International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-1, November 2019

Ensembled Adaboost Learning with Id3 Algorithm for Energy Aware Data Gathering in WSN

G. Kalaimani, G. Kavitha

Abstract: ***In this paper, data collection is the operation of gathering a lot of details from the sensor nodes and shipping it to the sink node. The use of Network is increasingly required to perform these processes, so, increases energy consumption. A lot of WSN architectures are designed to solve this complex problem. By using a technique called Decision Tree Classifier using Adaboost (DTCA) algorithm, can extension that data collection efficiency, as well as reducing the delay and Power consumption. In the proposed methodology, the power of each sensor node should be estimated at the outset. Then the mobile sink node receives the information from the high power sensor nodes with minimal delay. The mobile sink node classifies the data pockets using the Decision Tree Classifier. This classifies based on the relationship between the sensor nodes in WSN. That relationship is measure using the method of population Pearson product moment correlation coefficient. Adaboost algorithm is a combination of several weak non-linear classifiers to create a higher classification. Then finally, it sends classified particulars to Base Station. The operation of the DTCA system is convey out with divergent parameters such as classification time, EC, (Network Lifetime) NL, data collection capability, Classification Accuracy (CA), (FPR) false positive rate and delay.***

Keywords: ***Data Collection, WSN, Classification, Energy Consumption, DTCA, Pearson Product-Moment Correlation Coefficient***

I. INTRODUCTION

During this period, Network's applications are growing; this is because people are now using the Internet as much as maximum to exchange their opinion. Not only that, but people watch cricket, volley ball and other such games through their mobile phone online network at home. As the utility of the Internet increases, so will power. This is because all the devices that use the Internet need power. Nowadays, everyone is using WSN. A lot of data's goes online and it can save automatically. So, the need of storage is increased. Thus, in this paper, analyze the data and then classify it this method has been proposed to transport only the required information to the base station.

WSN is used to monitor physical or the ecosystem conditions. These monitoring data are sent to BS by WSN's. Directly or indirectly, the sink node collects information through the intermediate nodes. Recently, many methods have been developed to collect this data.

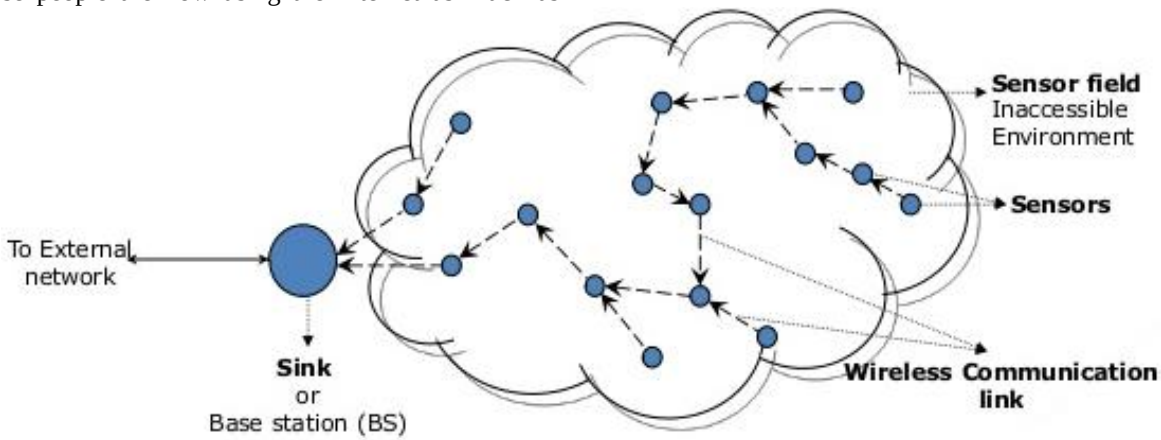

Fig.1: WSN

To achieve appropriate data collection, in [1] has been distributed and a data collection approach has been introduced. But the reality is that it categorizes the data packet. For Mobile Data Collection a two-step approach was designed.

