Mittal *et al. Journal of Cloud Computing (2022) 11:70*
https://doi.org/10.1186/s13677-022-00344-z

Journal of Cloud Computing:
Advances, Systems and Applications

RESEARCH

Open Access

An efficient quality of services based wireless sensor network for anomaly detection using soft computing approaches

Mohit Mittal[1†], Martyna Kobielnik[2†], Swadha Gupta[3†], Xiaochun Cheng[4†] and Marcin Wozniak[2*†]

Abstract

Wireless sensor network (WSN) is widely acceptable communication network where human-intervention is less. Another prominent factors are cheap in cost and covers huge area of field for communication. WSN as name suggests sensor nodes are present which communicate to the neighboring node to form a network. These nodes are communicate via radio signals and equipped with battery which is one of most challenge in these networks. The battery consumption is depend on weather where sensors are deployed, routing protocols etc. To reduce the battery at routing level various quality of services (QoS) parameters are available to measure the performance of the network. To overcome this problem, many routing protocol has been proposed. In this paper, we considered two energy efficient protocols i.e. LEACH and Sub-cluster LEACH protocols. For provision of better performance of network Levenberg-Marquardt neural network (LMNN) and Moth-Flame optimisation both are implemented one by one. QoS parameters considered to measure the performance are energy efficiency, end-to-end delay, Throughput and Packet delivery ratio (PDR). After implementation, simulation results show that Sub-cluster LEACH with MFO is outperforms among other algorithms.Along with this, second part of paper considered to anomaly detection based on machine learning algorithms such as SVM, KNN and LR. NSLKDD dataset is considered and than proposed the anomaly detection method.Simulation results shows that proposed method with SVM provide better results among others.

Keywords: LEACH, Quality of services, Soft computing techniques, Moth-Flame optimisation, Anomaly detection

Introduction

Nowadays, Wireless sensor network (WSN) is first priority for wireless communication systems that works on huge number of sensor nodes or motes. WSN is either have homogeneous or heterogeneous based sensor nodes (SN) [1]. These are fully depended on the applications where it is deployed. The sensor network is majorly dependent on batteries as they all are equipped with it. There exists many activities inside the each sensor node such as communicating with neighboring node, sensing the parameters such as temperature, humidity etc, processing of data such as collecting it, making it into various packets and transmitting to the destination [2]. The collected data are transmitted to sink node. Base station (BS) is the node which responsible for collecting all data from the network and transmitting further for further analysis. WSN is generally application based networks which are specifically designed to sense the particular parameter. These are deployed in any harsh environment and are easily configurable in nature that why are most popular and cheap in nature. The sensor nodes are mostly work for monitoring as well as tracking purposes. For instance, in hospitals , the patients need high intensive care so there are many sensors are equipped to store and analysis the patients health to monitor it per second [3].

†Mohit Mittal, Martyna Kobielnik, Swadha Gupta, Xiaochun Cheng and Marcin Wozniak contributed equally to this work.

*Correspondence: marcin.wozniak@polsl.pl

[2] Faculty of Applied Mathematics, Silesian University of Technology, Gliwice 44100, Poland
Full list of author information is available at the end of the article

Therefore, wireless sensor network in this case is highly recommended.In addition to this, sensor nodes are manage to cover wide area of sensing with low cost. Moreover, WSN is also used for tracking purposes. The animal tracking is mostly tough task in the wild life. The animal those are equipped with the sensors are easily tracked in the wide spread wild area [4]. WSN has changed the network communication field drastically because the major manufacturing industries are running continuously and required highly monitoring 24 hours so the sensor network works efficiently and effectively [5]. In enhancement in the recent hardware technologies with the introducing micro-electromechanical system (MEMS) [6]. The current scenario is drastically changed and provides more innovative platform for cost-effective as well as more efficient communication network. These sensor nodes (SNs) have abilities of detecting, figuring and imparting through radio frequencies [7, 8]. WSN deployed in brutal and unfriendly condition where human intercession is at zero level.WSN are normally sent for checking and following purposes. If there should be an occurrence of checking purposes like coming down level observing, gaseous tension observing, concoction vapor observing and so forth though if there should be an occurrence of following purposes like creature following, human following and so on [9, 10].

In Fig. 1, blue colored nodes are shown as cluster head nodes (CH). They are selected for particular round during the protocol runs. Every round the cluster head nodes are changed to improve the network lifetime. The yellow colored nodes represented by N-CH. The CH, N-CH and Base station all together made wireless sensor network which communicate via radio signals. All sensor nodes are used omini-channel based signals to communicate their neighboring nodes. N-CH nodes closer to base station are responsible for transmitting the data from all other CH node to base station. The red colored line represents communication between sensor node and CH node whereas black line represents communication between CH node and other CH node or Base station. There is need of efficient setup of communication to persist the network lifetime as all sensor nodes are battery equipped. To further enhancement in the network lifetime always need of very efficient and productive protocol [11]. The communication is via radio signals i.e. ubiquitous in nature. The signal is listening by all neighboring nodes, but the particular sensor only responds based on node id embedded message and other nodes discard the message. The protocol is main soul to manage all functionality such as sensing data, aggregation of data, control overheads, manage control messages, query generation and transmission strategies.

LEACH protocol

LEACH protocol [12], is an efficient way to manage the data in the form of clustering in wireless sensor network. It is highly inspired and come under hierarchical protocols. The formation of clusters round by round fashion is idea which effectively enhance the network lifetime. Cluster are the connectivity between the neighboring

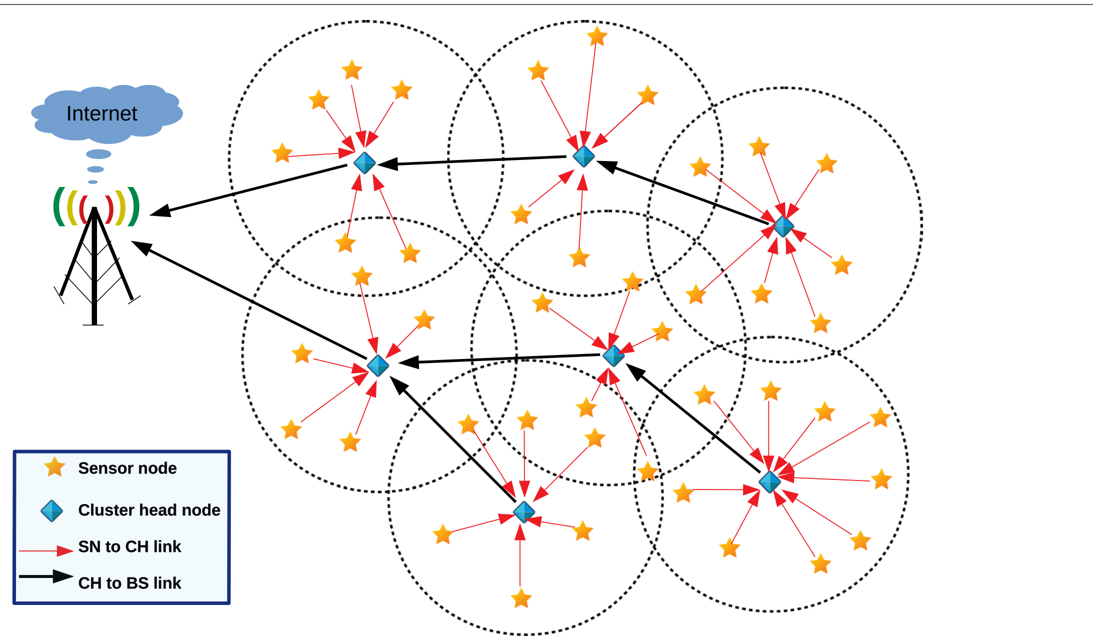

Fig. 1 Conceptual representation of WSN

sensor nodes to communicate each other and transmit their data to single sensor node who aggregate and take care to send aggregated data to base station at final stage [13]. The clustering is performed due to reducing the network congestion which traditionally a big problem in protocols but in LEACH it is solved by introducing the clustering. Clustering helps in managing the data into sub-groups rather than sending directly to base station. Sub-groups aggregates data from the connected neighboring sensors at single node that node is known as CH node [14, 15]. This node has specific functionality of collaborating with neighboring nodes, collecting the sensed data and finally transmitting this data to base station efficiently. The other nodes which is part of sub grouping is known as N-CH nodes. A non-persistent CSMA is used as the MAC protocol in LEACH. From N-CH nodes, i number of nodes elects itself as CH node with probability $P_i(t)$. The selection of CH node probability depends on whether node never act as CH node in past rounds and remaining energy is higher. Each CH node sends an advertisement message. In case of N-CH nodes, it receives many advertisement messages from many CH nodes to join the cluster for particular round. The nodes inform the CH using a join-request message [16]. Later, data is transmitted to intact CH nodes to BS.The collection of data at CH nodes from their neighboring N-CH nodes, the whole process is known as data aggregation at cluster head nodes. The data from CH node is not send as it is collected. It is first compressed to reduce the number of bytes to save energy consumption of transmission and send to BS via intermediate nodes.

According to prior study of protocols, LEACH protocol is very energy efficient protocol in terms of energy which is on of the most required in WSN as all SNs are equipped with battery power. But still, there is a lot of quality improvement required as these are implemented in wide spread area and applications.Therefore, we are proposing a new way to enhance the network lifetime of WSN [17]. The proposed approach is known as Sub-Cluster LEACH protocol. In this approach, the BS send a query to the network for data, then whole network forms the clusters based on the remaining energy of the SNs and concentration of the SNs. As the SNs are having high priority to form cluster are defined as Cluster Heads. The responsibility of the CH node is aggregation of the data from the neighboring SNs [18]. As data is collected at CH node and it transmit to the BS via intermediate nodes. Here, the CH node energy is drain at very high rate . To save the energy of the CH node and restrict the energy consumption at a greater level, A sub-cluster is introduced. For instance, A CH node is have 10 N-CH nodes then all N-CH nodes are transmitting their data to single Node so there is high burden on that node, in that case , CH node select two more node in its vicinity of particular round [19]. That nodes are known as Sub-CH nodes. These nodes are selected on the basis of the remaining energy. These two nodes are act as CH for all nodes in the vicinity of CH node. The neighboring nodes of Sub-CH nodes are selected bu N-CH nodes then they collect their data [20]. After all three nodes such as CH node and two Sub-CH node are act as CH nodes and N-CH nodes are sending their data to one of these node. After data aggregation on all three nodes, Sub-CH send their data to BS via intermediate node. Therefore, The data transmission rate high but the number of byte per CH node is reduced as Sub-CH nodes are transmitting the data to BS [21, 22]. Figure 2 representing the Sub-Cluster LEACH conceptual structure where sub-CH nodes are represented in green color, yellow color nodes are N-CH nodes and blue color nodes are CH nodes. The there is are close between Sub-CH nodes to CH nodes as well as N-CH nodes to Sub-CH nodes.

