International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6445
***Article Received**: 08 February 2023 **Revised**: 04 March 2023 **Accepted**: 22 March 2023*

Optimization of Energy-Efficient Cluster Head Selection Algorithm for Internet of Things in Wireless Sensor Networks

Rella Usha Rani[1], P. Sankara Rao[2], Kothapalli Lavanaya[3], Nimmala Satyanarayana[4], Sudula Lallitha[5], Phani Prasad J[6]

[1]Professor, CSE Department, CVR College of Engineering, Hyderabad, India
teaching.usha@gmail.com[1]
[2]Asst. Professor, CSE Department, GITAM School of Technology, Visakhapatnam, India
dr.sankararaop@gmail.com[2]
[3]Asst. Professor, IT Department, Sridevi Woman Engineering College, Hyderabad, India
teachinglavanyak@gmail.com[3]
[4]Professor, CSE Department, CVR College of Engineering and Technology, Hyderabad, India
satyauce234@gmail.com [4]
[5]Asst. Professor, IT Department, Sridevi Woman Engineering College, Hyderabad, India
shudulalalitha@gmail.com[5]
[6]Asst Professor, CSE Department, CVR College of Engineering, Hyderabad, India
phanimtechcse@gmail.com[5]

Abstract: The Internet of Things (IoT) now uses the Wireless Sensor Network (WSN) as a platform to sense and communicate data. The increase in the number of embedded and interconnected devices on the Internet has resulted in a need for software solutions to manage them proficiently in an elegant and scalable manner. Also, these devices can generate massive amounts of data, resulting in a classic Big Data problem that must be stored and processed. Large volumes of information have to be produced by using IoT applications, thus raising two major issues in big data analytics. To ensure an efficient form of mining of both spatial and temporal data, a sensed sample has to be collected. So for this work, a new strategy to remove redundancy has been proposed. This classifies all forms of collected data to be either relevant or irrelevant in choosing suitable information even before they are forwarded to the base station or the cluster head. A Low-Energy Adaptive Clustering Hierarchy (LEACH) is a cluster-based routing protocol that uses cluster formation. The LEACH chooses one head from the network sensor nodes, such as the Cluster Head (CH), to rotate the role to a new distributed energy load. The CHs were chosen randomly with the possibility of all CHs being concentrated in one locality. The primary idea behind such dynamic clustering was them resulted in more overheads due to changes in the CH and advertisements. Therefore, the LEACH was not suitable for large networks. Here, Particle Swarm Optimization (PSO) and River Formation Dynamics are used to optimize the CH selection (RFD). The results proved that the proposed method to have performed better compared to other methods.

Keywords: Internet of Things (IoT), WSN, Big Data, Clustering, LEACH, Particle Swarm Optimization (PSO), and River Formation Dynamics (RFD).

I. INTRODUCTION

IoT is an important topic that must be researched because it allows sensors in cars and other vehicles to communicate with one another without the need for human intervention. This is defined in the form of a network connection in which the sensors use different applications. In the case of remote health which has wearable body sensors for monitoring patients at home without continuous visits to the hospital, it ensures students can load their books by using their mobile. During the time the IoT related to the WSN, it was taken as communication that was network-to-network as several sensors will be able to communicate using the Internet for other applications such as health. This is for a smart home. When the person goes to work and forgets something like turning off the stove, remote control can be done through the Internet to turn it off without returning home [1].

WSNs include a large number of sensor nodes that monitor and record all physical conditions in the environment using sensor data collected by a sink node. The WSNs were employed for measuring various environmental conditions such as sound, pollution, wind, humidity, and temperature. But the limited capacity of one node, along with a narrow wireless link (in comparison with other typical networks), can result in problems while delivering sensor data to their sink node. However, a good system of data aggregation can be beneficial to various big data systems. Thus, a need to analyze these studies linking the WSNs to the Big Data systems, thereby overcoming its deficiencies, was felt [2].

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6445
***Article Received**: 08 February 2023 **Revised**: 04 March 2023 **Accepted**: 22 March 2023*

These advancements in the field of IoT have been generating plenty of data, referred to as Big Data. Based on a report that was published in the year 2012 by IBM, the world data had been generated only in the last two years. The result of this was Big Data that has risen widely in the form of a new trend. It is applied to several areas like querying, mining, processing, distributing, and modeling. Big Data consists of three Vs. which are Volume, Velocity, and Variety, and these can be a major challenge in organizing as they are difficult to obtain, store, analyze, or process with the currently used technology. Volume refers to plenty of data that needs to be aggregated to process and further analyze velocity indicating processing and analysis of high speed, such as health data, social websites, remote sensing, and online streaming. At the same time, Variety indicates the varied structures present, such as the WSN, IoTs, and Machine-to-Machine. Additionally, currently, existing services such as routers, network switches, and social websites can generate big data of a large volume [3]. It is also anticipated that most of such data will be generated using different sensors such as Infrared Sensors, Temperature Sensors, and Ultrasonic Sensors.

The IoT, as defined broadly, is similar to a brain storing real-world data (in databases or cloud services) and may be used for monitoring other real-world parameters. This will result in the meaningful interpretation of decisions for sensed data. The IoT is therefore responsible for decision-making, manipulation, and data processing. The WSN can be termed as the ears and eyes of the IoT. It is like a bridge connecting the real world and the digital world, passing values to the Internet [4]. Extracting any useful information from large amounts of data will need higher levels of processing and computation that are executed at the level of the sensor nodes. These are nothing but battery-driven devices that have very limited power. Therefore, the WSN can also have certain other forms of limitations, such as power and capacity to compute, and these will have to be optimized. At the same time, the IoT will be used for connecting a very large number of devices employed to process metadata. This can further result in using the available power that can affect the network and its lifetime simultaneously. To ensure the WSN network lifetime is maximized, the routing paths for data packets are chosen in a way that the consumed energy for the total path is reduced.

