15.89	15.89	41.27	41.24	0.55	0.53	29.00	29.29	-	-	-	-	-	-	-	-	-										
	std	6.77	6.79	13.88	13.47	0.16	0.15	3.48	3.14	-	-	-	-	-	-	-	-	-										
IESD-30	Variance	45.85	46.17	192.87	181.59	0.02	0.02	12.13	9.88	-	-	-	-	-	-	-	-	-										
	Min	5.65	5.38	14.64	15.80	0.30	0.29	21.05	19.06	-	-	-	-	8.44	8.96	-	-	-										
	Max	34.74	35.30	76.99	74.29	1.24	1.15	40.5	39.78	-	-	-	-	39.48	39.45	-	-	-										
	Mean	17.23	16.93	42.10	43.26	0.60	0.62	29.95	29.93	-	-	-	-	14.45	14.14	-	-	-										
	Median	17.02	16.43	43.53	45.18	0.59	0.59	30.16	29.57	-	-	-	-	14.03	13.27	-	-	-										
IESD-1H	std	5.96	5.55	12.21	12.25	0.19	0.17	3.92	3.63	-	-	-	-	4.23	3.87	-	-	-										
	Variance	35.63	30.87	149.10	150.27	0.03	0.03	15.37	13.19	-	-	-	-	17.95	15.05	-	-	-										
	Min	5.86	5.89	15.35	12.94	0.27	0.32	23.19	21.29	-	-	-	-	6.74	5.08	-	-	-										
	Max	40.15	38.71	81.99	85.31	1.49	1.38	45.08	43.85	-	-	-	-	34.52	37.34	-	-	-										
	Mean	16.64	16.63	42.78	42.79	0.58	0.59	30.62	30.95	-	-	-	-	15.66	15.64	-	-	-										
IESD-4H	Median	15.86	15.83	43.73	43.72	0.55	0.54	30.13	30.45	-	-	-	-	14.60	14.33	-	-	-										
	std	6.43	6.00	15.39	14.97	0.21	0.17	4.12	3.63	-	-	-	-	5.14	4.93	-	-	-										
	Variance	41.40	36.02	236.96	224.33	0.04	0.03	16.97	13.19	-	-	-	-	26.50	24.37	-	-	-										
	Min	4.23	4.85	14.18	13.44	0.26	0.24	23.36	22.23	6.53	6.53	1.56	1.56	6.81	6.76	-	-	-										
	Max	27.88	29.36	61.38	62.80	1.22	1.22	38.31	41.24	6.90	6.90	1.96	1.95	33.58	34.10	-	-	-										
IESD-8H	Mean	15.58	15.49	38.67	39.07	0.54	0.57	30.73	30.74	6.65	6.66	1.76	1.77	15.48	15.70	-	-	-										
	Median	15.65	15.38	38.51	39.31	0.52	0.56	30.92	30.77	6.67	6.67	1.80	1.80	14.70	14.70	-	-	-										
	std	5.99	5.31	13.11	12.92	0.18	0.18	3.78	3.51	0.06	0.05	0.10	0.10	4.96	5.00	-	-	-										
	Variance	35.97	28.23	172.10	167.09	0.03	0.03	14.33	12.35	0.004	0.003	0.01	0.01	14.33	12.35	-	-	-										
	Min	7.03	6.97	-	-	0.31	0.29	23.51	23.48	6.56	6.56	1.65	1.64	9.24	9.17	-	-	-										
IESD-24H	Max	25.40	26.30	-	-	1.07	1.15	39.22	39.40	6.79	6.78	1.94	1.96	48.10	49.70	-	-	-										
	Mean	15.72	15.75	-	-	0.56	0.58	30.84	30.37	6.66	6.67	1.79	1.80	17.05	17.10	-	-	-										
	Median	15.18	15.36	-	-	0.55	0.56	30.34	30.37	6.67	6.67	1.78	1.79	14.18	12.38	-	-	-										
	std	5.01	5.30	-	-	0.18	0.15	4.09	3.38	0.05	0.04	0.07	0.07	8.83	6.16	-	-	-										
	Variance	25.19	28.11	-	-	0.03	0.03	16.75	11.43	0.002	0.002	0.005	0.005	78.11	38.01	-	-	-										
Continued	Min	13.11	13.00	29.77	29.15	0.5	0.49	28.49	28.40	6.62	6.62	1.75	1.74	11.02	10.94	-	-	-										
	Max	23.92	24.04	53.90	54.17	0.83	0.84	37.91	38.28	6.74	6.72	1.88	1.86	29.89	29.73	-	-	-										
	Mean	16.97	17.07	40.12	40.58	0.62	0.61	32.20	31.64	6.67	6.66	1.80	1.79	18.14	17.97	-	-	-										
	Median	15.38	15.88	38.49	38.78	0.58	0.58	31.30	30.59	6.67	6.65	1.79	1.79	17.80	17.36	-	-	-										
	std	3.72	3.53	8.95	8.31	0.12	0.11	2.97	2.85	0.03	0.02	0.04	0.04	5.82	5.23	-	-	-										
	Variance	13.88	12.52	80.12	69.07	0.001	0.001	8.83	8.16	0.0012	0.0007	0.002	0.002	33.90	27.42	-	-	-										