Related work

LEACH protocol [23–27] follows the hierarchical structure where cluster based data aggregation is completed. LEACH protocol performs two phase to complete one round and rounds are ended up with last node dead from wireless sensor network. Cluster formation and data transmission are two phases. The cluster formation phase is generally related to selection sensor node that will act as CH nodes. The all other nodes have to select one of the nearest neighboring CH node for data transmission. In the second phase i.e. Data transmission phase whole data is collect are respective CH nodes from N-CH nodes and further compressed and send to base station. The data transmission responsibility is upon CH node only but that can be send either directly to BS if the BS is situated at the near to CH or transmitted via intermediate nodes i.e. CH or N-CH nodes which acts as intermediate for forwarding the data to BS [23]. The data transmission in LEACH protocol is multi-hop fashion. The data aggregation and transmission consumed maximum of energy of CH node. To improve the performance of network, the next very round previously select CH nodes are not taken under consideration for the selection CH node for next round. If this will not happen than CH nodes diminishes their energy beyond the threshold level as result of its CH node is no more in use for network. In the beginning this found to be severe problem, but it is handled by CH rotation [24]. In CH rotation, after every round the CH nodes are rotated i.e. eliminates that CH nodes which were already acted as CH node for next round selection. This found to be

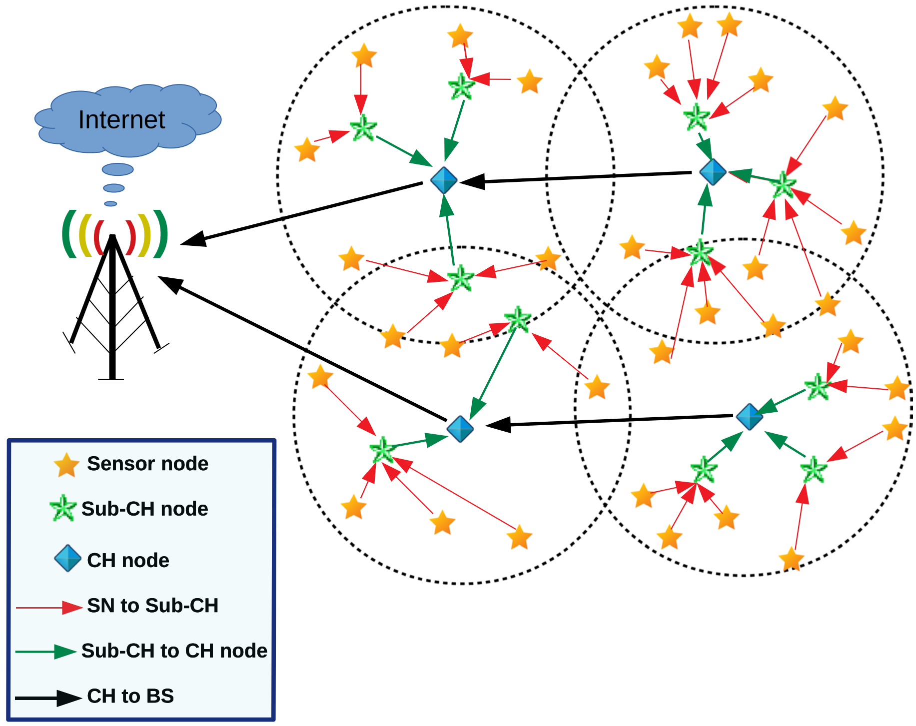

Fig. 2 Sub-cluster LEACH conceptual structure

as marginally increase in the overall network lifetime. But still, there is need of improvement in the protocol. LEACH-C is defined as exclusive version LEACH. This protocol provides majorly location specific information of N-CH nodes to CH nodes [25]. The location information helps in the accessing the sensor nodes location where if a sensor node will going to die and network generate a query for battery replacement or alternative to deploy new sensor node a specific location. In addition to this one simulated annealing approach is also embedded by the candidate node to reduce the objective function. This helps in the reducing the energy consumption of each N-CH while data transmission to their designated CH node. Then after aggregated data at CH nodes is transmitted to BS as like LEACH protocol [28]. The location information is generated by embedding the GPS in each sensor node. This helps in selected of CH node at the center of cluster formation through which proper aggregated on data and the greatest number of reachable N-CH nodes. The energy consumption is reduced for each node but due to GPS installation the energy consumption is not much impact on overall network lifetime. LEACH energy betweenness (LEACH-EB) approach is to enhance the network lifetime at a greater extent [26]. LEACH-EB [29] is very efficient protocol in which minimum variance is calculated during clustering formation phase. The energy consumption is reduced at notable remark as CH nodes selection in the distributive fashion [27, 30]. LEACH-E is famous protocol as it uses minimum spanning tree to direct the transmission from CH nodes to Base station. This step is proved to be very efficient as it provide enough information related to intermediate nodes transmission as shortest path. The outcome is enhancement in energy efficiency. IBLEACH is also come into picture as a energy efficient protocol [31]. The energy consumption is reduced at a greater extend in comparison to LEACH and LEACH-E [32, 33].

Sub-cluster LEACH protocol based on LMNN approach

The Sub-cluster LEACH is explained in detail in previous section. According to that clustering in LEACH protocol is done and other two sensor nodes are selected by CH node to act as CH nodes which are known as Sub-CH nodes. The duties of Sub-CH nodes are similar to CH nodes just the difference in the responding the final node. As Sub-CH nodes are responding to CH node only [34, 35]. Now to enhance the energy efficiency to greater extend than there is requirement to efficient approach for selection CH node.Therefore, Levenberg–Marquardt neural network is introduced to Sub-Cluster LEACH protocol for selection of Sub-CH nodes. Figure 3 provides

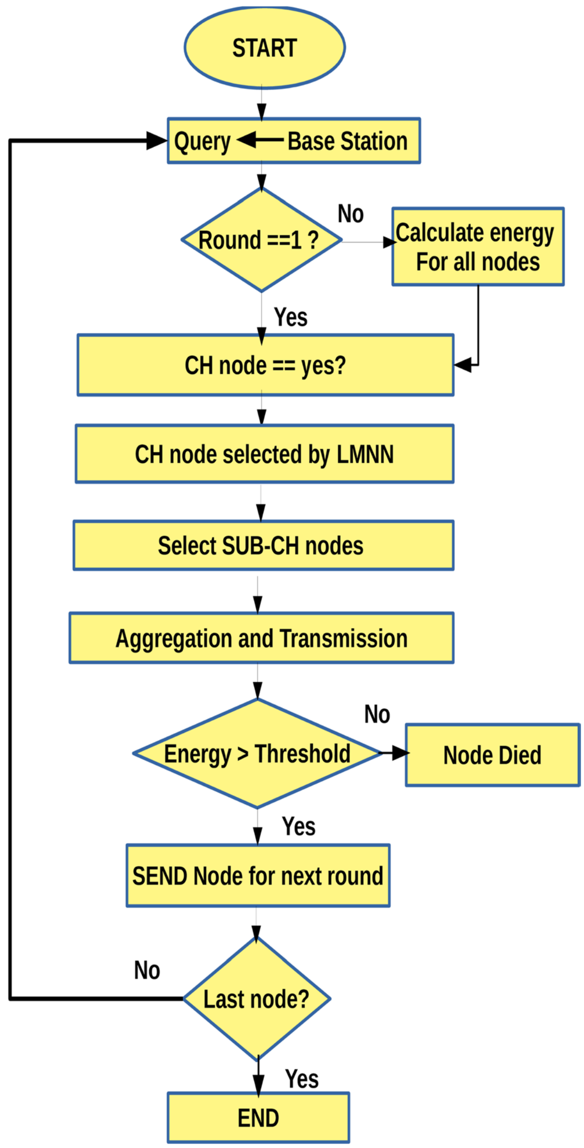

Fig. 3 Step by step representation of Sub-cluster LEACH-LMNN protocol

the step by step description of implementation of protocol. According to this, The base station made a query for data. After this, the network has to be sure whether it is for the first round or not. If yes and proceed further otherwise, there is requirement of evaluation energy of whole sensor nodes before selection CH nodes. If it is first round than all sensor have high and equal energy [36, 37]. Than If the algorithm proposed whether it is CH node or not. If the it is identified as CH node than LMNN is come into picture. CH nodes is over burdened of aggregation of huge amount of data to reduce the burden Sub-CH nodes are selected from N-CH nodes which is successfully done with the help LMNN approach. After selection procedure completed, Aggregation and transmission of data is completed by Sub-CH nodes and CH nodes [38]. After this there is check of energy level of each node. If the energy level of every is higher than threshold energy than it is eligible for next round otherwise the node is declared to be died that means it cannot be considered as part of network anymore. LMNN approach is further in detail described in Algorithm 1.

Algorithm 1 Levenberg Marquardt algorithm

Step 1 Initialize k $\leftarrow$ 0.
Step 2 $\theta_k + \theta(X_k)$ and $\alpha_k + \alpha(X_k)$.

$$\lambda_k = \mu \parallel \theta_k \parallel^{\delta k} \ with \ \delta_k = \begin{cases} \frac{1}{\|\theta_k\|} & \parallel \theta_k \parallel \geq 1, \\ 1 & otherwise \end{cases} \tag{1}$$

Step 3 Solve

$$(\alpha_k^T \alpha_k + \lambda_k I)d = -\alpha_k^T \theta_k, \tag{2}$$

$$(\alpha_k^T \alpha_k + \lambda_k I)d = -\alpha_k^T \theta(y_k), \tag{3}$$

$$(\alpha_k^T \alpha_k + \lambda_k I)d = -\alpha_k^T \theta(z_k), \tag{4}$$

Step 4 Calculate

$$\parallel \theta(x_k + d_k) \parallel \leq \rho \parallel \theta_k \parallel, \tag{5}$$

Step 5 Set

$$d_k = \begin{cases} d_{1k} + d_{2k} + d_{3k} \ \theta_k^T J_k(d_{1k} + d_{2k} + d_{3k}) \leq -\gamma \\ d_{1k} \qquad otherwise. \end{cases} \tag{6}$$

Step 6 Repeat step 1

Sub cluster LEACH protocol based on Moth-flame optimization algorithm

Figure 4 provides the step by step description of Sub cluster LEACH protocol implementation of Mouth flame optimization approach. According to this, The base station made a query for data. After this, the network has to be sure whether it is for the first round or not. If yes and proceed further otherwise, there is requirement of evaluation energy of whole sensor nodes before selection CH nodes. If it is first round than all sensor have high and equal energy. Than If the algorithm proposed whether it is CH node or not. If the it is identified as CH node than MFO is come into picture. CH nodes is over burdened of aggregation of huge amount of data to reduce the burden Sub-CH nodes are selected from N-CH nodes which is successfully done with the help MFO approach. After selection procedure completed, Aggregation and transmission of data is completed by Sub-CH nodes and CH nodes. After this there is check of energy level of each node. If the energy level of every is higher than threshold energy than it is eligible for next round otherwise the node is declared to be died that means it cannot be considered as part of network anymore.

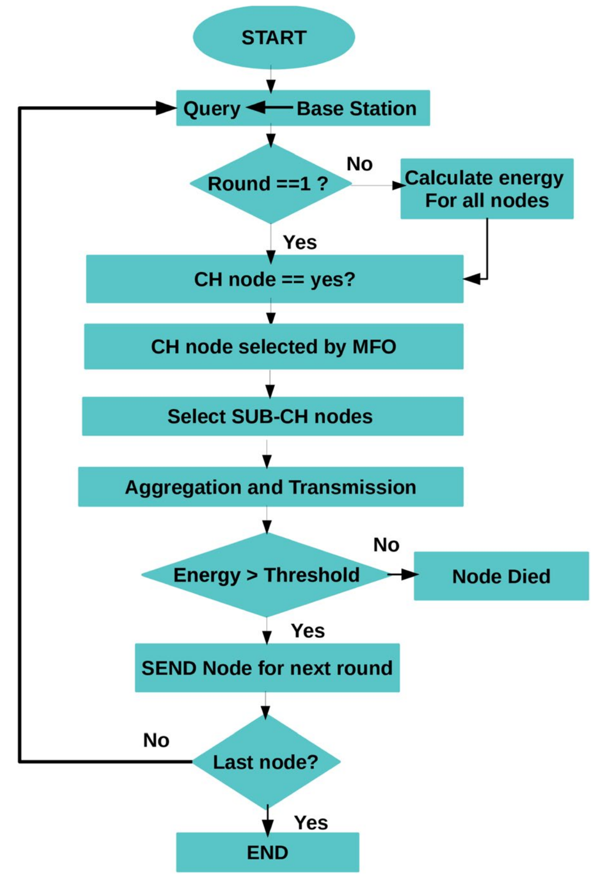

Fig. 4 Step by step representation of Sub-cluster LEACH-MFO protocol

Anomaly detection system

Anomaly detection system has capabilities to detect malicious attacks over energy efficient network [39]. Wireless sensor networks are low-cost network that is why most of people, industry prefer to install or deploy in the wide areas. But there are always fear to attack on the network that may interrupt the network [40, 41]. The intruders are keep their eye on the network which carry very important information to steal from it or down-grading the network by cyber attacking.The intruders are deploying the bug into the network that is also called malicious attack which reduce the efficiency of network and ultimately shutdown the network from working [42, 43]. To overcome this problem there is a need of anomaly detection system which should be robust in nature so that it will not degrade easily and also very secure in nature. For that anomaly detection system is proposed to provide a strong shield against the malicious attacks.