Revised Manuscript Received on November 30, 2019.
***** Correspondence Author
Dr. G. Kalaimani*, Professor, Department of Computer Science and Engineering, Shadan Women's College Of Engineering & Technology, Hyderabad, Telangana, India. E-mail:kalaimaniphd@gmail.com,
Dr. G. Kavitha, Professor, Department of Computer Science and Engineering, Muthayammal Engineering College, Kakkaveri, Rasipuram, Tamil Nadu, India. E-mail:kavitha032003@yahoo.co.in

Even so, the delay in collecting data was high. Lacking a structure and the energy balanced data collection protocol designed in [3]. However the combined data are not classified. To collect data with low EC Multi-Layer Clustering Program developed in [4]. But, EC's performance is low. Communication Awareness ETX (CA-ETX) was introduced. But reliability does not improve performance.

In [6], a tree-cluster-based data-gathering algorithm (TCBDGA) was designed. Though, data loss was not considered. An anchor selection based on tradeoff between neighbor amount and residual energy (AS-NAE) technique was introduced in [7]. But data collection efficiency was not enlarged. An energy efficient structure-free data aggregation and delivery (ESDAD) protocol was planned in [8]. However, it does not have efficient measurement in the sensing department. A distributed data compression framework was established in [9]. But the classification was not implementing. A Cell-based Path Scheduling (CPS) algorithm was designed in [10]. But, the power of sensor nodes was not measured.

So above given problems are overcome by using DTCA technique.

II. PROPOSED METHODOLOGY

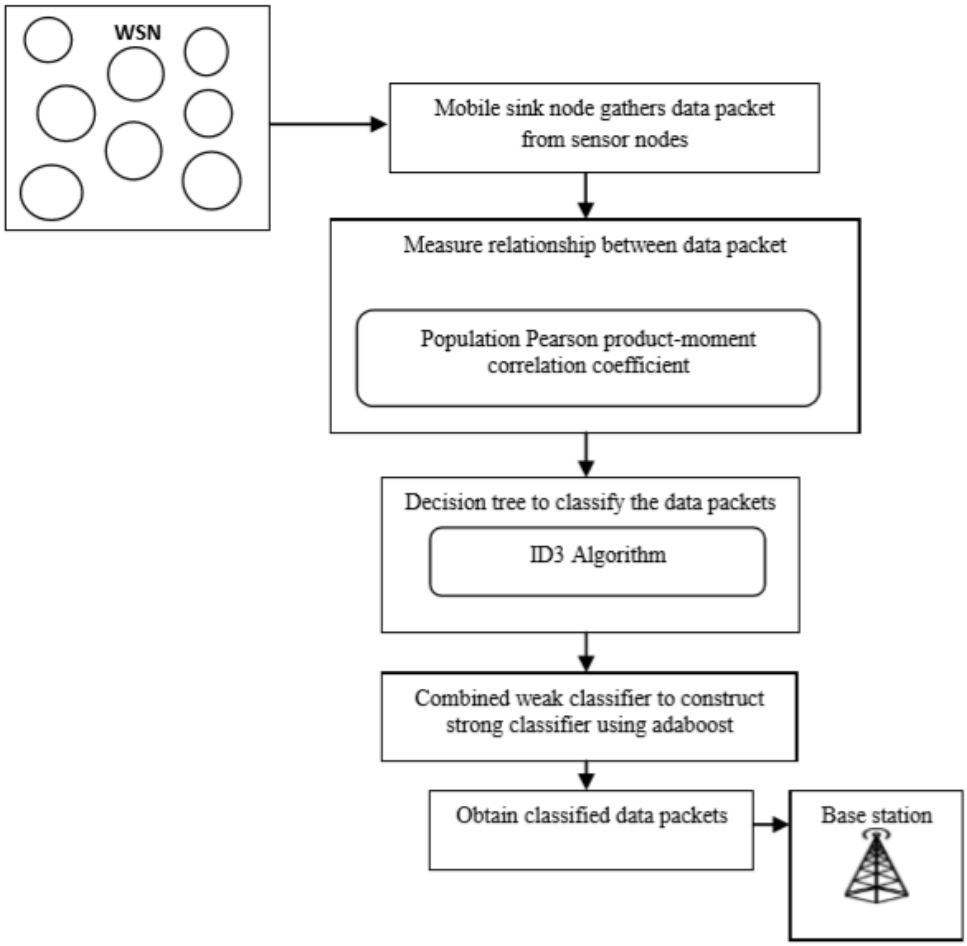

Fig.2: Flow diagram of proposed method

1) Mobile sink node gathers data packet from sensor nodes

Sensors in WSN are used to collect available data on the environment surrounding them, those sensors are hardware tools. They are all used to measure changes in the physical state, such as pressure or temperature. Sensors monitor and collect physics data very accurately and sensitively. But it requires more energy. So the minor the size of the sensor node, the less energy it takes.