The process of clustering sensor nodes can be listed as a very popular approach to collecting data in an energy-efficient manner. For this approach, the sensor nodes will be divided into clusters, with each cluster containing a coordinator node known as the Cluster Head (CH). Cluster Members are the other nodes in the cluster (CMs). Every CH will be assigned the task of collecting data from the cluster's CMs. Once sensed data from the CMs have been collected, each CH will aggregate it and then transmit it to the sink node via multi-hop communication with other CMs or directly [5]. The selection of an optimal number of clustering nodes, such as CHs distributed uniformly over the area of interest, can be a significant challenge in this process. This is an NP-hard problem since there was a C_m^n possible selection of *m* cluster heads from the *n* sensor nodes. Large-scale WSNs will have non-polynomial computational complexity for the problem. Due to the brute-force approach's inefficiency in problem resolution, researchers have sought to offer better solutions in the published literature via the proposal of other computing paradigms which are nature-inspired as well as based on meta-heuristic algorithms [6].

For these homogeneous WSNs, the capacity of transmission will be similar for every node. Due to the sensor node's limited energy budget, the aggregated data's direct transmission was made from the CHs to their sink. However, for a large-scale WSN, this is not an energy-efficient solution. Therefore, there is a requirement for the multi-hop routing algorithm to carry out inter-cluster communication for transferring the aggregated data from the CH to its sink. This will involve the determination of an energy-balanced shortest route, which in turn, will be a problem that is NP-hard. For the currently existing networks, the problems in clustering and routing were taken into consideration by the researchers. For this work, the problems are jointly considered, and an integrated protocol for clustering as well as routing in the large-scale WSNs is included. The remainder of this investigation has been divided into the sections given below. The literature's associated work is detailed in Section two. All methods employed were explained in Section three. The experimental results were discussed in Section four, and the conclusion was made in Section five

II. Related Works

Cui et al., [7] introduced another new variant of the Bat Algorithm (BA) which in turn was integrated with the centroid strategy. There was the introduction of three other centroid strategies with the utilization of six designs. Additionally, an inertia-free update equation was provided for this. The final performance of optimization had been verified using the CEC2013 benchmarks in the designs using their comparison with the conventional BA. The superiority of the Bat Algorithm with Weighted Harmonic Centroid (WHCBA) was evident from the simulated outcomes. With the integration of the WHCBA into the LEACH protocol, there was the development of a two-stage strategy for CH node selection which had more energy conservation capability in comparison to the LEACH protocol.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6445
***Article Received**: 08 February 2023 **Revised**: 04 March 2023 **Accepted**: 22 March 2023*

To enhance the efficiency of energy in the sensor node lifespan, this research included a protocol that was energy efficient. Referred to as the Low Energy Adaptive Clustering Hierarchy (LEACH), it also included the Genetic Algorithm (GA). Its proposal was put forward by Bhola et al., [8]. The LEACH is a hierarchical protocol type that would involve the conversion of all the sensor nodes to the CH whereas, in the meantime, the CH would forward the data to a target node through aggregation as well as further compression. Moreover, the GA would employ the fitness function to aid in optimal route identification. Upon MATLAB simulation of the code, there was a reduction in the rate of energy consumption by up to 17.39%. Finally, a comparison was made to the currently existing works, and it was observed that the proposed work was efficient.

The LEACH is a very important load-balancing algorithm but is not capable of establishing a performance that is satisfactory as it may be enhanced using other metaheuristic approaches. Zivkovic et al., [9] proposed another refined dragonfly algorithm that was applied to enhance the WSN lifetime. The performance of this algorithm was assessed by comparing it to the original one, the traditional LEACH algorithm, and the PSO. It was evident from the simulated outcomes that the proposed algorithm performed better and was capable of retrieving certain valuable results in the domain. The current protocols that employed a non-optimal CH selection together with the IoT's frequent re-clustering had resulted in a significant level of energy consumption. To a large extent, it would to possible to avoid re-clustering in case there is prior knowledge about the CH lifetime among devices (or nodes). Maratha and Gupta [10] assessed the devices' entire life as the CHs through the resolution of the linear optimization problem for an extension of its first node death to the extent possible as well as for postponement of the process of frequent re-clustering to reduce the energy consumption. A uniform CH distribution was employed by the authors to make sure there was balanced consumption of energy among the IoT devices. The proposed technique of clustering,i.e., the Efficient Clustering using Fuzzy logic based on Estimated Lifetime (ECFEL) for the IoT, had outperformed all the other existing protocols like the LEACH, Dynamic k-LEACH (DkLEACH), the MODified LEACH (MOD-LEACH), the M-IWOCA, the Novel-PSO-LEACH as well as the FM-SCHEL. This included techniques in the First Node Death (FND), the Last Node Death (LND) as well as the Half Node Death (HND). The outcomes of the simulation proved that the ECFEL had a better lifetime, especially concerning its FND, LND as well as HND. Also, it was confirmed that the energy consumed by the ECFEL during the maintenance of the Packet Delivery Ratio (PDR) was less.

There have been several clustering algorithms that were implemented in the recent decade, primarily aiming to ensure a balance in energy consumption for every node and to increase its efficiency. This is known as load balancing. An important representative of such traditional algorithms is the continuous use of LEACH. At the same time, other swarm intelligence meta-heuristics were applied to several NP-hard problems in the WSN domain. Zivkovic et al., [11] presented another improved Grey Wolf Algorithm (GWA) that was applied for improving the optimization of network lifetime. This was employed to form clusters in the process of CH selection. In this research, the authors further evaluated the proposed exploration and enhanced the GWA by making a comparison of this to the traditional PSO, LEACH, and the basic Grey Wolf approach. The results obtained from the simulation proved that the performance was better.

Jagan and Jesu Jayarin [12] proposed a new Fully Connected Energy Efficient Clustering (FCEEC) mechanism that used an electrostatic discharge algorithm to build a network using the shortest path routing from sensor nodes to the CH within a new multi-hop environment. Moreover, the Electro-Static Discharge Algorithm (ESDA) was capable of further improvement of the network lifetime and also attained energy-efficient connectivity among sensors. The result of the ESD was the reduction of the dead node count to ensure the longevity of the network. Finally, the results of the simulation showed an improved performance in metrics like network latency, packet delivery, dead node count, and energy efficiency when compared to other conventional approaches.