Figure 5 represents the step-wise conceptual block diagram of proposed anomaly detection system. In this system, the initial step is collect of data, here wireless sensor network is providing a data via base station to database. After collecting the data, it is prepossessed so that it can be encoded in a efficient way to secure it. After this features are extracted to embed into the machine learning algorithms (ML) [44]. Three ML algorithms are used to solve the complex problem i.e LR, SVM and KNN. All these are algorithms are very efficient in nature which can help to detect whether the test case is belongs to normal category or anomaly category. There are binary class distribution to detect the test case.

Machine learning techniques

Support vector machine approach

Support vector machine is also denoted as SVM, a algorithm to solve the complex problems. It practices the supervised learning. So, SVM algorithm considers target values also to evaluate the problem and results to be error-rate. It is very efficiently worked on classification problems [45]. Figure 6 represents the classification of two different classes and support vectors.There is a decision boundary with decides exact mutual separability between two classes.

Logistic regression approach

Logistic regression (LR) is generally used for solving complex problem. Mostly the problems are of supervised learning. LR supports a function which is used over binary prediction. this function is known as logistic regression [46]. There existed two types of variables i.e dependent variable and independent variable. Independent variable can be many in number but dependent variable are restricted to 2 [47]. Sigmoid function is used for evaluation as shown in Fig. 7.

K-nearest neighbor approach

K-nearest neighbor is also denoted as KNN. It is most popular algorithm to solve the complex problems. It can be used as classifier as well as regressor depending on the problem that need to evaluate. KNN lies under the category of supervised learning [48]. So, It tackle the dataset having label with it. It generally calculate the euclidean distance between two set of point. for instance we have two points such as m and n. and our aim is to identify the unknown observation m with respect to the identical output n that predict positively. It can be accomplished with the help of defining the function such as f: $m \rightarrow n$. KNN algorithm pseudo-code is represented in Algorithm 2.

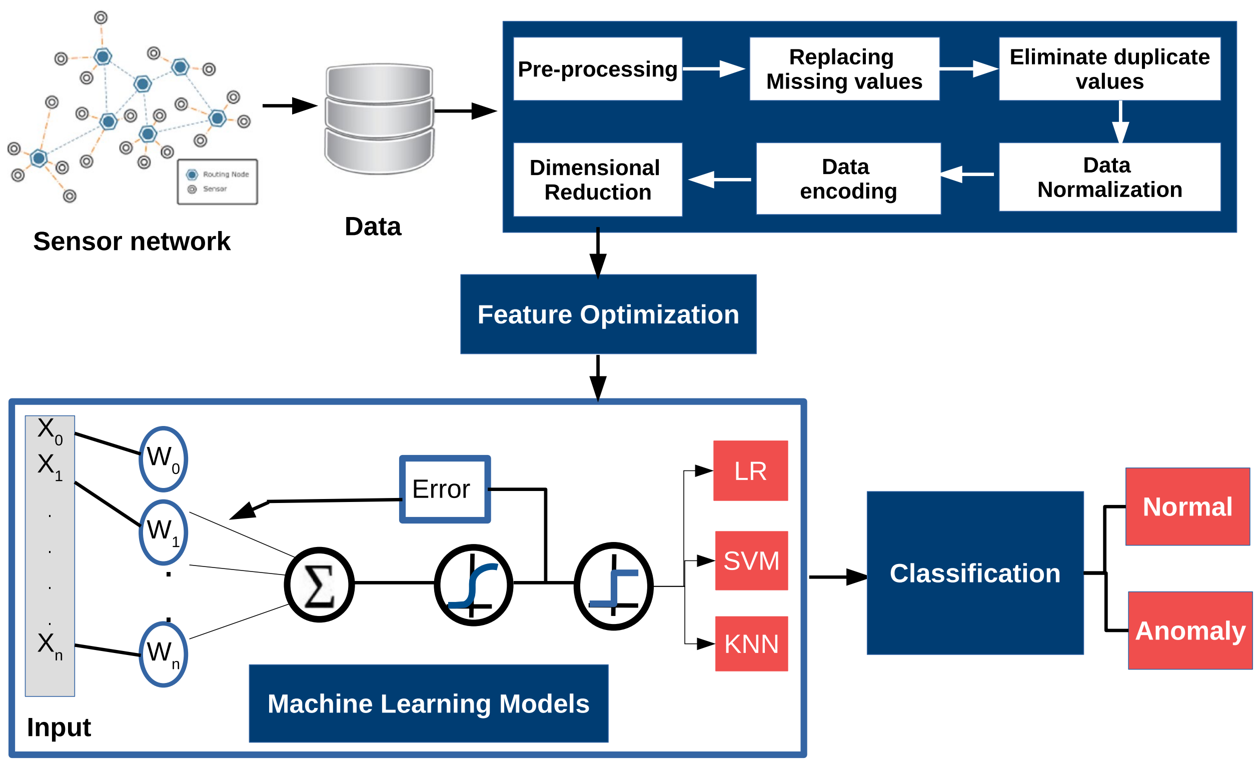

Fig. 5 Proposed anomaly detection model

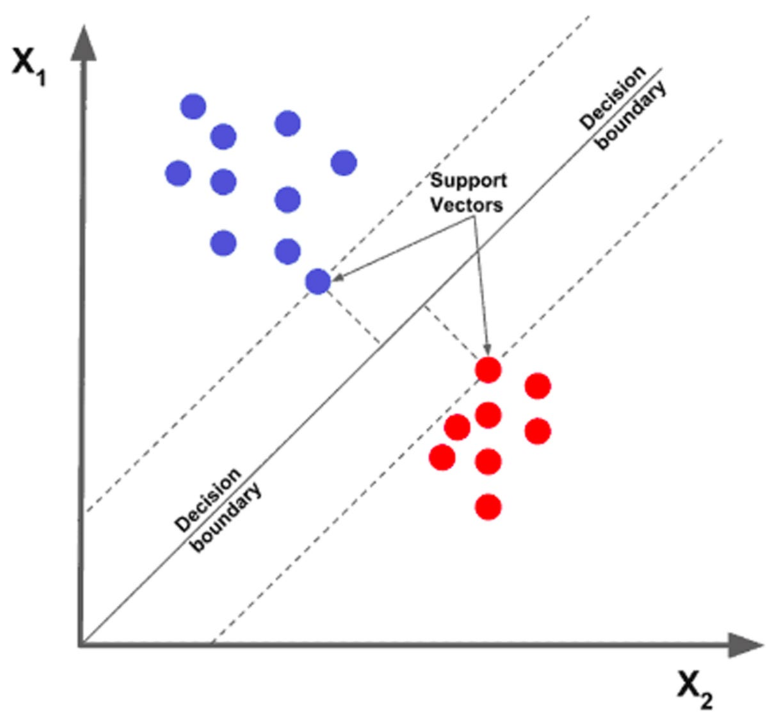

Fig. 6 Support vector machine model

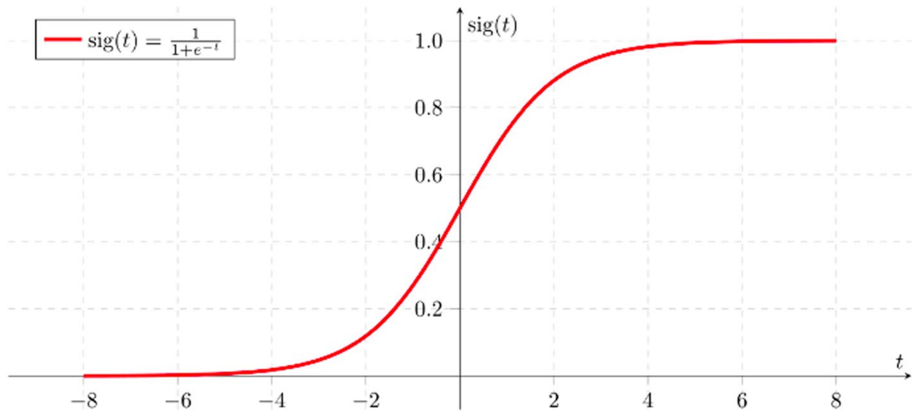

Fig. 7 Sigmoid Activation Function for logistic regression explaination

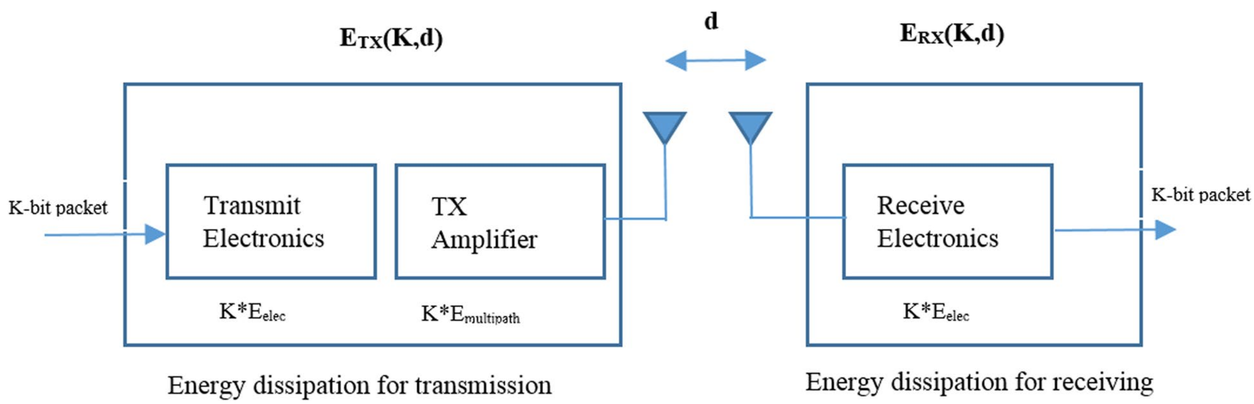

Fig. 8 Energy model

Algorithm 2 KNN Algorithm Pseudocode

1. Calculate distance as follows:

$$D(\alpha,\beta)=\sqrt{\sum_{i=1}^{j}(\beta_i-\alpha_i)^2}$$

Where D denotes Euclidean distance between two point i.e. α and β

2. SORT $\rightarrow$ β Euclidean distances.

3. SELECT $\rightarrow$ k distances from this sorted list. Where k>0

4. FIND $\rightarrow$ k number of points according to distance

5.Let k_i represents points related to i^{th} class; where $k \geq 0$

6.If $k_i > k_j, \forall$ i$\neq$ j then put m in class i.

Results and discussion

Experimental setup

The proposed system is implemented on MATLAB 2009 version. The calculation of energy is based on the energy transmission model which is represented in Fig. 8. The K-bit packet is induced with which transmit energy is calculated. It is amplified with the help of amplifier to transmit to other neighbouring node. The distance between two nodes is d. At the receiving end K-bit is received and also energy consumption is calculated. Hence in the both ends the energy dissipation is calculated.