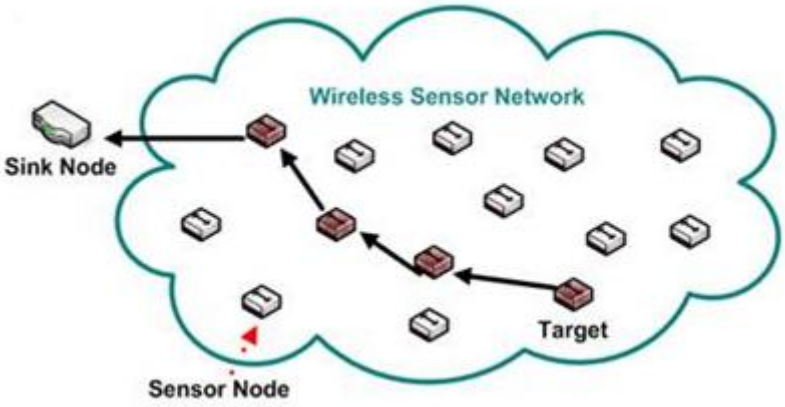

Fig.3: Sensor Node in WSN

2) Population Pearson Product-Moment Correlation Coefficient

The regression analysis is a statistical process for finding the relationship between data packets using population Pearson product-moment correlation. It is utilized to compute the correlation between the data packets. It has a point between the gamut from +1 to −1,

Where,

1 = positive correlation,

0 = no correlation

−1 = negative correlation between the data packets.

Pearson's correlation is the proportion of covariance of two data packets and the multiplicand of their standard deviations. Hence the name involves a product moment correlations. The sensor collects information about nodes, pressure, temperature, humidity and various ambient conditions. Each sensor node in WSN broadcasts a number of data packets. So sums are created in data packets. In order to efficient classify data packets; values regarding the relationships between data packets are very important.

Let us appraise the no. of data packets $DP_1, DP_2, DP_3, \ldots DP_n$. The relationship between the data packets are measured as follows,

$$\rho_{(DP_1,DP_2)} = \frac{C(DP_1,DP_2)}{\sigma_{DP_1},\sigma_{DP_2}} \quad (1)$$

From (1),

$\rho_{(DP_1,DP_2)}$ = a population Pearson product-moment correlation coefficient

C = a covariance between data packets DP_1 and DP_2.

σ_{DP_1}&σ_{DP_2}= a standard deviation of two data packets DP_1 and DP_2.

The above equation (5), is derived and obtaining the final correlation results,

$$\rho_{(DP_1,DP_2)} = \frac{\sum DP_1,DP_2 - \frac{\sum DP_1 * \sum DP_2}{n}}{\sqrt{\left(\sum DP_1{}^2 - \frac{(\sum DP_1)^2}{n}\right)} * \sqrt{\left(\sum DP_2{}^2 - \frac{(\sum DP_2)^2}{n}\right)}} \quad (2)$$

Where,

n = no. of data packets

DP_1&DP_2= two data packets

$\sum DP_1$= the sum of DP_1

$\sum DP_2$ = the sum of DP_2

$\sum DP_1 * DP_2$= the sum of cross multiplicand of DP_1 andDP_2

$\sum DP_1{}^2$ = the sum of squared score of DP_1

$\sum DP_2{}^2$ = the sum of squared score of data packet 2.

The population Pearson product-moment correlation provides positive correlation +1 which means the correlation between the data is discovered. The negative correlation allocates fake relationship between the data.

3) Decision Tree Classifier

Decision trees are some of a kind of a supervised learning algorithm. That is, Researchers have to describe what the output is for the given input. In this case, a data is sustainably separated up to a definite certain parameter. The decision tree can be splitter into two piece, that is, the details can be split and sustained by the leaves and nodes. Result nodes are where the data is separated. There are several methods to create decision tree. But there is a better mechanism than those steps. Its name is ID3 Algorithm. Its full name is Iterative Dichotomiser 3.