Dataset	Actual(A) vs Predicted(P) Sensor Parameters																Benzene(A)	Benzene(P)	
	PM2.5(A)	PM2.5(P)	PM10(A)	PM10(P)	CO(A)	CO(P)	SO2(A)	SO2(P)	NO(A)	NO(P)	NO2(A)	NO2(P)	NH3(A)	NH3(P)	Ozone(A)	Ozone(P)			
BTM-L-B	Min	-	-	-	-	-	-	10.17	10.41	5.20	5.10	25.32	25.12	-	-	-	-	-	-
	Max	-	-	-	-	-	-	17.65	17.52	18.17	18.17	39.32	39.54	-	-	-	-	-	-
	Mean	-	-	-	-	-	-	13.05	13.06	8.05	8.07	35.12	35.17	-	-	-	-	-	-
	Median	-	-	-	-	-	-	13.34	12.22	8.14	8.17	36.17	36.55	-	-	-	-	-	-
	std	-	-	-	-	-	-	2.72	2.53	2.02	1.94	4.07	3.55	-	-	-	-	-	-
IITM-D	Variance	-	-	-	-	-	-	7.44	6.42	4.10	3.78	16.63	12.67	-	-	-	-	-	-
	Min	70.7	70.1	79	78.27	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Max	333.2	331.6	417.8	414.93	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Mean	140.58	139.69	249.95	254.23	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Median	125.86	124.87	256.47	255.83	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AN-RSPCB	std	53.53	53.39	51.62	52.30	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Variance	2865.69	2851.21	2665.40	2736.21	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Min	33.10	32.99	34.70	33.96	0.57	0.568	-	-	5.63	5.47	14.40	14.12	8.22	8.61	6.7	6.68	2.51	2.60
	Max	121.4	122.56	188.9	189.53	2.40	2.47	-	-	126.97	126.66	112.10	111.73	131.52	130.85	70.25	69.94	30.75	31.43
	Mean	69.54	69.11	108.87	108.13	1.15	1.22	-	-	29.90	29.75	65.19	65.35	49.09	48.78	24.77	24.61	9.32	9.38
	Median	66.30	66.96	102.04	102.59	1.15	1.18	-	-	12.41	11.98	67.18	65.16	38.67	38.87	20.48	19.79	9.03	8.58
	std	24.11	23.75	40.20	38.66	0.40	0.38	-	-	31.83	31.34	30.02	29.48	31.70	31.54	16.32	15.37	5.51	5.37
	Variance	581.72	564.21	1616.47	1495.25	0.16	0.15	-	-	1013.72	982.59	901.43	869.44	1004.9	994.78	266.62	236.53	30.39	28.89

Table 6. Statistical measures of actual vs predicted sleep sensor parameters value using *C_cBPS* approach among all datasets.

Dataset	Figure	Category 1	Category 2	Category 3
IESD-15	5	PM2.5, PM10, CO, SO ₂	-	-
IESD-30	6	PM2.5, PM10, SO ₂	CO, NH ₃	-
IESD-1H	7	PM2.5, PM10, NH ₃	CO, SO ₂	-
IESD-4H	8	PM10, CO, NO ₂ , NH ₃	PM2.5, SO ₂ , NO	-
IESD-8H	9	PM2.5, NH ₃	CO, SO ₂ , NO ₂	NO
IESD-24H	10	PM2.5, PM10, CO, SO ₂	NO, NO ₂	NH ₃
BTM-L-B	11	SO ₂	NO	NO ₂
IITM-D	12	PM2.5, PM10,	-	-
AN-RSPCB	13	PM2.5, PM10, NO ₂ , NH ₃	CO, Ozone	NO, Benzene

Table 7. Categorization of actual vs. predicted graphs based on their average prediction error among all datasets.

- The third category shows a fair approximation, demonstrating the model's ability to catch essential patterns despite inevitable differences and having the average error between the actual and predicted parameters greater than 0.5%.

Table 7 shows the parameters that lie within these categories over all the datasets. These results highlight the effectiveness of the proposed approach because it accomplishes precise signal reconstruction with minor errors for all sleep parameter values.