Rajesh and Ponmuthu ramalingam [13] proposed another secure as well as energy-aware WSN optimal routing scheme by learning the sensor nodes' dynamic traits with a Bidirectional search based on the Harris Hawk optimizations (LDCSN-BSHHO). The following were the four distinct steps involved in the performance of the optimal routing: i) the clustering, ii) the CH's selection, iii) the data's encryption as well as iv) the routing. In the initial stages, there was the utilization of a method of Weigh Utility-based Stratified Sampling (WUSS) for expansion of the Network Life-Time (NLT). Following that, an Elite Opposition and Ranking Mutation-based Butterfly Optimization Algorithm (EORM-BOA) method was used to determine the clusters' optimal CH choice. Later, there was the application of an Improved Blowfish Algorithm (IBFA) for the Data Packet's (DP) encryption so as provide data security. Lastly, the LDCSN-BSHHO will use an optimum path for forwarding the encrypted DP toward all the base stations. There was a dynamic study made on the behavior of the nodes to choose an optimal path by using the BSHHO algorithm in

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6445
***Article Received**: 08 February 2023 **Revised**: 04 March 2023 **Accepted**: 22 March 2023*

transferring data. This energy, as well as security-centered WSN routing method, is referred to as a secure as well as Energy-Aware Routing (EAR). This method's outcome was examined against other presently-employed techniques, and its efficiency in routing was thus proved.

Dayalan and Kuppusamy [14] presented an improved Evaporation Rate Water Cycle (ER-WC) algorithm to analyze different factors like the location, energy, and network of the CH. The work was to ensure energy efficiency was attained and network throughput was increased. This work provided an optimal method of clustering for the Fuzzy C-means (FCM), in which there was improved efficiency among the WSNs. There were empirical evaluations that were conducted to identify the lifespan of the network, total residual energy, and finally, network stabilization.

III. METHODOLOGY

Present day, clustering is seen as an extremely efficient technique that can save energy. However, in the case of a WSN based on hierarchical clusters, the CHs will end up consuming more additional energy due to an overload that can receive and aggregate data from sensor nodes before sending it to the base station. Thus, the right CH selection has a major role in conserving the energy of the sensor nodes as well as in prolonging the WSN lifetimes. Discussions are provided here on an energy-efficient algorithm for CH selection that is based on both the RFD and the PSO.

A. Low-Energy Adaptive Clustering Hierarchy (LEACH)

The LEACH refers to the initial network protocol which employs the WSN hierarchical routing for the network lifetime extension. All nodes within the network will arrange themselves into local clusters that in turn, will have a single node acting as their CH. While all the other non-CH nodes will transmit data towards the CH, in turn, the CH node will receive data from the cluster members in performing a function for signal processing which is based on data like the data aggregation as well as its transmission towards the BS. Thus, more energy is required for being a CH node in comparison to a non-CH node. So on the death of the cluster head node, other nodes will die and lose their ability to communicate. The LEACH further incorporates a randomized rotation of its high-energy CH position to ensure it can rotate this among sensors to prevent any further battery drainage [15]. As a result, the entire energy load associated with this node being a CH will have an even distribution amongst all the nodes. As the CH node is aware of its cluster members, the CH node will then form a TDMA schedule for informing the node of the data transmission time as well as for the prevention of any intra-cluster collisions.

The LEACH protocol was the first among hierarchical wireless sensor routing protocols [16] that was proposed in the year 2002 by Wendi B. Heinzelman. The architecture of the LEACH is depicted in Figure 1.

Architecture of LEACH

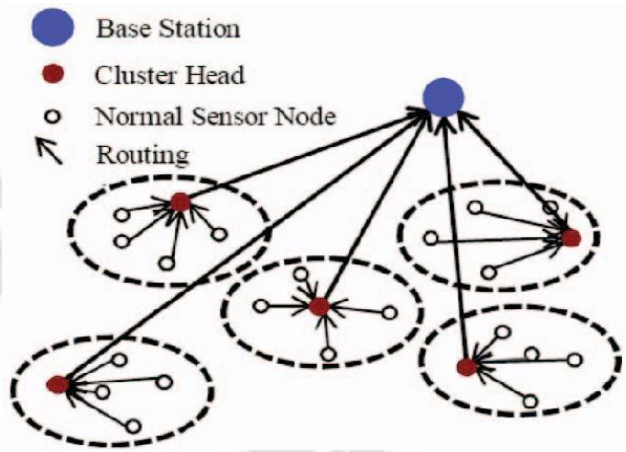


Figure 1

There are many rounds to the LEACH operation, each of which will have the following two distinct phases: the setup phase as well as the steady-state phase. Each cluster's CH will receive this data and will aggregate this data from the cluster members for transmission of such aggregated data towards its BS. Later, there is the performance of the CH selection for each round. Every sensor node will then independently decide about the other sensor nodes that claim to be CH. This is done via the generation of a random number that lies between 0 and 1 as well as drawing a comparison utilizing a threshold $T(n)$. There will be the node's selection as the current round's CH if the generated number is lower than the threshold value, $T(n)$. Computation of this threshold value is given in (1):

$$T(n)=\begin{cases}\dfrac{P}{1-P\left(r \bmod \left(\dfrac{1}{P}\right)\right)}, & n \in G \\ 0, & n \notin G\end{cases} \quad (1)$$

Wherein,n will indicate the actual node number while P will indicate the percentage of the node that is chosen as the CH. r will be the round in which the CH is selected, and G refers to a set of nodes that were not accepted in the form of a CH in the final $1/P$ rounds. A broadcast is used to inform all other nodes of the CH node selection.

B. Particle Swarm Optimization (PSO) Algorithm

The PSO algorithm refers to a nature-inspired one employing swarm intelligence. The PSO has been modeled on observing the flock of birds and their behavior in looking

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6445
***Article Received**: 08 February 2023 **Revised**: 04 March 2023 **Accepted**: 22 March 2023*

out for food sources. This algorithm aimed to identify the particle and its position with a consequence that estimated the cost function. In the process of looking out for food, The bird that finds the food will inform the other birds in the flock of the exact location of the food. The PSO reduces any intra-cluster distance between nodes as well as their CHs. These will now begin transmitting location information as well as residual energy. Such transmissions increase network congestion resulting in the wastage of energy [17].

The PSO algorithm is incorporated using the LEACH in the setup phase and will undergo execution in the BS as its fully centralized. The PSO algorithm is as below:

1. Initialize the S particle after arbitrarily selecting the CHs (2):

$$Xij(0) = \left(xi,\ j(0),\ yi,\ j(0)\right) \tag{2}$$

The due positions of all sensor nodes

2. Compute each particle's cost function as in (3):

a) $\forall ki,\ i = 1, 2, ..., N$ (3)

Assess the distance,d (ki, CHp, q), amongst the nodes ki as well as CHp, q.