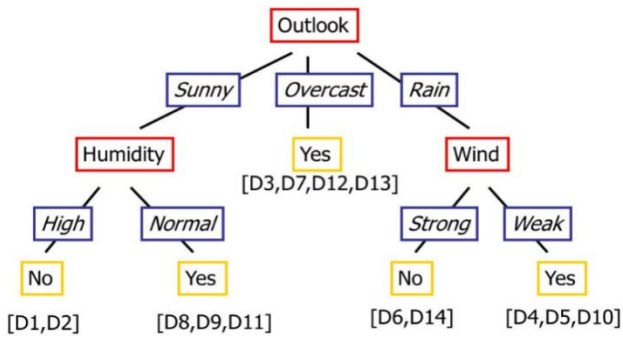

Fig.4: Basic tree classification

Shannon calls entropy the entropy. It is a finite set S is indicated by H(S). This is a compute of the inconsistency or uncertainty in the data.

Mathematically, entropy can be written as:

$$H(S) = \sum_{x=X} Dp(x) \log 2\ 1/Dp(x)$$

For a particular s attribute the leading change in entropy after verification is called Kullback-Leibler divergence, indicated by IG(S, A) for a set.

This relates to independent variables, measuring the change in entropy.

$$IG [S, A] = H[S] – H[S, A]$$

Alternatively,

$$IG[S, A] = H(S) - \sum_{i=0}^{n}(Dp(x) * H(x))$$

By using feature A, Information gain is IG[S, A]. H is the entropy of the whole set, after using the second word a feature, Computing entropy. Here, P(x) probability is the probability of x.

Steps for ID3 Algorithm:

1. For tree design the root node.
2. If the exemplifications are positive, return the leaf node to the positive.
3. If the exemplifications are negative, return the leaf node to the negative.
4. The temporary constant must evaluate the entropy of H[S].
5. For each characteristic, the entropy must be quantified with respect to the x characteristic indicated by H[S, x].
6. Choose the character with the maximum point of IG[S, x].
7. From the set of characteristics, the possessions that deliver the most IG should be banished.
8. Perform again all steps, until run out or in the decision tree, all leaf solutions will be.

4) Adaboost Algorithm

Boosting is a common ensemble system, this, from many weak classifiers, creates a strong classification. By generating a model from the training ground, this is done, and then make-up a 2nd model to correct errors from the 1st model. Until the maximum number of samples is added, or until the training set is correctly predicted, new models are added. In the classification of data problems, Adaboost is best for maximizing the effectiveness of decision trees. The most appropriate and common method used with the Adaboost method is the decision tree. Every step in the training data set, weighs.

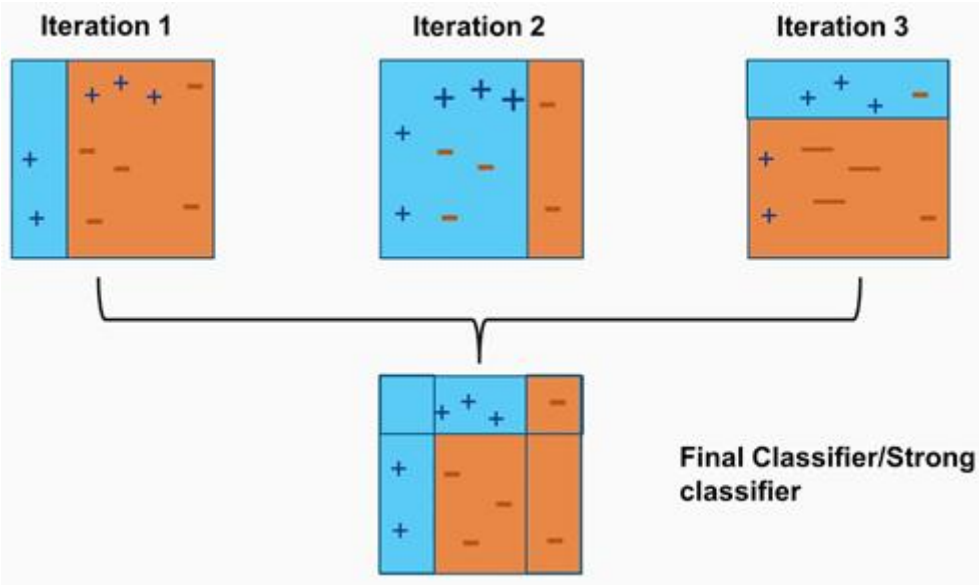

Fig.5: Create a strong classifier

The initial weight can be group as follows:

$$Weight (xi) = 1/n$$

Where,

xi = i'th training step

n = no. of training steps

The false classification rate is enumerated for the trained sample. From past time, it is enumerated as follows:

$$error = [errorless – N] / N$$

Where,

error = misclassification rate

N = total number of training steps.