For the simulations, 100 m × 100 m as well as 200 m × 200 m are considered for monitoring. In general, if 200m network size is considered with same network density as in 100m area, then, we need to increase the number of sensor nodes. However, we fix the number of sensor nodes for both network areas in our study to evaluate the protocol efficiency in different field areas. To see the dependency criteria on performance, we fix the total number of sensor nodes as 100. We have considered homogeneous sensor nodes only. Table 1 summarizes the other parameters used in the simulations. In this paper, we have chosen two protocols such LEACH and Sub cluster LEACH (Sub-LEACH). These approaches are quiet similar but are considered as energy efficient. Both approach are widely acceptable and have flexibly modified for specific environment. To enhance performance, we have embedded with Levenberg-Marquardt neural

Table 1 Initial parameter for network

Parameters	Values
Field Dimensions	100m X 100m/ 200m X 200m
Number of nodes	100
Base Station	50m X 50m/ 100m X 100m
Battery energy	0.5 Joules
Energy model parameter: $\epsilon_f s$	$10pJ/bit/m^2$
Energy model parameter: $\epsilon_m p$	$0.0013pJ/bit/m^4$
Electronics Energy: $E_E lec$	50 nJ/bit
Data packet length	4000 bits
Control packet length	200 bits

network (LMNN) and Moth-flame optimisation (MFO). To measure the performance various approaches such LEACH, LEACH-LMNN, LEACH-MFO, Sub-LEACH, Sub-LEACH-LMNN and Sub-LEACH-MFO, a comparative analysis has been done in terms network lifetime, end-to-end delay, throughput and pack delivery ratio (PDR).

Network lifetime

Figures 9 and 10 illustrates the network lifetime for the network size of 100m and 200m respectively. From both figures, we see that our proposed Sub-LEACH-LMNN as well as Sub-LEACH-MFO protocols perform better in terms of network lifetime compared to the other protocols.

The protocols are embedded with LMNN and MFO approaches which actually tried to reduce the network data traffic that ultimately reduce the network congestion by channelization data towards the specified sensor nodes or sub-cluster head nodes. The dependency of overall network lifetime is on energy consumption during execution of protocols and transmission of data. For this, the criteria of enhancement of time duration of network is done by selecting the next round CH node with highest remaining energy and nearer to base station in LEACH protocol with help of LMNN and MFO approaches. Similarly,in Sub-LEACH protocol, LMNN and MFO helps selection of nearest and highest remaining energy sensor node which further act sub-cluster head to compensate the consumption of battery life of cluster head node. For network lifetime, we considered last node dead (LND).The network lifetime is evaluated for 100m and 200m. The reduction in consumption of battery is notable in Sub-LEACH-LMNN and Sub-LEACH-MFO. This means that both approach i.e. LMNN and MFO work efficient in selection of cluster head selection and sub-cluster head selection. Also handles the network congestion at the grater extend from Cluster head nodes to base station. Hence overall lifetime is prolonged.

Analysis for delay

The delay analysis is evaluated by the following equation:

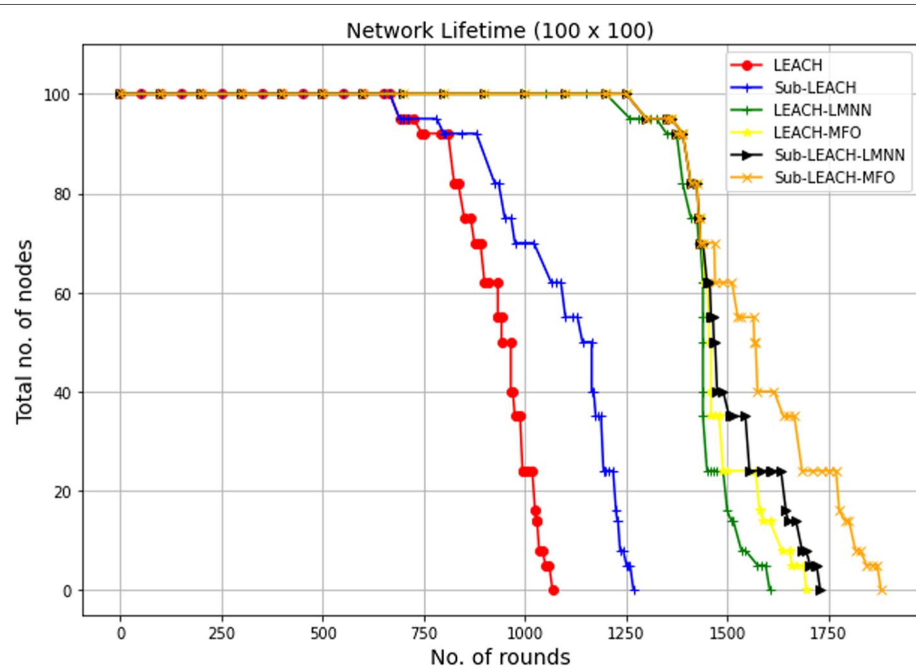

Fig. 9 Graphical representation of network lifetime for 100m

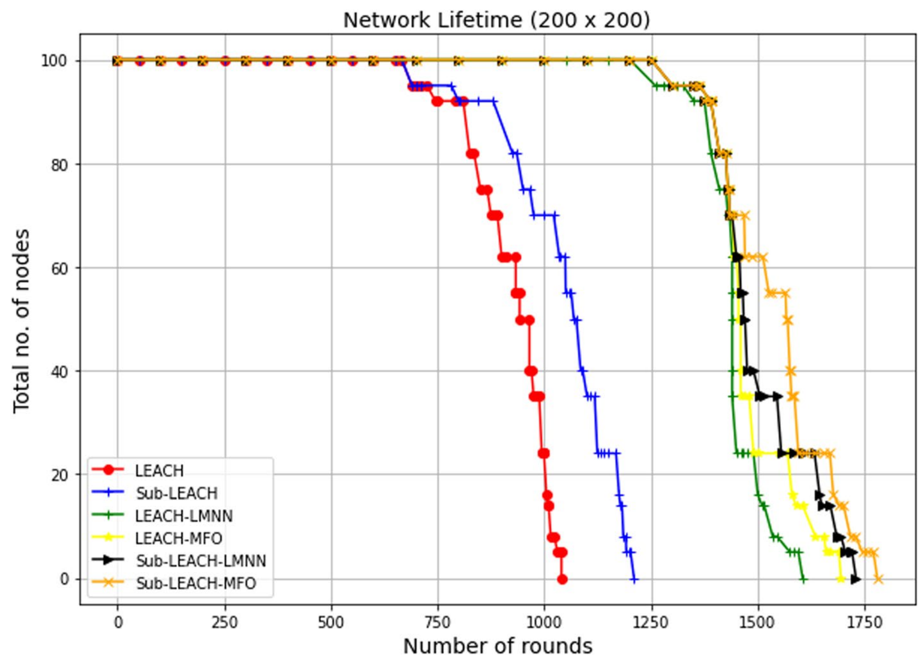

Fig. 10 Graphical representation of network lifetime for 200m

$$Delay_{E-to-E} = \frac{\sum(\theta - \alpha)}{N} \tag{1}$$

where θ denotes to arrive time, α denotes to send time and N denotes to number of connections.

From above equation, It can easily evaluate the delay. If any protocol will take lesser delay time that means it actually take less time for aggregation of data at CH node as well as forward data to BS. Therefore, the performance can be computed if any protocol boosted to speed up its aggregation process and transferring its aggregated data to BS. The end-to-end delay is evaluated for all six protocols. Delay is computed with each round for all protocols. From Figs. 11 and 12 depicts the delay analysis for all considered algorithms.It is clearly observed from the results that our proposed Sub-LEACH-MFO protocol outperforms the state-of-the-art in both network sizes. This signifies that Sub-LEACH-MFO performs efficiently in both fields and robust in nature. The end-to-end delay parameter helps any protocol to give better reliability. For this LMNN and MFO both provides good strategic approach for instant solution of selection CH node or Sub-Cluster head node. LMNN and MFO approaches help in proper categorization of nodes to get optimal solution of data transmission at minimum delay rate. Interestingly, in both cases the Sub-LEACH-MFO performs amazing for energy efficiency and delay in 100m as well as 200m.

Analysis of packet delivery ratio

Packet delivery ratio (PDR) is defined as ratio between number of packet received to number of packet sent. The performance of network is directly proportional to PDR. As PDR increases the performance of the network is also enhanced. The PDR can be calculated with the following equation:

$$PDR = \frac{\sum N_R}{N_S} \tag{2}$$

N_R denotes to number of packet receive where as N_S denotes to number of packet send.

From Figs. 13 and 14 represents the PDR of the proposed Sub-LEACH-LMNN and Sub-LEACH-MFO with other protocols. It is clear that Sub-LEACH-MFO has higher packet delivery ratio for both 100m and 200m. In $100\,m \times 100\,m$ PDR evaluation, the PDR is reduced for Sub-LEACH-MFO upto 97.18 % which is considerably high in comparison to Sub-LEACH-LMNN(96.88 %), LEACH-MFO (96.76 %), Sub-LEACH (96.62 %), LEACH-LMNN (95.89 %) and LEACH (95.54 %). Similarly, for $200\,m \times 200\,m$ PDR values in Sub-LEACH-MFO (97.22 %) is higher than Sub-LEACH-LMNN(96.91 %), LEACH-MFO (96.56 %), Sub-LEACH (96.11 %), LEACH-LMNN (95.77 %) and LEACH (95.44 %).

Throughput analysis

Throughput analysis is evaluated when each data packet is successfully received in a unit time. It is one

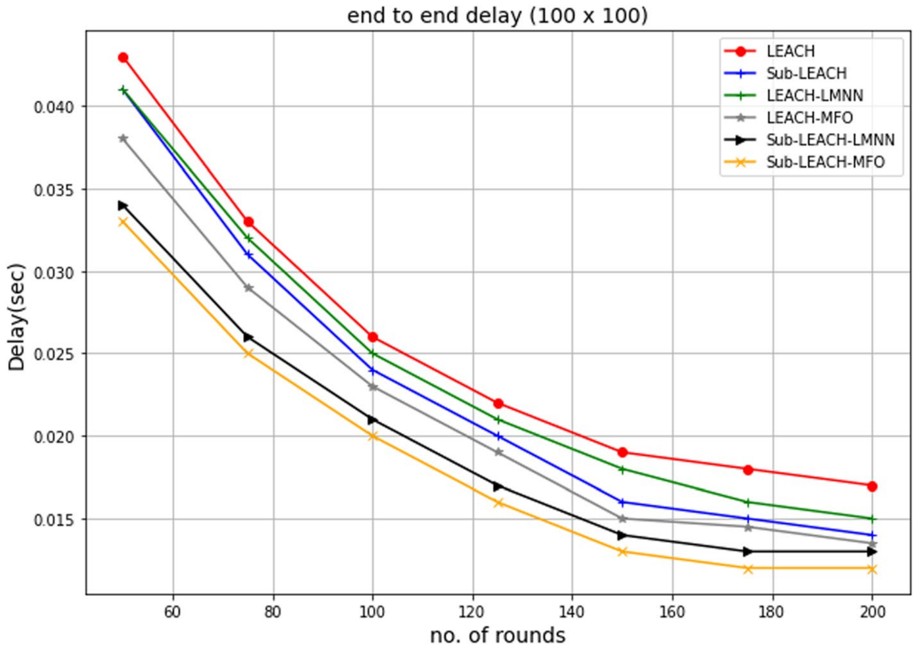

Fig. 11 Graphical representation of delay for 100m

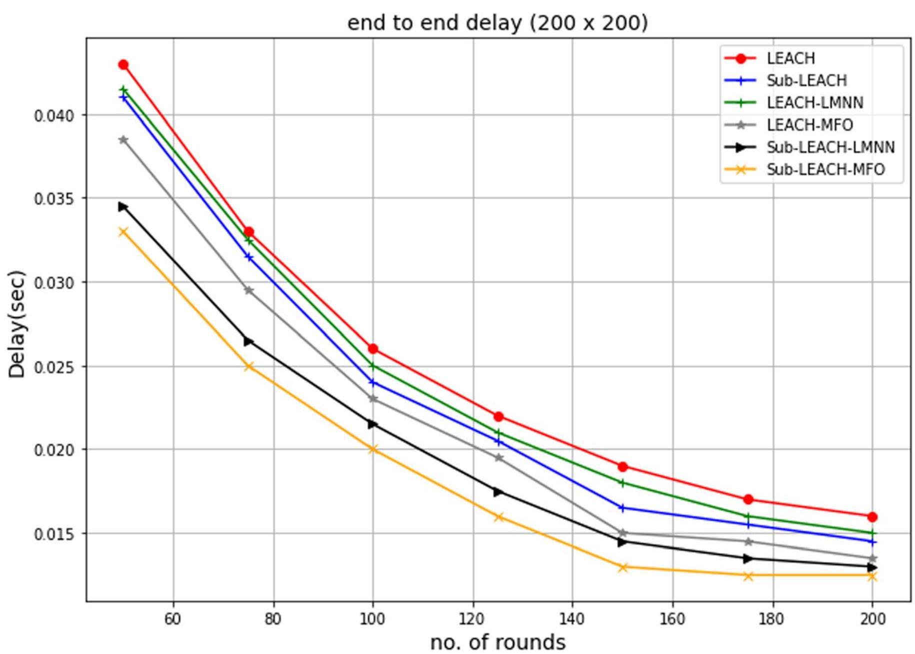

Fig. 12 Graphical representation of delay for 200m

of important parameter to measure the traffic analytic for any network. In network, maximum number of traffic per unit time is considered easily the more stable network will be. So, During steady stage in sensor network, throughput analysis place major role to measure the stability of the sensor network. With help of LMNN approach, size of query from sensor nodes are restricted and equalized the network traffic. From Figs. 15 and 16 shows that Sub-LEACH-MFO better results among other for both 100m and 200m areas.