Allot the node ki to CHp, q (4):

$$D\left(ki, CHp, q\right) = min\forall q1,\ 2,\ ...,\ \left\{d\left(ki, CHp, q\right)\right\} \tag{4}$$

b) Determine the final cost function as in (5 and 6):

$$cost\ function = \alpha\ .C1 + \left(1 - a\right) \tag{5}$$

$$C1 = \max q = 1, 2, ..., q\{\sum d(ki, CHp, q)\}$$
$$C2 = \sum_{i=1}^{N} E(ki) / \sum_{q=1}^{Q} E(CHp, q) \tag{6}$$

3. For every such particle, the global and personal best has to be determined

4. Change the particle's velocity as well as a position using (7 to 10):

$$Vid(t) = W.Vid(t) + L1.H1(Pbestid - Xid(t)) + L2.H2(Gbest - Xid(t)) \tag{7}$$

$$Xid(t) = Xid(t-1) + Vid(t) \tag{8}$$

If not,

$$Vid(t+1) = W.Vid(t-1) + L1.H1(Pbestid - Xid(t-1)) + L2.H2(Gbest - Xid(t-1)) \tag{9}$$

$$Xid(t+1) = Xid(t) + Vid(t+1) \tag{10}$$

Wherein, X will indicate each particle's position, V will indicate each particle's velocity, t will indicate the time, $L1$, as well as L2, will indicate the learning factors, $H1$as well as $H2$ will indicate acceleration coefficients that are arbitrary numbers that lie between 0 and 1. *Pbestid* will indicate the particle's best position, *Gbest* will indicate the global best position while $W(0 < W < 1)$ will indicate the inertia weight.

There will be a repetition of the updating procedure of V as well as X till it can achieve a good level of value for the global best position (*Gbest*). The article will assess the cost function and also will update the the*Pbestid* as well as the *Gbest* in (11 and 12).

$$Pbestid = \begin{matrix} Pid \\ Pbestid \end{matrix}\Bigg) \begin{matrix} if\ \ cost\ function\ of\ \ Pi < cost\ function\ of\ \ Pbestid \\ else \end{matrix} \tag{11}$$

$$Gbest = \begin{matrix} Pid \\ Gbest \end{matrix}\Bigg) \begin{matrix} if\ \ cost\ function\ of\ \ Pi < cost\ function\ of\ \ Gbest \\ else \end{matrix} \tag{12}$$

5. There is a mapping of the revised positions onto the nearest (x, y) coordinates.

6. There will be reiteration from step 2 to Step 5 to attain the maximum number of iterations.

C. River Formation Dynamics (RFD) Algorithm

RFD refers to a method of heuristic optimization along with a swarm intelligence subset topic. It is based on simulating how water drops are combined to form rivers, and eventually, be combined for joining the sea. This is done by looking out for the path that is the shortest based on the land altitudes from which they flow. For this river formation process, there is the flow of the water drops from a higher altitude towards a lower altitude. As the slope for both of these positions is higher, the water that flows from the higher positions towards the lower positions will erode and will carry away this eroded soil for depositing it at the lower positions. As a result, there will be an increase in a lower position's altitude, and also the formation of the shortest path from a higher position to a lower position [18].

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6445
***Article Received**: 08 February 2023 **Revised**: 04 March 2023 **Accepted**: 22 March 2023*

Algorithm 1 General RFD algorithm
procedure RFD Algorithm
// Stage I : Initialization Stage
Initialization of Drops generating positions;
Initialization of Intermediate positions;
Initialization of Destination(*Sea*) *positions*;
// Stage II : River Formation Stage
while (*not all drops Flow The Same Path*)
and (*not other Ending Condition*) *do*
select _ Forward _ Position();
move _ Drops();
erode _ Path();
add _ Sediments();
end
Analyze the paths;
end procedure

Algorithm 1 will depict the basic RFD algorithm which will have two different stages. The initialization stage will involve the assignation of initial values to three different positions: the Source(S), the intermediate positions (I) as well as the destination (D). Representation of these positions will be done with various altitude values. While D has a zero-value altitude, S and I will have positive values of the altitude. The positions that generated water drops will continue to do so. All intermediate positions will receive water drops from the source and then will forward them to the sea. During the river formation stage, there will creation of a river between the positions that generate the drop and the Sea by using an iterative process with functions such as Forward Position (), erode Path (), move Drops () as well as add Sediments (). There is this iterative procedure's repetition till such time the drops either will follow this path or will meet the termination conditions like limited time for execution or limited iterations.

The drop-generating positions will choose the subsequent neighbor positions to forward the drops in the select Forward Position () function based on a probability function, P I j), as shown in Equation (13), where I and j will be positioned such that I S or I I) and (j I or j D). P I j) indicate that position I has a chance of selecting position j as its next hop position to forward drops.

$$P(i,j)=\begin{cases}\dfrac{DG(i,j)}{\sum_{l\in Nb(i)}DG(i,l)} & if\ \ j\in Nb(i)\\ 0 & Otherwise\end{cases} \quad (13)$$

Wherein, Nb (i) will indicate the position i's neighbors while DG (i; j) will indicate the Decreasing Gradient that is observed between node i as well as node j. This is computed by employing Equation (14) given below:

$$DG(i,j)=\frac{\left(altitude(i)-altitude(j)\right)}{distance(i,j)} \quad (14)$$

For the function erode Path (), the drop movements will be placed, and the paths get eroded. In case the drop moves from A to B, it erodes A and will deposit the soil to B by using a function and adding Sediments (). This means the altitude for position A will be reduced, and the altitude value for position B will be increased based on its current gradient falling between A and B. In case the downslope falling between A and B is found to be higher, the erosion will also be higher. The destination position and its altitude (SEA) will not be modified and will be equal to 0 at the time of execution. Lastly, the paths are analyzed and will be formed using drops or stores of an optimized path.