The following equation has been modified to utilize the weight of the training steps:

$$error = sum (w [i] . terror [i]) / sum [w]$$

Where,

w = weight for training step i

terror = prognostication error for training step i

The value of the trained model is enumerated as follows:

$$stage = ln [(1-error) / error]$$

Where,

stage = The location value is utilized for weight estimation from the model.

ln () = natural logarithm

error = misclassification error for the model.

III. SIMULATION RESULTS AND DISCUSSIONS

An efficient DTCA technique is implemented in NS2.34 network simulator.

1) Energy Consumption

Figure 4 shows the results of EC versus number of sensor nodes in WSN. As a result, EC of ABNRTC technique considerably reduced by 10%, 16%, 24% and 29% as compared to existing [1], [2] , [3] and [4] respectively.

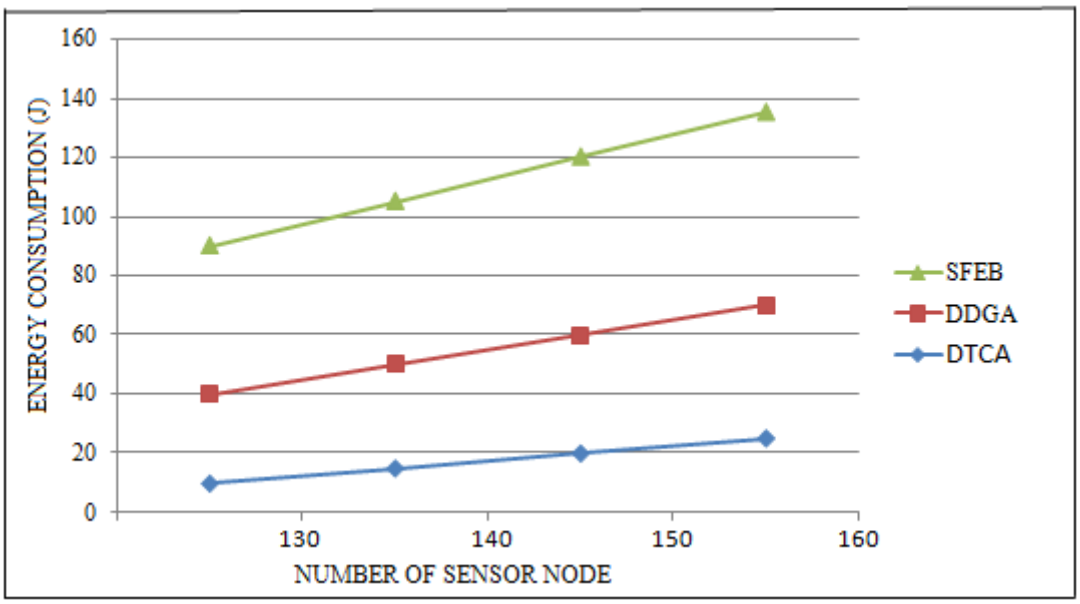

Fig.6: Simulation result of energy consumption

2) Network Lifetime

The use of efficient sensor nodes, which utilize very little power for data collection in WSN, extensions the period of the network lifetime. It is enumerated in percentage terms and below.

$$\mathrm{NL} = \frac{\text{No.of energy efficient sensor nodes are selected}}{\mathrm{n}} * 100 \quad (3)$$

From (3),

NL = network lifetime

n = no. of sensor nodes in network.

Figure 7 characteristics a simulation result of NL versus no. of senor nodes. As shown in the figure, DTCA technique gives better performance in NL than the other existing methods. This shows the noteworthy advancement of simulation results with energy efficient sensor nodes.