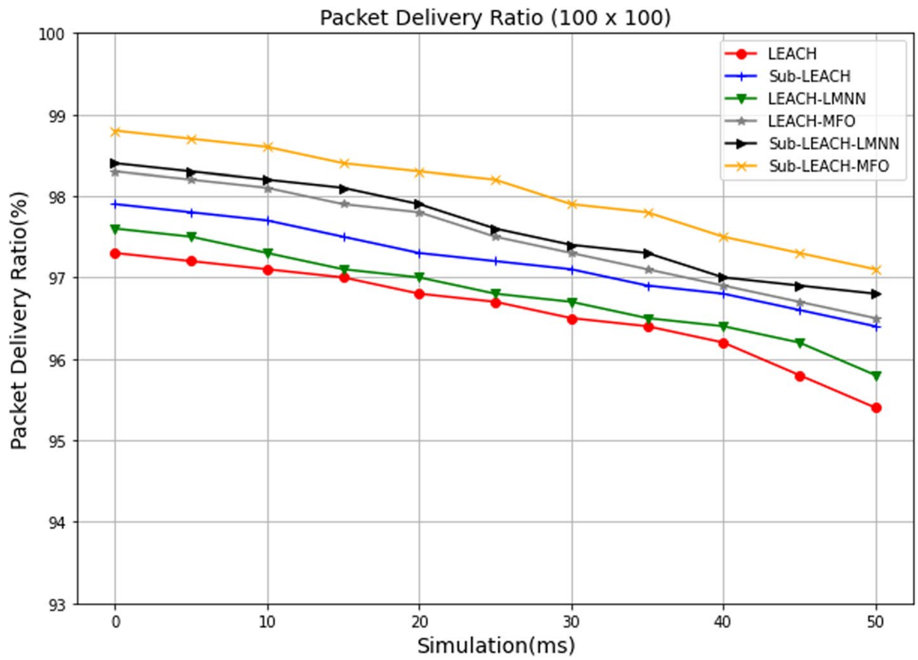

Fig. 13 Graphical representation of PDR for 100m

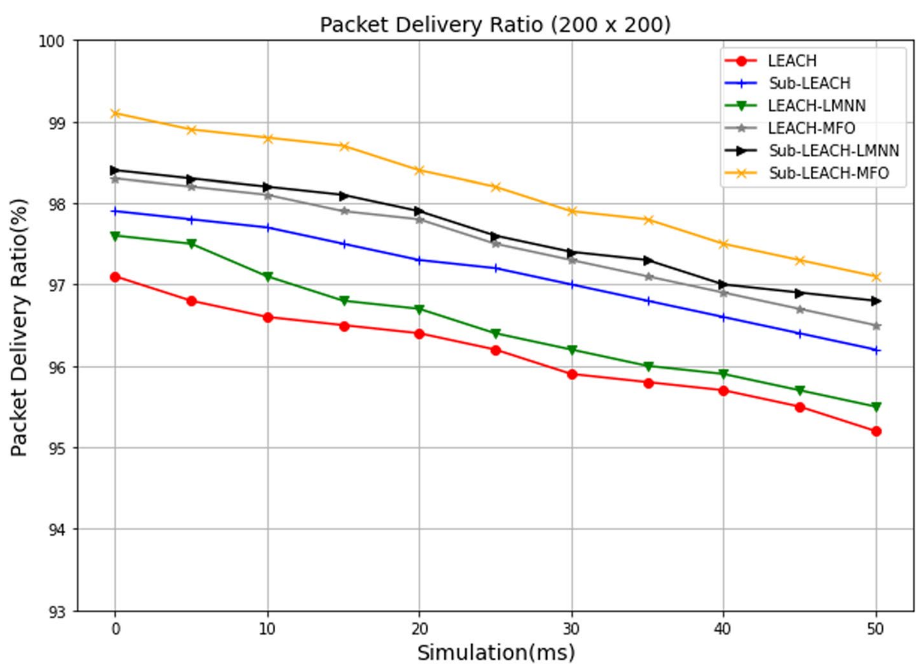

Fig. 14 Graphical representation of Packet PDR for 200m

In 100 m × 100 m throughput evaluation, the throughput is increased for Sub-LEACH-MFO upto 4.4 Mbps which is considerably high in comparison to Sub-LEACH-LMNN(4.2 Mbps), LEACH-MFO (4.1 Mbps), Sub-LEACH (4.0 Mbps), LEACH-LMNN (3.6 Mbps) and LEACH (3.3 Mbps). Similarly, for 200 m × 200 m throughput values in Sub-LEACH-MFO (4.5 Mbps) is higher than Sub-LEACH-LMNN(4.2 Mbps), LEACH-MFO (3.8 Mbps), Sub-LEACH (3.5 Mbps), LEACH-LMNN (3.2 Mbps) and LEACH (3.1 Mbps).

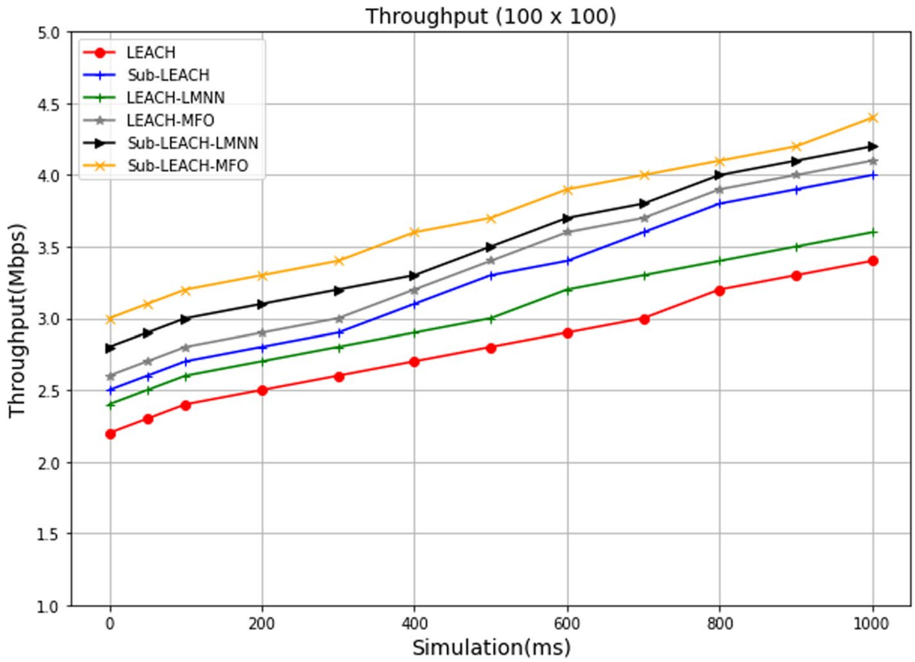

Fig. 15 Graphical representation of throughput analysis for 100m

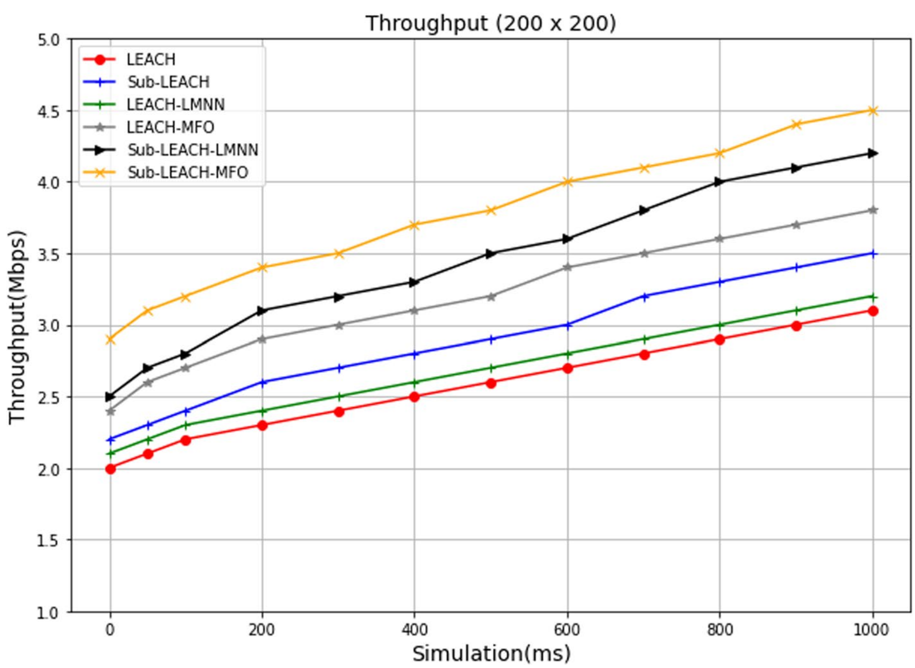

Fig. 16 Graphical representation of throughput analysis for 200m

Anomaly detection results

The anomaly detection for data collect is implemented with the help of three efficient machine learning algorithms i.e LR, SVM and KNN. The vision is to check whether the proposed system reliable or not, so the standard dataset is considered to perform the experiments. NSLKDD dataset is one of the standard and popular dataset to detect the anomaly detection. This dataset is easily available online. This dataset consists of total 148517 instances as well as 42 attributes. The normal packets are 77,054 and 71,463 are anomaly packets i.e. 148,517 total packets are available. For the training and testing, our system considered 80% of dataset for training and 20% for testing.

The proposed model is shown in Fig. 5. In which, after processing the data, system provides features which further embedded to machine learning three technique such as LR, SVM and KNN. These technique are implemented one by one and know as Proposed+LR, Proposed+SVM and Proposed+KNN . The training and testing loss for Proposed+LR is represented in Fig. 17 in which 100 epochs are considered and loss is drastically decrease from 1.6 to 0.152. Similarly, loss for KNN and SVM are represented in Figs. 18 and 19 respectively. The accuracy for the proposed anomaly detection system from all three proposed approaches are represented. Figure 20 (for Proposed+LR), Fig. 21 (for Proposed+SVM) and Fig. 22 (for Proposed+KNN) shows the accuracy for both training and testing. In comparison to the accuracy, for Proposed+SVM is the highest among others. Further Receiver operating characteristic i.e. ROC curve is evaluated for the proposed anomaly detection system. The ROC curve for Proposed+LR is represented in Fig. 23. Similarly, ROC curve for Proposed+SVM and Proposed+KNN are represented in Figs. 24 and 25 respectively. From ROC curve results, it is clear that Proposed+SVM is highest i.e 0.88 than other which are 0.81 and 0.72.