It is possible to observe similarities between the RFD as well as the WSN's data collection procedure: for the RFD, the source (which is drop-generating), as well as their positions, will produce water drops that are interested in meeting the destination or the Sea. Therefore, while the sensor data will act as the water drops, the source positions will act as either the sensor nodes or the Base station,i.e., the Sea. Upon combination of these drops, they will flow from the source towards the sea for the formation of rivers based on the RFD's position as well as its altitude value. So, the sensor nodes will forward the WSN data by forming a path toward the BS. The work proposed a new RFD approach for choosing an optimal location for the CH. But, the algorithm's main objective had been to minimize any intra-cluster distance, ignoring its distance towards the sink, and this is a desirable metric for enhancement of the network's energy efficiency. Further, the assignation of non-CH to the CHs can result in an imbalance of energy in the network [19].

D. Naïve Bayes Classifier

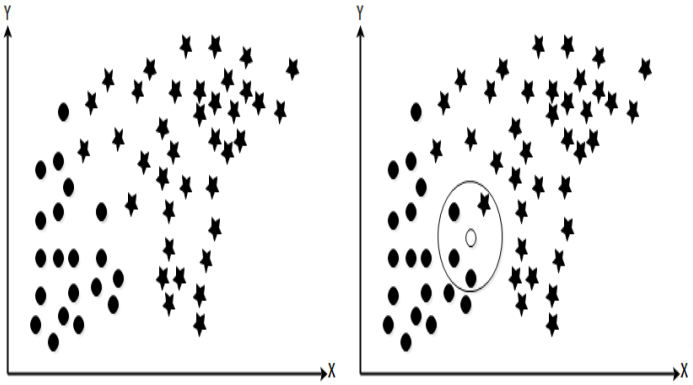
The Nave Bayesian classification method will be a supervised machine learning algorithm that will classify observations using the algorithm's own rules. This new classification tool will be trained using a learning dataset to display the required entries. In the learning phase, the algorithm will develop the rules of classification for the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6445
Article Received: 08 February 2023 Revised: 04 March 2023 Accepted: 22 March 2023

dataset and will apply them again to the prediction dataset. This Naïve Bayesian classifier implied that this learning dataset's classes are provided, and therefore, will result in this tool's supervised nature. The classifier is yet another simple method that has been employed in supervised learning based on Bayes' theorem [20].

For classifying and further providing a new concept to this method, another explanatory example is given in Figure 2 a. There are new objects that are classified as either Circles or Stars. The primary task in this was for the classification of all the new cases upon their arrival, and also for the determination of their class labels based on their objects.

Classification example using Naïve Bayes



(a) (b)

Figure 2

Thus, the task of classification will be recommended for prior probability, posterior probability, and likelihood. For the Bayesian analysis, the prior probabilities were based on experiences where the percentage of their Circles or Stars objects have been employed for advanced prediction of their results.

Once the prior probabilities are calculated, there is an application of a likelihood principle for a new object's (Noted WHITE circle) classifications depicted in Figure 2 b. For its measurement, there will be the definition of a circle around *X*, which is also inclusive of several points that are selected a priori (which will imply their independence of the class labels). After this, the circle's number of points will be computed based on each of these class labels. The likelihood is subsequently computed as below:

The Bayesian analysis's classification is done by combining the following two distinct information sources: the prior as well as the likelihood to yield a new posterior probability.

Lastly, there will be the classification of *X* (white circle) as Stars due to it having the highest posterior probability for its membership with the class.

E. Adaboost Classifier

Adaboost [21] is an abbreviation for adaptive boosting, which refers to a dichotomy classification algorithm that can train weak classifiers and combine them to form a new and strong classifier that can meet classification needs. The weak classifier's misclassification of the sample's weight to increase the sample's weight for training the subsequent weak classifier is referred to as adapting the Adaboost. There will be no determination of the last strong classifier until either a low error rate or the maximum number of iterations is met. Adaboost's weak classifiers, unlike the Gradient Boosting Decision Tree (GBDT) algorithm, are independent of one another. Because of the connection with the GBDT's weak classifiers, an additional set of parameter control functions is essential to ensure the prevention of any errors in training. It was proved from the result of the Adaboost Algorithm that it is chosen when quick and precise classification is needed.

Suppose that there is an Adaboost algorithm [22] which is having N training points (x_i; y_i), they will be $x_i \in X$ as well as $y_i \in \{-1, +1\}$. For round *m*, wherein m = 1,……M, it will fit G_m (x), a new as well as weak classifier, for the dataset's new version, which in turn will undergo reweighting with w_m, a weighting vector. It further computes the weighted misclassification rate for the chosen learner for updating its weighting measure in the subsequent round, $w_m + 1$. The last classifier will symbolize a new weighted linear classifier combination from every part of the algorithm. It is practically possible to use regularization for limiting the actual number of rounds.

F. Support Vector Machine (SVM) Classifier

The development of the SVM originated from the statistic learning theory concept of the late 70s. It was focused on the following two-class classification problems: a linear line or hyperplane that had been built as a decision boundary for two classes. The term support vectors are used to refer to those data points nearest to the hyperplane imparting its construction [23] [24]. Therefore, the algorithm will be an SVM. The mathematical expression of this optimized hyperplane is provided below as (15):

$$w^T x + b = 0 \tag{15}$$

Wherein,*w* will be the vector of weights, *x* will be the input vector and *b* will be the bias. Given below are the support vector equations for each class (16):

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6445
***Article Received**: 08 February 2023 **Revised**: 04 March 2023 **Accepted**: 22 March 2023*

$$w^T x + b = +1, \ for \ d_i = +1$$
$$w^T x + b = -1, \ for \ d_i = -1 \quad (16)$$

Wherein, d_i refers to its class,i.e., for class A, $d_i = +1$, and class B, d_i = -1. For a certain training sample, $\{(x_i, d_i)\}_{i=1}^{k}$, the optimization problem involved the identification of its optimal hyperplane as shown in (17):

$$\min \Phi(w) = \frac{1}{2} w^T w, \quad (17)$$

In such a way that, $d_i(w^T x_i + b) \geq 1, \ for \ i = 1, 2, ..., k,$

Acquisition of a final decision function will be as per (18):

$$f(x) = sign\left(\sum_{i=1}^{N} \alpha_{0,i} (x^T x_i) + b \right) \quad (18)$$

Wherein,*x* will be the classified input vector while *N* will be the actual number of support vectors within its training phase. For the definition of the support vectors from the input vectors, there is a utilization of all the non-negative parameters, $\alpha_{0,i}$. There is the transformation of all the linearly non-separable patterns into a new feature space with a mapping function, $\varphi(x)$, that will permit data classification with a linear hyperplane. A decision function as in Equation (19) is now modified to:

$$f(x) = sign\left(\sum_{i=1}^{N} \alpha_{o,i} (\varphi(x) \varphi_i(x)) + b \right) \quad (19)$$

$K(x; y) = \varphi(x)\varphi(x_i)$, an inner-product kernel function, has been employed for the reduction of the complexity of high-dimensional numerical optimization [25]. This will involve an update of the decision function as in (20):

$$f(x) = sign\left(\sum_{i=1}^{N} \alpha_{0,i} K(x, x_i) + b \right) \quad (20)$$

The SVM will employ various t kernel functions such as the sigmoid, the linear, the polynomial, and the Radial-Basis functions (RBF) for the non-linear classification of patterns.