Conclusion

This paper presents a cross-correlation-based parameter selection (C_cBPS) approach. The proposed approach utilizes parameter cross-correlation to select the highly correlated parameters to predict each sleep sensor parameter at the edge node. The effectiveness of the C_cBPS approach has been validated theoretically. Simulations are also performed on nine publicly available environmental datasets, collected from different geographical locations and time intervals. The proposed approach outperforms the existing state-of-the-art parameter selection approaches and demonstrates an improvement of 6.5% - 34.2% in energy efficiency in the prediction of sleep sensor parameters at the edge node among all datasets.

Data Availability

Data is provided with supplementary files.

Received: 14 May 2025; Accepted: 19 September 2025

Published online: 27 October 2025

References

- Ali, I. et al. Data collection in sensor-cloud: A systematic literature review. *IEEE Access* **8**, 184664–184687 (2020).
- Zhu, T., Li, J., Gao, H. & Li, Y. Latency-efficient data collection scheduling in battery-free wireless sensor networks. *ACM Trans. Sensor Networks (TOSN)* **16**, 1–21 (2020).
- Ghosh, S., De, S., Chatterjee, S. & Portmann, M. Learning-based adaptive sensor selection framework for multi-sensing wsn. *IEEE Sens. J.* **21**, 13551–13563 (2021).
- Ghosh, S. et al. Energy aware smart sensing and implementation in green air pollution monitoring system. In *ICC 2023 - IEEE International Conference on Communications*, 2153–2158, <https://doi.org/10.1109/ICC45041.2023.10279138> (2023).
- Ghosh, S., De, S., Chatterjee, S. & Portmann, M. Edge intelligence framework for data-driven dynamic priority sensing and transmission. *IEEE Trans. Green Commun. Netw.* **6**, 376–390 (2021).
- Gupta, V. & De, S. Collaborative multi-sensing in energy harvesting wireless sensor networks. *IEEE Trans. Signal Inf. Process. Netw.* **6**, 426–441 (2020).
- Gupta, V. & De, S. An energy-efficient edge computing framework for decentralized sensing in wsn-assisted iot. *IEEE Trans. Wireless Commun.* **20**, 4811–4827 (2021).
- Gopika, N. & ME, A. M. K. Correlation based feature selection algorithm for machine learning. In *2018 3rd international conference on communication and electronics systems (ICCES)*, 692–695 (IEEE, 2018).
- Yu, L. & Liu, H. Feature selection for high-dimensional data: A fast correlation-based filter solution. In: *Proc. 20th international conference on machine learning (ICML-03)*, 856–863 (2003).
- Bhuyan, H. K., Chakraborty, D. C., Pani, S. K. & Ravi, V. Feature and subfeature selection for classification using correlation coefficient and fuzzy model. *IEEE Trans. Eng. Manage.* **70**, 1655–1669. <https://doi.org/10.1109/TEM.2021.3065699> (2023).
- Tripathi, G., Singh, V. K., Sharma, V. & Vinodbhai, M. V. Weighted feature selection for machine learning based accurate intrusion detection in communication networks. *IEEE Access* (2024).
- Tu, T., Su, Y., Tang, Y., Tan, W. & Ren, S. A more flexible and robust feature selection algorithm. *IEEE Access* (2023).
- Belhaouari, S. B., Shakeel, M. B., Erbad, A., Oflaz, Z. & Kassoul, K. Bird's eye view feature selection for high-dimensional data. *Scientific Reports* (2023).
- Alhussan, A. A. et al. A binary waterwheel plant optimization algorithm for feature selection. *IEEE Access* (2023).
- Zhan, J., Huang, X., Qian, Y. & Ding, W. A fuzzy c-means clustering-based hybrid multivariate time series prediction framework with feature selection. *IEEE Transactions on Fuzzy Systems* (2024).
- Fan, Y. et al. Learning correlation information for multi-label feature selection. *Pattern Recogn.* **145**, 109899 (2024).
- Jadaliha, M., Xu, Y., Choi, J., Johnson, N. S. & Li, W. Gaussian process regression for sensor networks under localization uncertainty. *IEEE Trans. Signal Process.* **61**, 223–237 (2012).

18. Rasmussen, C. E. & Williams, C. K. Gaussian processes for machine learning cambridge. MA: the MIT Press.[Google Scholar] (2006).
19. Williams, C. K. & Rasmussen, C. E. *Gaussian processes for machine learning* Vol. 2 (MIT press Cambridge, MA, 2006).
20. Schober, P., Boer, C. & Schwarte, L. A. Correlation coefficients: appropriate use and interpretation. *Anesthes. Analgesia* **126**, 1763–1768 (2018).
21. CPCB. Air quality data. <https://airquality.cpcb.gov.in> (2025).

Author contributions

Ruchir Gupta and Sumit Kumar reviewed the Manuscript and contributed to propose the idea Vipin Maurya proposed and implemented the idea Sonali Raj reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-21171-7>.

Correspondence and requests for materials should be addressed to R.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025