Anomaly detection discussion

The Proposed+SVM model evaluated and generates scores of 95.30% (training) and 95.81% (testing) as shown in Table 2. It is higher than Proposed+SVM and Proposed+KNN. For training and testing scores of Proposed+LR are 91.32% and 91.88% respectively and for Proposed+KNN are training (93.76%) and testing (93.21%). The values related to Proposed+SVM for various parameters for normal class are precision (94.00%), recall (98.00%) and F1-Score (96.00%). The values for anomaly class are precision (98.00%), recall (94.00%) and F1-score (96.00%).In case of Proposed+LR, for normal class the values are precision (91.00%), Recall (91.50%) and F1-score(92.00%) and on the other hand, the values of anomaly class are precision (91.65%), Recall (91.73%) and F1-score (92.11%). In last case i.e Proposed+KNN for normal class the values are precision (92.54%), Recall (92.89%) and F1-score (93.12%) and on the other hand, for anomaly class the values are precision (92.99%), Recall (93.11%) and F1-score (93.34%).

Table 3 represents the comparison with existing and proposed model in terms of TPR and FPR. For Proposed+SVM the values of FPR is 7.29 and TPR is 96.15% which is highest among all. In comparison to Proposed+LR which is FPR (6.19) and TPR (92.62%)

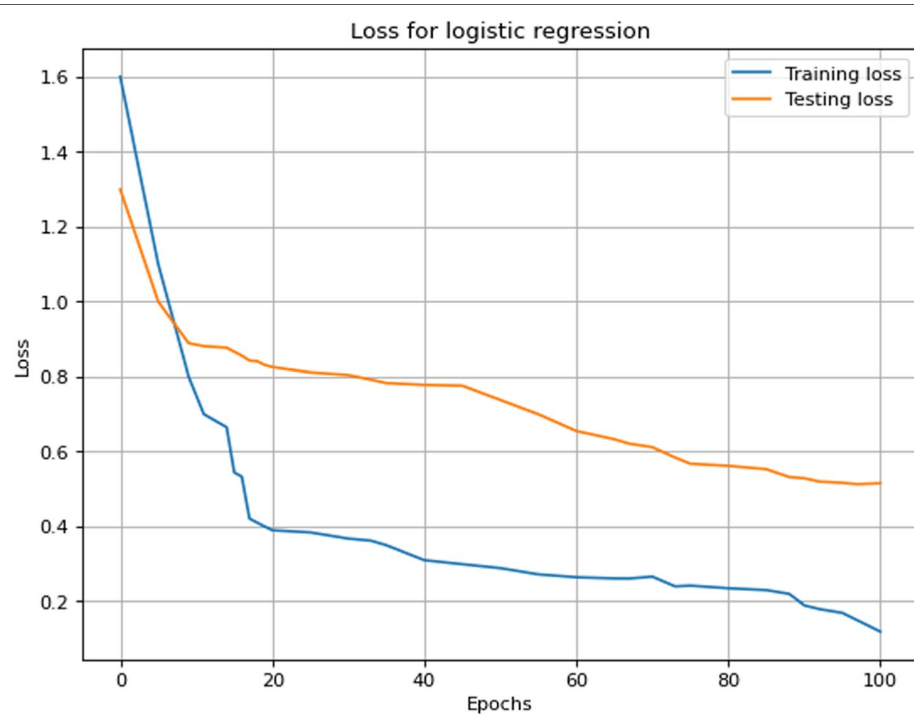

Fig. 17 Loss for our proposed model with logistic regression

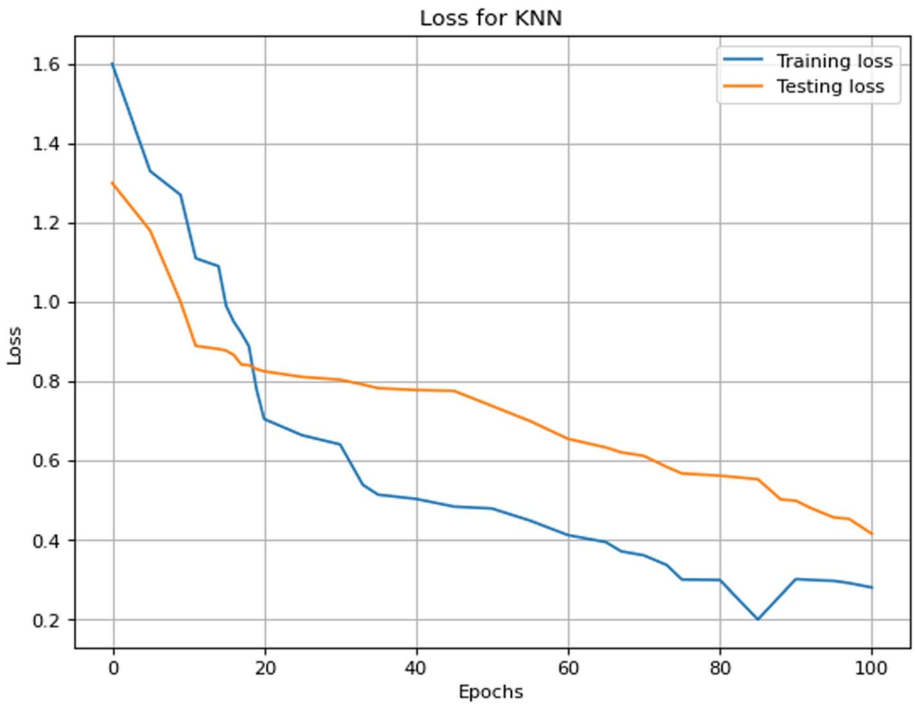

Fig. 18 Loss for our proposed model with KNN

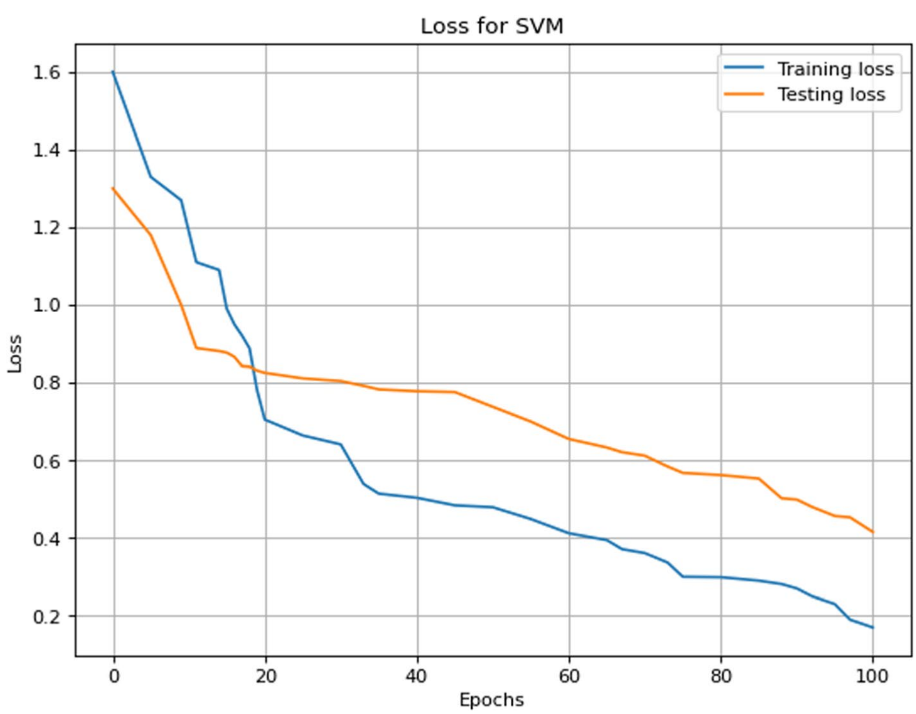

Fig. 19 Loss for our proposed model with SVM

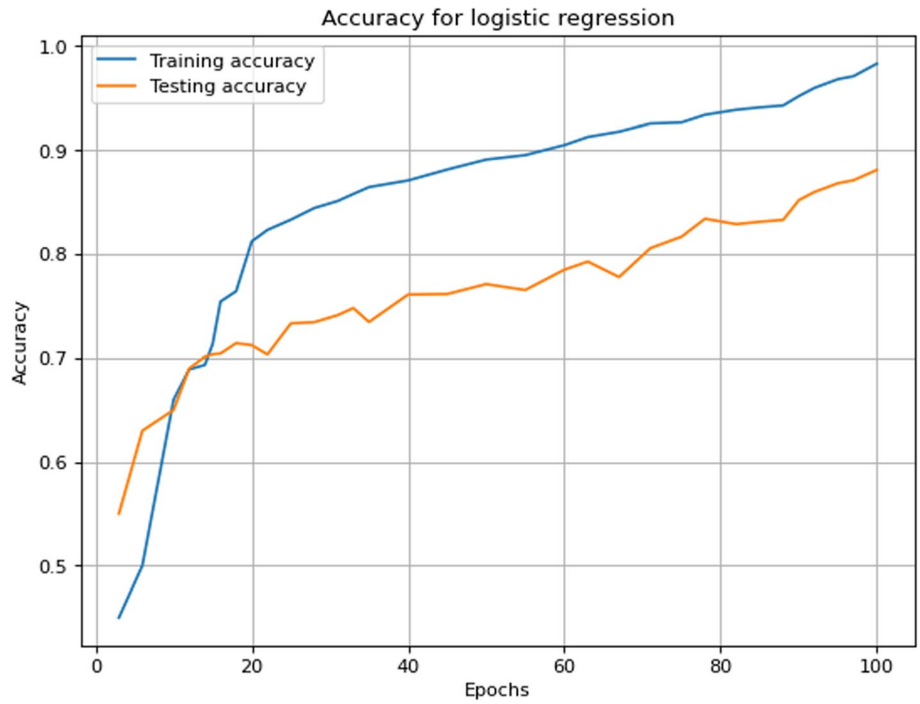

Fig. 20 Accuracy for our proposed model with logistic regression

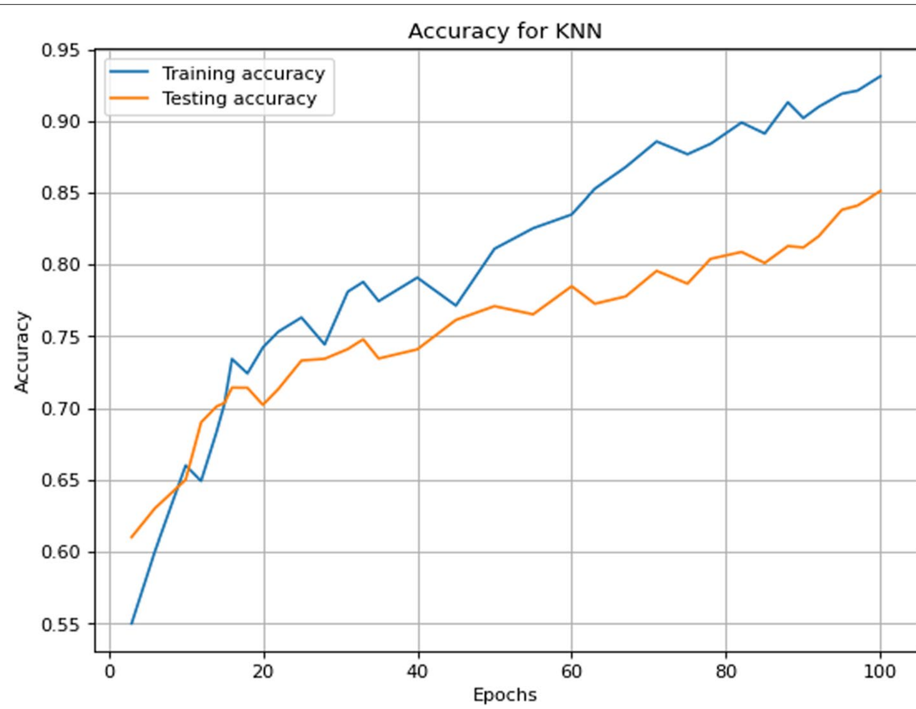

Fig. 21 Accuracy for our proposed model with SVM

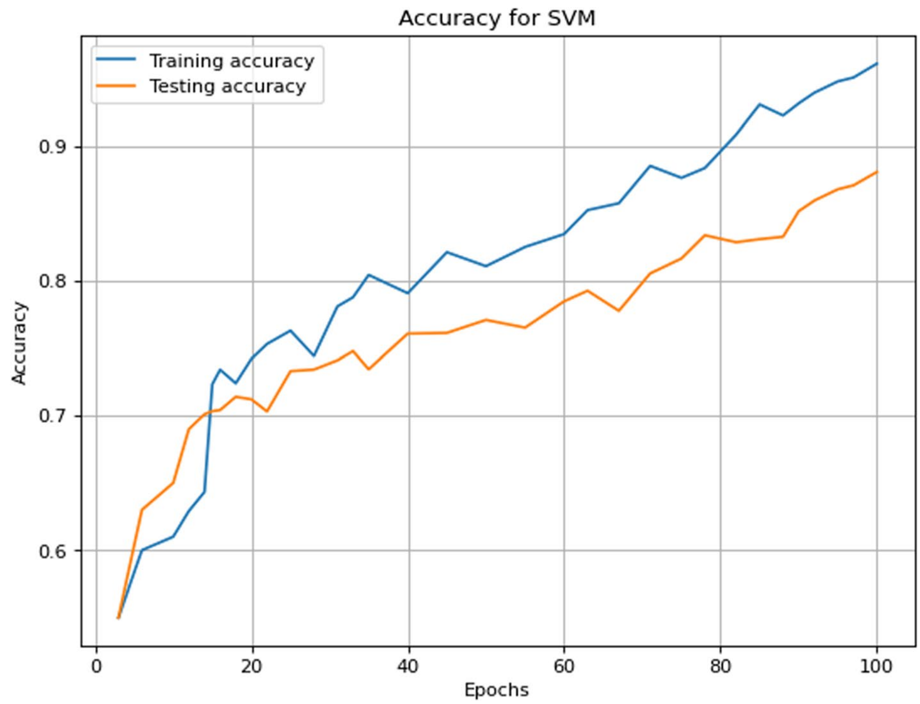

Fig. 22 Accuracy for our proposed model with KNN

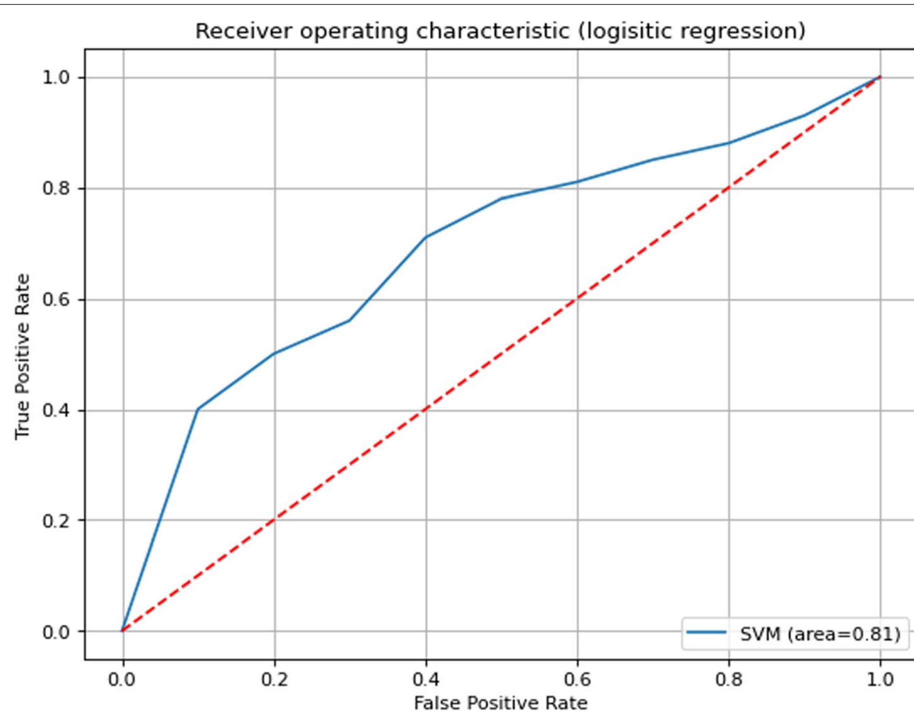

Fig. 23 ROC curve for our proposed model with logistic regression

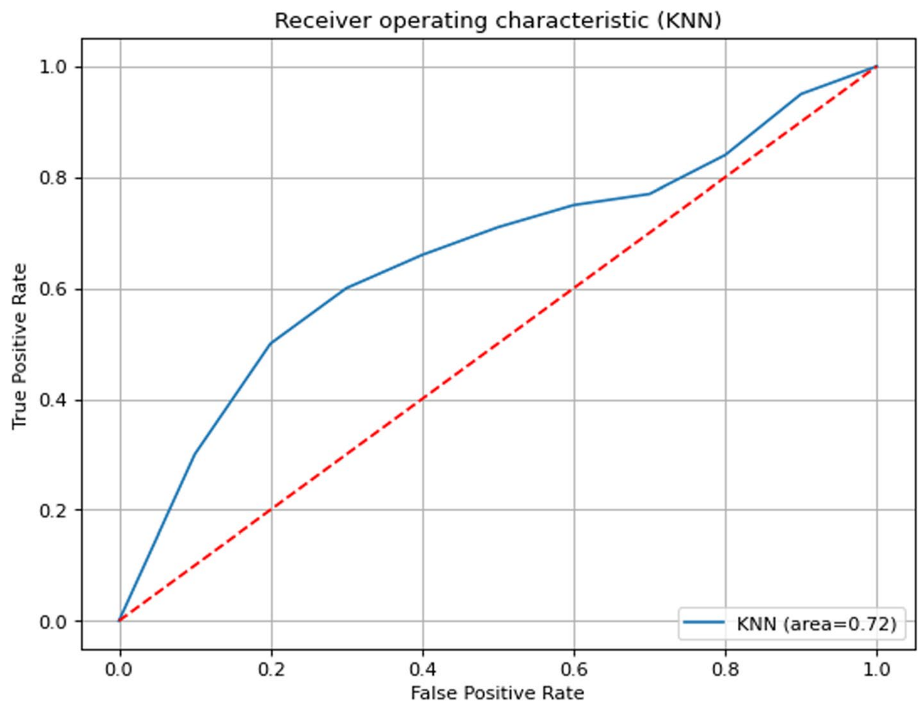

Fig. 24 ROC curve for our proposed model with SVM

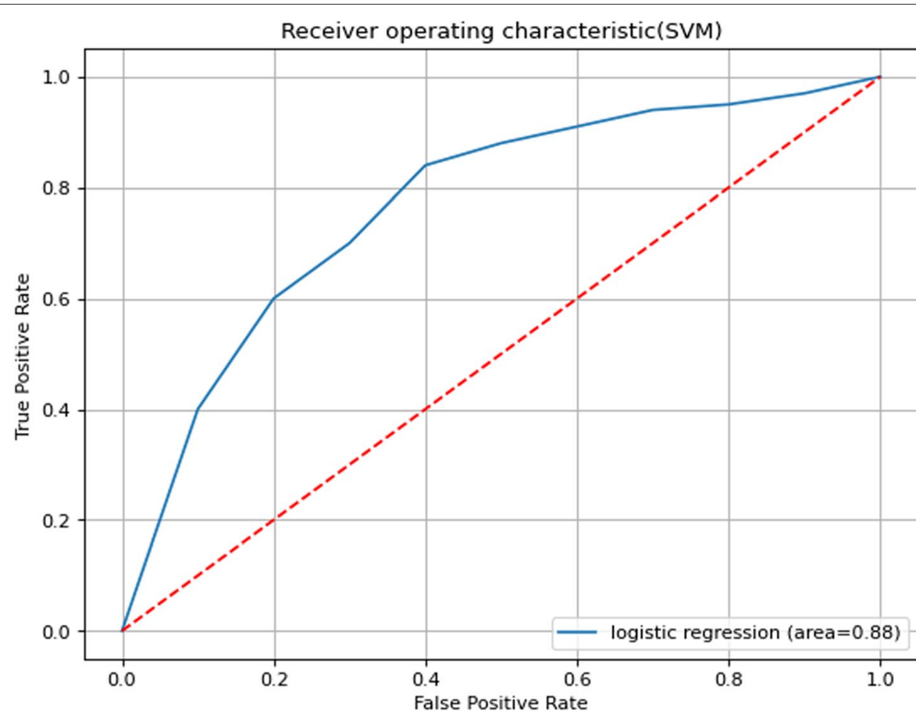

Fig. 25 ROC curve for our proposed model with KNN

Table 2 Normal and Anomaly classification experimental results

Algorithms		Train ACC%	Train Loss	Test ACC%	Test Loss
	Proposed+LR	91.32	0.16	91.88	0.21
	Proposed+KNN	93.76	0.13	93.21	0.15
	Proposed+SVM	95.30	0.11	95.81	0.10