IV. RESULTS AND DISCUSSION

In this section, the LEACH, PSO, RFD, naïve Bayes, Adaboost, and SVM methods are discussed. Experiments are carried out using 500 to 3000 nodes and 0 to 800 rounds. The average end-to-end delay, average Packet Delivery Ratio (PDR), percentage of nodes alive, average recall, average precision, and average f measure as shown in tables 1 to 6 and figures 3 to 8.

TABLE 1: Average End-to-End Delay for RFD

Nodes	LEACH	PSO	RFD
500	0.0017	0.0016	0.0016
1000	0.0016	0.0015	0.002
1500	0.0169	0.0157	0.019
2000	0.0297	0.0272	0.023
2500	0.0633	0.0599	0.0518
3000	0.0682	0.0642	0.0542

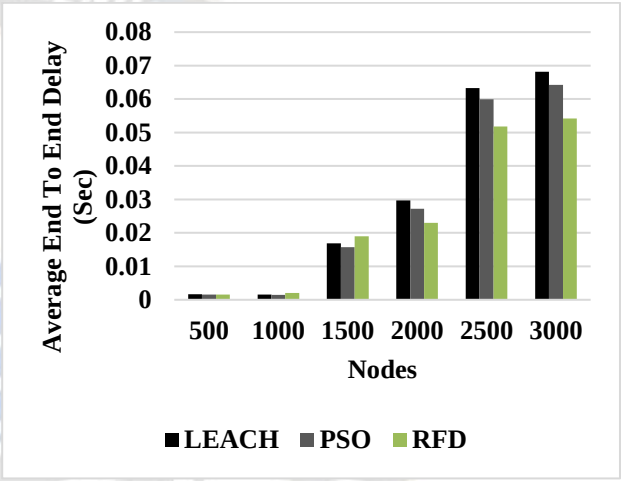


Figure 3

From figure 3, it can be observed that the RFD has a lower average end-to-end delay of 6.06% & no change for 500 nodes, 22.22% & 28.57% for 1000 nodes, by 11.69% & 19.02% for 1500 nodes, 25.42% & 16.73% for 2000 nodes, by 19.98% & 14.5% for 2500 nodes and by 22.87% & 16.89% for 3000 nodes when compared with LEACH and PSO respectively.

TABLE 2: Average Packet Delivery Ratio for RFD

Nodes	LEACH	PSO	RFD
500	0.8417	0.8738	0.9326
1000	0.7966	0.8316	0.9114
1500	0.784	0.8075	0.8857
2000	0.7499	0.7746	0.8273
2500	0.6793	0.7127	0.7802
3000	0.5881	0.6143	0.6889

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6445
***Article Received**: 08 February 2023 **Revised**: 04 March 2023 **Accepted**: 22 March 2023*

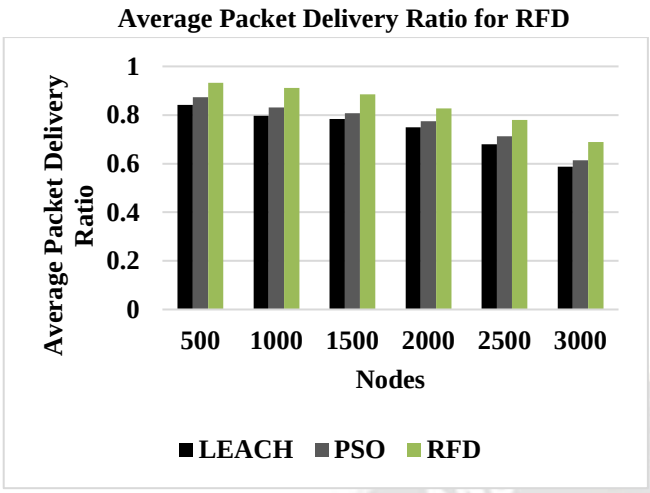


Figure 4

From figure 4, it can be observed that the RFD has a higher average PDR of 10.24% & 6.51% for 500 nodes, by 13.44% & 9.15% for 1000 nodes, 12.18% & 9.23% for 1500 nodes, 9.81% & 6.57% for 2000 nodes, by 13.83% & 9.04% for 2500 nodes and by 15.78% & 11.45% for 3000 nodes when compared with LEACH and PSO respectively.

TABLE 3: Percentage of Nodes Alive for RFD

Number of rounds	LEACH	PSO	RFD
0	100	100	100
100	89	92	100
200	70	79	87
300	59	72	81
400	12	34	48
500	3	11	22
600	0	2	11
700	0	0	7
800	0	0	2

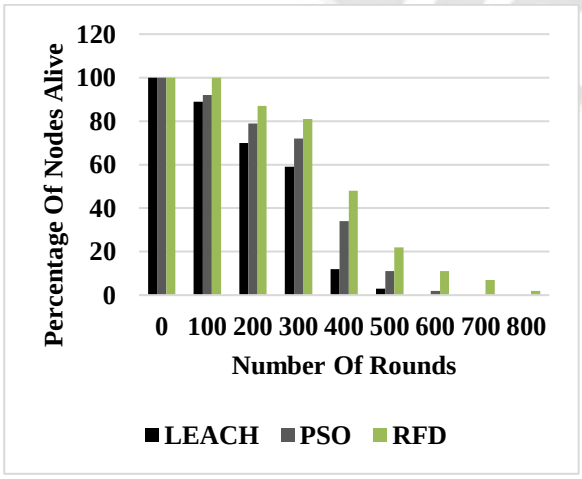


Figure 5

From figure 5, it can be observed that the RFD has a higher percentage of nodes alive by 11.64% & 8.33% for 100 rounds, 21.65% & 9.63% for 200 rounds, 31.43% & 11.76% for 300 rounds, 120% & 34.14% for 400 number of rounds and by 152% & 66.67% for 500 number of rounds when compared with LEACH and PSO respectively.