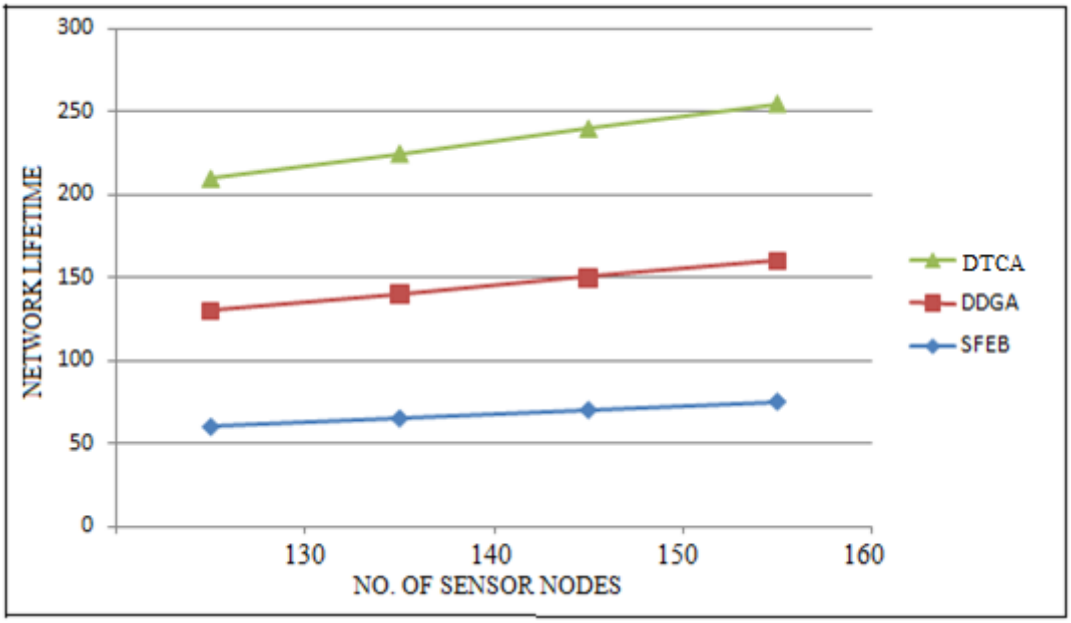

Fig.7: Simulation result of Network Lifetime

3) Data Gathering Efficiency

The efficiency of data collection is calculated based on the data packet loss rate, these outcomes in the no. of data packets being lost to the total no. of data packets. So this is calculated as follows:

$$\mathrm{DGE} = \frac{\text{Number of data packet lost}}{\text{Number of data packets}} * 100 \quad (4)$$

From (4) DGE regarded as data gathering efficiency. It is enumerated in percentage, in table 1, from 9 to 90, the about states the capacity to fetch details based on the number of packets. This enhances the capacity to collect information utilizing DTCA compared to other techniques.

Table 1: Tabulation for Data gathering efficiency

Number of data packets	Data Gathering Efficiency		
	DTCA	DDGA	SFEB
9	22	32	54
18	24	35	57
27	26	38	59
36	30	41	61
45	33	43	63
54	35	45	65
63	37	48	67
72	39	51	68
81	42	53	70
90	45	55	71

4) Delay

Delay is defined as contrast between expected time and actual time for data gathering in WSN. The mathematical formula for measuring the delay is expressed as,

$$\text{Delay} = \text{Expected time} - \text{actual time}$$

Table 2: Tabulation for Delay

Number of data packets	Delay (ms)		
	DTCA	DDGA	SFEB
9	13	18	25
18	16	20	28
27	18	22	33
36	21	25	35
45	23	29	38
54	25	32	42
63	28	35	44
72	32	38	47
81	35	42	50
90	38	46	53

5) Classification Accuracy

Rate of no. of data packets, with the overall number of data packets are Classification increases the rate of CA. The appropriate mathematical equation for CA is given below.

$$CA = \frac{Number\ of\ data\ packets\ are\ classified}{Number\ of data\ packets} * 100 \quad (5)$$

From (5), CA indicates classification accuracy which is considered in percentage (%).