Confusion Metric					
Algorithms		**TP**	**FP**	**FN**	**TN**
	Proposed+LR	19942	1022	563	22112
	Proposed+KNN	19989	1154	512	22367
	Proposed+SVM	20098	1354	443	22661
Classification Report					
Algorithms		**Class Labels**	**Precision**	**Recall**	**F1 Score**
	Proposed+LR	Normal Class	91.00	91.50	92.00
		Anomaly Class	91.65	91.73	92.11
	Proposed+KNN	Normal Class	92.54	92.89	93.12
		Anomaly Class	92.99	93.11	93.34
	Proposed+SVM	Normal Class	94.00	98.00	96.00
		Anomaly Class	98.00	94.00	96.00

Table 3 Comparison between our model and existing approaches

Models	TPR%	FPR%
DAR Ensemble [49]	78.88	N/A
Naive Bayes [50]	82.00	5.43
GAR Forest [51]	85.00	15.00
Proposed Model+LR	**92.62**	**6.19**
Proposed Model+KNN	**94.12**	**6.67**
Proposed Model+SVM	**96.15**	**7.29**

and for Proposed+KNN which is FPR (6.67) and TPR (94.12%). The comparison to other existing models TPR for DAR Ensemble (78.88%) , Naive Bayes (82.00%) and GAR Forest (85.00%) is lower than our Proposed+LR, Proposed+KNN and Proposed+SVM.

Conclusion

In this paper, we evaluate four QoS -based parameters, such as remaining energy, delay, throughput and PDR in WSNs. We considered two protocols such as LEACH and Sub-cluster LEACH (Sub-LEACH) protocols for same set of fields i.e. $100m \times 100m$ and $200m \times 200m$. To enhance the overall performance we have embedded two popular optimization approaches which are Levenberg-Marquardt neural network (LMNN) and Moth-flame optimisation (MFO). These both approaches are embedded to LEACH and Sub-cluster LEACH and we got LEACH-LMNN, LEACH-MFO, Sub-LEACH-LMNN and SUB-LEACH-MFO. These protocol are evaluated into two area of fields $100m \times 100m$ and $200m \times 200m$ one by one and got interesting results. Sub-LEACH-MFO is performed best as compared to remaining algorithms in terms of energy efficiency, end-to-end delay, throughput and PDR as shown in Figs. 9 to 16. Further, in this paper anomaly detection system with three ML protocols named as Proposed+LR, Proposed+KNN and Proposed+SVM. Proposed+SVM is performed well among other as shown in Table 2. Proposed+SVM is also compared to existing systems and shows outstanding results in terms of TPR and FPR represented in Table 3. In future, we will try to consider more parameters and huge area of field to analyse the results. We will also focus on the more optimiser approaches to get better results.