TABLE 4: Average Recall for SVM

	Naïve Bayes	Adaboost	SVM
25k events/second	0.73	0.79	0.88
50k events/second	0.73	0.81	0.88

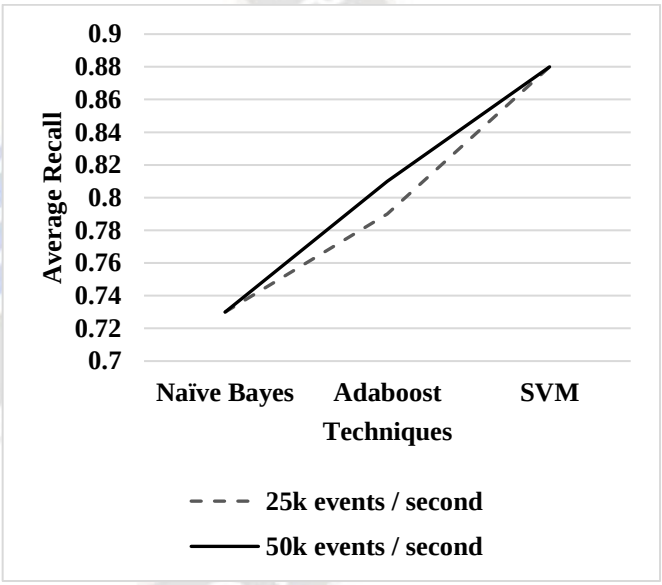


Figure 6

From figure 6, it can be observed that the SVM has a higher average recall of 18.63% & 10.77% for 25k events/second by 18.63% & 8.28% for 50k events / second when compared with naive Bayes and Adaboost, respectively.

TABLE 5: Average Precision for SVM

	Naïve Bayes	Adaboost	SVM
25k events/second	0.61	0.68	0.8
50k events/second	0.58	0.66	0.78

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6445
***Article Received**: 08 February 2023 **Revised**: 04 March 2023 **Accepted**: 22 March 2023*

Average Precision for SVM

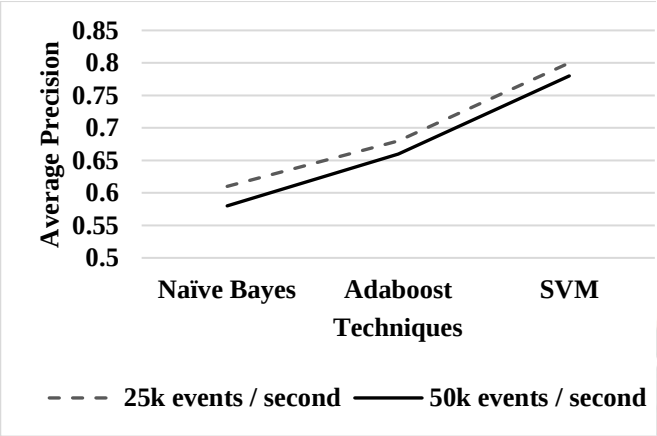


Figure 7

From figure 7, it can be observed that the SVM has higher average precision of 26.95% & 16.22% for 25k events/second by 29.41% & 16.67% for 50k events / second when compared with naive Bayes and Adaboost, respectively.

TABLE 6: Average F Measure for SVM

	Naïve Bayes	Adaboost	SVM
25k events/second	0.66	0.73	0.84
50k events/ second	0.65	0.73	0.81

Average F Measure for SVM

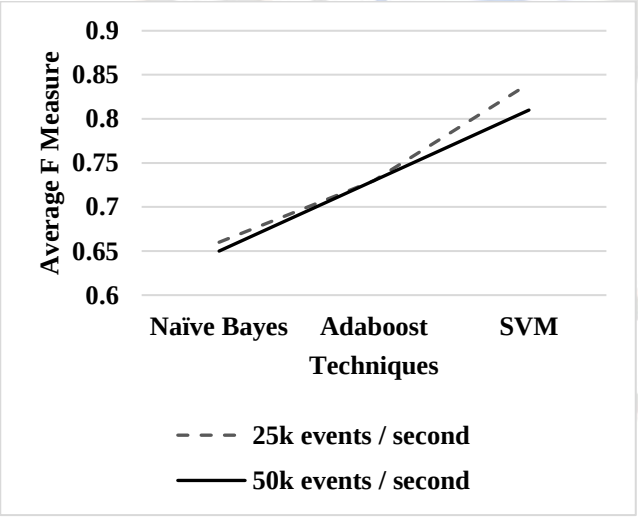


Figure 8

From figure 8, it can be observed that the SVM has a higher average f measure by 24% & 14.01% for 25k events/second by 21.92% & 10.38% for 50k events / second when compared with naive Bayes and Adaboost, respectively.

V. CONCLUSION

The number of devices that are connected to the Internet has been on the increase, thus anticipating the era of the IoT. But handling such Big Data from the IoT networks can pose a major challenge to decision-makers. The WSNs are a major source of data, and in this, there has been a wider range of avenues monitored by many thousands of sensors wherein all gathered data can be forwarded to the sink node. However, the WSNs also pose challenges to other networks. Clustering is an effective technique that is energy-efficient, and for this work, a CH selection algorithm based on the RFD and PSO has been proposed. The PSO is very efficient as a nature-inspired algorithm with higher solution quality and the capacity to escape the local optima aside from its quick convergence. The RFD constructs solutions by a modification of values connected to the graph nodes. Gradient orientation for this will provide certain important features like quick reinforcements of shortcuts, a focus on eliminating blind alleys, and natural cycle avoidance. So, the algorithm is useful in choosing an energy-aware CH that is based on a new fitness function taking into consideration the node residual energy. When compared to the LEACH and PSO, the proposed RFD showed a higher average PDR by about 10.24% and 6.51% for the 500 nodes, about 13.44% and 9.15% for the 1000 nodes, about 12.18% and 9.23% for the 1500 nodes, about 9.81% and 6.57% for the 2000 nodes, about 13.83% and 9.04% for the 2500 nodes, and finally, about 15.78% and 11.45% for the 3000 nodes.