Table 3: Tabulation for Classification Accuracy

Number of data packets	Classification Accuracy		
	DTCA	DDGA	SFEB
9	88	78	56
18	89	80	62
27	90	82	65
36	91	83	67
45	92	84	69
54	93	85	72
63	94	86	73
72	95	87	75
81	96	88	77
90	97	89	80

6) False Positive Rate

FPR counts because the no. of data packets is classify incorrectly with the amount of total data packets. It is defined as follows:

$$FPR = \frac{Number\ of\ data\ packets\ are\ incorrectly\ classified}{Number\ of\ data\ packets} * 100 (6)$$

Table 3 narrates simulation results of FPR versus amount of data packets.

Table 3: Tabulation for False Positive Rate

Number of data packets	False Positive Rate		
	DTCA	DDGA	SFEB
9	21	33	53
18	23	34	54
27	25	36	55
36	28	38	56
45	30	40	58
54	31	42	60
63	33	43	61
72	35	45	62
81	38	47	64
90	40	49	65

7) Classification Time

CT is defined as an amount of time required to classify gathered data packet. It is measured milli seconds (ms) and expressed as follows,

$$CT = Number\ of\ data\ packets * time\ (classify\ the\ data\ packets) \quad (7)$$

From (7), CT represents a classification time.

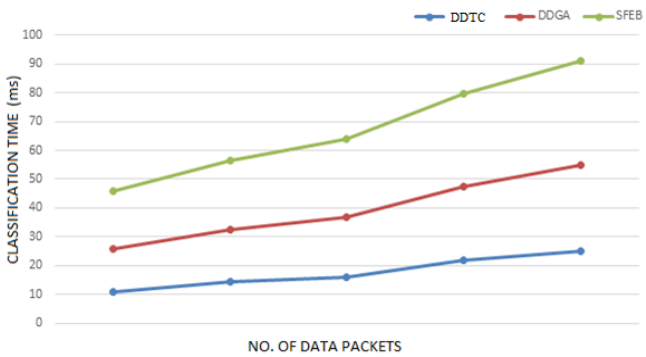

Fig.8: Simulation result of Classification Time

The base classifiers are combined based on weight value to classify the data packets. Let us consider, the no. of i/p data packet is 9, the CT of ABNRTC technique is 11ms whereas 15ms, 18ms,20ms and 22ms CT of existing [1] and [2], [3] and [4] respectively. Therefore, the ABNRTC technique takes minimum time for classification.

As a result, CT is significantly decreased by 16%, 23%, 30% and 36% as compared to existing [1] and [2], [3] and [4] respectively.

IV. CONCLUSIONS

In this paper, the technique of DTCA was performed with the aim of decreasing the energy required for data collection in WSN. There are three types of methods used to improve the data collection capability. DTCA technique uses the mobile sink node to assemble data from sensor nodes. This minimized the need for power and enlarged the ability to assemble data. This paper, then utilized decision tree classifier was magnified by Adaboost method to collect the data. This Adaboost algorithm generates an excellent, very robust classifier and classifies the data using it. This makes it possible to collect data very quickly and accurately, increasing efficiency. Finally, we have found some Parameters and have simulated some Parameters.