Authors' contributions

All authors read and approved the final manuscript.

Funding

Authors would like to acknowledge contribution to this research from the Rector of the Silesian University of Technology, Gliwice, Poland under grant No. 09/010/RGJ22/0068.

Availability of data and materials

The data presented in this study are available on request from the corresponding author. Some data are publicly available and some are not.

Declarations

Ethics approval and consent to participate

This research does not consider any of human and/ or animal studies which may cause any ethical concern, thus the research does not need to recieve ethical approvals.

Competing interests

The author declares that they have no competing interests.

Author details

[1]Scool Team, Centre de Recherche en Informatique, Signal et Automatique de Lille (INRIA), Villeneuve-d'Ascq 59655, Lille, France. [2]Faculty of Applied Mathematics, Silesian University of Technology, Gliwice 44100, Poland. [3]Department of Computer Science and Engineering, Thapar Institute of Engineering And Technology, Patiala 147004, Punjab, India. [4]Department of Computer Science, Middlesex University, London NW4 4BT, The Burroughs, United Kingdom.

Received: 12 September 2022 Accepted: 8 October 2022
Published online: 27 October 2022

References

1. Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E (2002) Wireless sensor networks: a survey. Comput Netw 38:393–422
2. Yick J, Mukherjee B, Ghosal D (2008) Wireless sensor network survey. Comput Netw 52:2292–2330
3. Awad A, German R, Dressler F (2011) Exploiting Virtual Coordinates for Improved Routing Performance in Sensor Networks. IEEE Trans Mob Comput 10(9):1214–1226
4. Ahmad M, Li T, Khan Z, Khurshid F, Ahmad M (2018) A Novel Connectivity-Based LEACH-MEEC Routing Protocol for Mobile Wireless Sensor Network. Sensors 18(12):1–20
5. Intanagonwiwat C, Govindan R, Estrin D (2007) Direct Diffusion: a Scalable and Robust Communication Paradigm for Sensor Networks. Proc ACM (Mobi-Com) 56–67

6. Engmann F, Katsriku FA, Abdulai JD, Adu-Manu KS, Banaseka FK (2018) Prolonging the lifetime of wireless sensor networks: a review of current techniques. Wirel Commun Mob Comput 2018:1–24
7. Guo J, Jafarkhani H (2019) Movement-Efficient Sensor Deployment in Wireless Sensor Networks With Limited Communication Range. IEEE Trans Wirel Commun 18(7):3469–3484
8. Liu X, Wu J (2019) A method for energy balance and data transmission optimal routing in wireless sensor networks. Sensors 13:3017
9. Sharma H, Haque A, Jaffery ZA (2019) Maximization of wireless sensor network lifetime using solar energy harvesting for smart agriculture monitoring. Ad Hoc Netw 94:1–14
10. Nguyen PL, Nguyen TH, Nguyen K (2020) A Path-Length Efficient, Low-Overhead, Load-Balanced Routing Protocol for Maximum Network Lifetime in Wireless Sensor Networks with Holes. Sensors 1-32
11. Guo W, Yan C, Lu T (2019) Optimizing the lifetime of wireless sensor networks via reinforcement-learning-based routing. Int J Distrib Sens Netw 15:1–20
12. Oh H, Chae K (2007) An Energy-Efficient Sensor Routing with low latency, scalability in Wireless Sensor Networks. Int Conf Multimed Ubiquit Eng 147–152
13. Mittal M, Kumar K (2014) Network Lifetime Enhancement of Homogeneous Sensor Network Using ART1 Neural Network. Sixth Int Conf Comput Intell Commun Netw 472–475
14. Fausett L (1994) Fundamentals of neural networks: architecture. algorithm and applications. 1st edn. Pearson, p 480
15. Mittal M, Kumar K (2015) Delay Prediction in Wireless Sensor Network Routing Using ART1 Neural Network. Afr J Comput ICT 8:175–180
16. Enami N, Moghadam RA (2010) Energy Based Clustering Self Organizing Map Protocol For extending Wireless Sensor Networks lifetime and coverage. Can J Multimed Wirel Netw 1:42–54
17. Mittal M, Kumar K (2015) Quality of services provisioning in wireless sensor networks using artificial neural network: a survey. Int J Comput Appl 117:28–40
18. Guo Z, Malakooti S, Sheikh S, Najjar CA, Lehman M, Malakooti B (2017) Energy aware proactive optimized link state routing in mobile ad-hoc networks. Appl Math Model 35:4715–4729
19. Robinson YH, Julie EG, Balaji S, Ayyasamy A (2017) Energy Aware Clustering Scheme in Wireless Sensor Network Using Neuro-Fuzzy Approach. Wirel Personal Commun 95:703–721
20. Akkaya K, Younis M (2005) Energy and QoS Aware Routing in Wireless Sensor Networks. Clust Comput 8:179–188
21. Mohemed RE, Saleh AI, Abdelrazzak M, Samra AS (2017) Energy-efficient routing protocols for solving energy hole problem in wireless sensor networks. Comput Netw 114:51–66
22. Aslam M, Munir EU, Rafique MM, Hu X (2016) Adaptive energy-efficient clustering path planning routing protocols for heterogeneous wireless sensor networks. Sustain Comput Inf Syst 12:57–71
23. Heinzelman WB, Chandrakasan A, Balakrishnan H (2000) Energy- efficient communication protocol for wireless microsensor networks. Int Conf Syst Sci 2:1–10
24. Heinzelman WB, Chandrakasan A, Balakrishnan H (2002) Application-specific protocol architecture for wireless microsensor networks. IEEE Trans Wirel Commun 1:660–670
25. Arbab E, Aghazarian V, Hedayati A, Motlagh NG (2012) LEACH-Based Clustering Algorithm for Optimizing Energy Consumption in Wireless Sensor Networks. In: 2nd International Conference on Computer Science and Information Technology (ICCSIT). pp 147-150
26. Feng YF, Pan SG, Huang ZY, Lin HC (2019) Improvement of energy efficiency in wireless sensor networks using Low-energy Adaptive Clustering Hierarchy (LEACH)-based energy betweenness ModelSens. Mater 31:2691–2702
27. Xu J, Jin N, Lou X, Peng T, Zhou Q, Chen Y (2012) Improvement of Leach protocol for WSN. In 2012 9th International conference on fuzzy systems and knowledge discovery. IEEE Conference, Chongqing, p 2174–2177
28. Anand G, Balakrishnan R (2013) Leach-Ex protocol - A comparative performance study and analysis with leach variants of wireless sensor networks. IEEE Conference, Malaysia, p 196–196
29. Liu JL, Ravishankar CV (2011) LEACH-GA: Genetic Algorithm-based energy efficient adaptive clustering protocol for Wireless Sensor Networks. Int J Mach Learn Comput 79–85
30. Mittal M, Srinivasan S, Rani M, Vyas OP (2017) Type-2 fuzzy ontology-based multi-agents system for wireless sensor network. In TENCON 2017 - 2017 IEEE Region 10 Conference. pp 2864–2869
31. Salim A, Osamy W, Khedr AM (2014) IBLEACH: Intra-balanced Leach protocol for Wireless Sensor Networks. Wirel Netw 20:1515–1525
32. Mittal M, Iwendi C, Khan S, Rehman Javed A (2020) Analysis of security and energy efficiency for shortest route discovery in low-energy adaptive clustering hierarchy protocol using Levenberg-Marquardt neural network and gated recurrent unit for intrusion detection system. Trans Emerging Tel Tech 1:e3997
33. Lindsey S, Raghavendra C (2002) PEGASIS: Power Efficient Gathering in Sensor Information Systems. IEEE Aerosp Conf 1125–1130
34. Gupta S, Kumar P, Tekchandani RK (2022) Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models. Multimed Tools Appl. https://doi.org/10.1007/s11042-022-13558-9
35. Gupta S, Kumar P (2021) Attention Recognition System in Online Learning Platform Using EEG Signals.Emerging Technologies for Smart Cities. Lecture Notes in Electrical Engineering, vol 765. Springer, Singapore. https://doi.org/10.1007/978-981-16-1550-4_15
36. Sharma A, Gupta S, Kaur S, Kumar P (2019) Smart Learning System Based on EEG Signals. ICACDS 2019. Communications in Computer and Information Science, vol 1046. Springer, Singapore. https://doi.org/10.1007/978-981-13-9942-8_44
37. Gupta S, Gouttam D (2017) Towards changing the paradigm of software development in software industries: An emergence of agile software development. In 2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM). pp 18-21. https://doi.org/10.1109/ICSTM.2017.8089120
38. Sharma S, Gupta S (2015) A Correction Model for Real-word Errors. In Proceedings of the 4th International Conference on Eco-friendly Computing and Communication Systems, vol.70. pp 99–106. https://doi.org/10.1016/j.procs.2015.10.047
39. Iwendi C, Khan S, Anajemba JH, Mittal M, Alenezi M, Alazab M (2020) The use of ensemble models for multiple class and binary class classification for improving intrusion detection systems. Sensors 20:2559
40. Mittal M, Saraswat LK, Iwendi C, Anajemba JH. (2019) A neuro-fuzzy approach for intrusion detection in energy efficient sensor routing. In Proceedings of the 4th International conference on internet of things: Smart innovation and usages (IoT-SIU). pp 1–5. https://doi.org/10.1109/IoT-SIU.2019.8777501
41. Anajemba JH, Iwendi C, Mittal M, Yue T (2020) Improved advance encryption standard with a privacy database structure for IoT nodes. In Proceedings of the 9th international conference on communication systems and network technologies (CSNT). IEEE Conference, Gwalior, p 201–206
42. Mittal M, de Prado RP, Kawai Y, Nakajima S, Muñoz-Expósito JE (2021) Machine learning techniques for energy efficiency and anomaly detection in hybrid wireless sensor networks. Energies 14:3125
43. Javed AR, Sarwar MU, Khan S, Iwendi C, Mittal M, Kumar N (2020) Analyzing the effectiveness and contribution of each axis of tri-axial accelerometer sensor for accurate activity recognition. Sensors 20:1–20
44. Chauhan S, Mittal M, Woźniak M, Gupta S, Pérez de Prado R (2021) A technology acceptance model-based analytics for online mobile games using machine learning techniques. Symmetry 13:1545
45. Payal A, Rai CS, Reddy BVR (2013) Comparative analysis of bayesian regularization and levenberg-marquardt training algorithm for localization in wireless sensor network. In Proceedings of the 15th International Conference on Advanced Communications Technology (ICACT). pp 191–194
46. Kaur N, Singh S (2017) Optimized cost effective and energy efficient routing protocol for wireless body area networks. Ad Hoc Netw 61:65–84
47. Haseeb K, Abbas N, Saleem MQ, Sheta OE, Awanand K, Islam N (2019) RCER: Reliable Cluster-based Energy-aware Routing protocol for heterogeneous Wireless Sensor Networks. PLoS ONE 14:9
48. Du YC, Stephanus A (2018) Levenberg-Marquardt neural network algorithm for degree of arteriovenous fistula stenosis classification using a dual optical photoplethysmography sensor. Sensors 7:1–18
49. Gaikwad D, Thool R (2016) DAREnsemble: Decision tree and rule learner based ensemble for network intrusion detection system. Proc First Int Conf Inf Commun Technol Intell Syst 50:185–193

50. Pajouh HH, Dastghaibyfard G, Hashemi S (2017) Two-tier network anomaly detection model: A machine learning approach. J Intell Inf Syst 48:61–74
51. Kanakarajan N, Muniasamy K (2016) Improving the accuracy of intrusion detection using gar-forest with feature selection. Proceedings of the 4th International Conference on Frontiers in Intelligent Computing: Theory and Applications. Springer, New Delhi, pp 539–547

Publisher's Note