REFERENCES

[1] Yassen, M. B., Aljawaerneh, S., & Abdulraziq, R. (2016, September). Secure low energy adaptive clustering hierarchal based on Internet of things for wireless sensor network (WSN): Survey. In 2016 International Conference on Engineering & MIS (ICEMIS) (pp. 1-9). IEEE.

[2] Kim, B. S., Kim, K. I., Shah, B., Chow, F., & Kim, K. H. (2019). Wireless sensor networks for big data systems. Sensors, 19(7), 1565.

[3] Din, S., Ghayvat, H., Paul, A., Ahmad, A., Rathore, M. M., & Shafi, I. (2015, December). An architecture to analyze big data in the Internet of things. In 2015 9th International Conference on Sensing Technology (ICST) (pp. 677-682). IEEE.

[4] Behera, T. M., Samal, U. C., & Mohapatra, S. K. (2018). Energy-efficient modified LEACH protocol for IoT application. IET Wireless Sensor Systems, 8(5), 223-228.

[5] Jagadeesh, S., & Muthulakshmi, I. (2022). Hybrid Metaheuristic Algorithm-Based Clustering with Multi-Hop Routing Protocol for Wireless Sensor Networks. In Proceedings of Data Analytics and Management (pp. 843-855). Springer, Singapore.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6445
***Article Received**: 08 February 2023 **Revised**: 04 March 2023 **Accepted**: 22 March 2023*

[6] Gupta, G. P., & Jha, S. (2018). Integrated clustering and routing protocol for wireless sensor networks using Cuckoo and Harmony Search-based metaheuristic techniques. Engineering Applications of Artificial Intelligence, 68, 101-109.

[7] Cui, Z., Cao, Y., Cai, X., Cai, J., & Chen, J. (2019). Optimal LEACH protocol with modified bat algorithm for big data sensing systems in the Internet of Things. Journal of Parallel and Distributed Computing, 132, 217-229.

[8] Bhola, J., Soni, S., & Cheema, G. K. (2020). Genetic algorithm-based optimized leach protocol for energy-efficient wireless sensor networks. Journal of Ambient Intelligence and Humanized Computing, 11(3), 1281-1288.

[9] Zivkovic, M., Zivkovic, T., Venkatachalam, K., & Bacanin, N. (2021). Enhanced dragonfly algorithm adapted for wireless sensor network lifetime optimization. In Data intelligence and cognitive informatics (pp. 803-817). Springer, Singapore.

[10] Maratha, P., & Gupta, K. (2022). Linear optimization and fuzzy-based clustering for WSNs assisted the Internet of things. Multimedia Tools and Applications, 1-25.

[11] Zivkovic, M., Bacanin, N., Zivkovic, T., Strumberger, I., Tuba, E., & Tuba, M. (2020, May). Enhanced grey wolf algorithm for energy-efficient wireless sensor networks. In 2020 zooming innovation in consumer technologies conference (ZINC) (pp. 87-92). IEEE.

[12] Jagan, G. C., & Jesu Jayarin, P. (2022). Wireless Sensor Network Cluster Head Selection and Short Routing Using Energy Efficient ElectroStatic Discharge Algorithm. Journal of Engineering, 2022.

[13] Rajesh, A. E., & Ponmuthuramalingam, P. (2021). Secure and energy-efficient optimal routing scheme for wireless sensor networks using IBFA and LDCSN-BSHHO algorithms. Turkish Journal of Computer and Mathematics Education, 12(10), 2640-2656.

[14] Dayalan, V., & Kuppusamy, M. (2022). An improved evaporation rate water cycle algorithm for energy-efficient routing protocol in WSNs. International Journal of Intelligent Computing and Cybernetics.

[15] Malik, M., Singh, Y., & Arora, A. (2013). Analysis of LEACH protocol in wireless sensor networks. International Journal of Advanced Research in Computer Science and Software Engineering, 3(2), 178-183.

[16] Quynh, T. P. T., & Viet, T. N. (2021). Improvement of LEACH based on K-means and Bat Algorithm. International Journal of Advanced Engineering Research and Science, 8, 2.

[17] Gambhir, A., & Payal, A. (2019). Analysis of particle swarm and artificial bee colony optimization-based clustering protocol for WSN. International Journal of Computational Systems Engineering, 5(2), 77-81.

[18] Guravaiah, K., & Velusamy, R. L. (2015). RFDMRP: River formation dynamics-based multi-hop routing protocol for data collection in wireless sensor networks. Procedia Computer Science, 54, 31-36.

[19] Rao, P. C., Jana, P. K., & Banka, H. (2017). A particle swarm optimization based energy efficient cluster head selection algorithm for wireless sensor networks. Wireless networks, 23(7), 2005-2020.

[20] Messaoud, S., Bradai, A., Bukhari, S. H. R., Quang, P. T. A., Ahmed, O. B., & Atri, M. (2020). A survey on machine learning in the Internet of things: algorithms, strategies, and applications. Internet of Things, 12, 100314.

[21] Wang, F., Jiang, D., Wen, H., & Song, H. (2019). Adaboost-based security level classification of mobile intelligent terminals. The Journal of Supercomputing, 75(11), 7460-7478.

[22] Wyner, A. J., Olson, M., Bleich, J., & Mease, D. (2017). Explaining the success of AdaBoost and random forests as interpolating classifiers. The Journal of Machine Learning Research, 18(1), 1558-1590.

[23] Jan, S. U., Ahmed, S., Shakhov, V., & Koo, I. (2019). Toward a lightweight intrusion detection system for the Internet of things. IEEE Access, 7, 42450-42471.

[24] S. Nimmala, Y. Ramadevi, B. A. Kumar, and R. Sahith, "An intelligent AAA approach to predict high blood pressure using PARP classifier," Clinical Epidemiology and Global Health, vol. 7, no. 4, pp. 668–672, 2019.

[25] K. Dinesh and S. K. SVN, "Trust aware secured energy efficient rule-based fuzzy clustering protocol with modified sunflower optimization algorithm in wireless sensor networks," 2022.