REFERENCES

1. Yongmin Zhang, Shibo He and Jiming Chen, "Near Optimal Data Gathering in Rechargeable Sensor Networks with a Mobile Sink", IEEE Transactions on Mobile Computing, Volume 16, Issue 6, June 2017, Pages 1718-1729
2. Cong Wang, SongtaoGuo, and Yuanyuan Yang, "An Optimization Framework for Mobile Data Collection in Energy-Harvesting Wireless Sensor Networks", IEEE Transactions on Mobile Computing, Volume 15, Issue 12, December 2016, Pages 2969-2986
3. Chih-Min Chao and Tzu-YingHsiao, "Design of structure-free and energy-balanced data aggregation in wireless sensor networks", Journal of Network and Computer Applications, Elsevier, Volume 37, January 2014, Pages 229-239
4. Jin-Tao Meng, Jian-Rui Yuan, Sheng-Zhong Feng, Yan-Jie Wei, "An Energy Efficient Clustering Scheme for Data Aggregation in Wireless Sensor Networks", Journal of Computer Science and Technology, Springer, Volume 28, Issue 3, 2013, Pages 564–57
5. Shusen Yang, Usman Adeel, Yad Tahir, and Julie A. McCann, "Practical Opportunistic Data Collection in Wireless Sensor Networks with Mobile Sinks", IEEE Transactions on Mobile Computing, Volume 16, Issue 5, May 2017, Pages 1420 – 1433
6. Chuan Zhu , Shuai Wu , Guangjie Han, Lei Shu , Hongyi Wu, "A Tree-Cluster-Based Data-Gathering Algorithm for Industrial WSNs With a Mobile Sink", IEEE Access , Volume 3.
7. Ping Zhong, Ya-Ting Li, Wei-RongLiu ,Gui-Hua Duan, Ying-Wen Chen, and Neal Xiong, "Joint Mobile Data Collection and Wireless Energy Transfer in Wireless Rechargeable Sensor Networks", Sensors, Volume 17, Issue 8, 2017, Pages 1-23
8. PrabhuduttaMohanty andManasRanjanKabat, "Energy efficient structure-free data aggregation and delivery in WSN", Egyptian Informatics Journal, Elsevier, Volume 17, Issue 3, 2016, Pages 273-284
9. Soobin Lee and Howon Lee, "Energy-Efficient Data Gathering Scheme Based on Broadcast Transmissions in Wireless Sensor Networks", the ScientificWorld Journal, Hindawi Publishing Corporation, Volume 2013, July 2013, Pages 1-7
10. Shouling Ji, Raheem Beyah, and ZhipengCai, "Snapshot and Continuous Data Collection in Probabilistic Wireless Sensor Networks", IEEE Transactions on Mobile Computing, Volume 13, Issue 3, 2014, Pages 626-637
11. Ngoc-TuNguyena, Bing-Hong Liu, Van-Trung Pham, Yi-Sheng Luo, "On maximizing the lifetime for data aggregation in wireless sensor networks using virtual data aggregation trees", Computer Networks, Elsevier, Volume 105, 2016, Pages 99-110
12. Liang He, Jianping Pan, and Jingdong Xu, "A Progressive Approach to Reducing Data Collection Latency in Wireless Sensor Networks with Mobile Elements", IEEE Transactions on Mobile Computing, Volume 12, Issue 7, 2013, Pages 1308 – 1320
13. Soo-HoonMoona, Sunju Park, Seung-jae Han, "Energy efficient data collection in sink-centric wireless sensor networks: A cluster-ring approach", Computer Communications, Elsevier, Volume 101, 2017, Pages 12-25
14. Ilkyu Ha, MamurjonDjuraev, and ByoungchulAhn, "An Energy-Efficient Data Collection Method for Wireless Multimedia Sensor Networks", International Journal of Distributed Sensor Networks, Hindawi Publishing Corporation, Volume 2014, September 2014, Pages 1-8
15. Liangshan Jiang, Anfeng Liu, Yanling Hu, Zhigang Chen, "Lifetime maximization through dynamic ring-based routing scheme for correlated data collecting in WSNs", Computers & Electrical Engineering, Elsevier, Volume 41, 2015, Pages 191-215
16. Leandro Aparecido Villas, AzzedineBoukerche, HeitorSoares Ramos, Horacio A.B. Fernandes de Oliveira, Regina Borges de Araujo, and Antonio Alfredo Ferreira Loureiro, " DRINA: A Lightweight and Reliable Routing Approach for In-Network Aggregation in Wireless Sensor Networks", IEEE Transactions on Computers Volume 62, Issue 4, April 2013, Pages 676–689
17. MouloodHeidari, Akbar Nooriemamzade, Hamid Reza Naji, "Effect of using mobile sink and clustering on energy consumption in wireless sensor networks", Turkish Journal of Electrical Engineering & Computer Sciences, 2015, Pages 1-16
18. Weigang Wu, Jiannong Cao, HejunWu ,Jingjing Li, "Robust and dynamic data aggregation in wireless sensor networks: A cross-layer approach", Computer Networks , Elsevier, Volume 57, 2013, Pages 3929–3940
19. MihaelaMitici, Jasper Goseling, Maurits de Graaf, Richard J. Boucherie, "Energy-efficient data collection in wireless sensor networks with time constraints", Performance Evaluation, Elsevier, Volume 102, August 2016, Pages 34–52
20. Yu Lasheng, Li Jie , Liu Renjie, "An effective data collection algorithm for wireless sensor network", Computing, Springer, Volume 95, Issue 9, 2013, Pages 723–738
21. Wei Wang, Dan Wang , Yu Jiang, "Energy efficient distributed compressed data gathering for senso