

Modelling an Adaptive Cluster Head Positioning Based Map Reducing Strategy for Data Transmission in Medical IoT



Gameil S. H. Ali, A. Nithya

Abstract- Internet-of-things (IoT) based health monitoring systems have turned out to be an interesting topic to enhance quality of health care services. Moreover, there is no advanced IoT based continuous monitoring of glucose systems in real time and some prevailing techniques have numerous limitations. Here, a continuous and invasive glucose monitoring system for transmitting the condition of individuals simultaneously utilizing IoT is modelled and a general system architectural design for processing back end systems to provide body temperature, real time glucose and contextual data in human readable and graphical forms to the physicians or patients is anticipated. As well, a protocol designs for monitoring the continuous data from IoT devices in order to overcome the short comings of existing methods is provided. The design of an energy efficient routing algorithm is a hot topic in the research of IoT based Data mining. Cluster Heads (CH) form backbone of inter-cluster communication. The selection of reliable and efficient cluster head is another important issue. In most of the clustering process, failure of CH occurs due to energy depletion and if the distance between sink and CH is more, it ultimately leads to failure in transmission. During transmission, nodes may fail that means sudden energy loss or node gets out of coverage. Due to relaying high data traffic, some of nodes quickly exhaust their energy and increase the risk of node failure. As a baseline to node failure, data packet loss also occurs in a CH due to congestion and poor link quality. Hence, one of the most crucial feature in designing a protocol is to minimize energy consumption for betterment of network functioning. Here, a clustering routing protocol based on data mining techniques is applied for sensor nodes in medical field called Adaptive Positioning of Cluster Head based Map reducing (APCH-MR) is proposed. Routing table based code book is generated for privacy concern, in which the process of mapping and reducing the data for dissemination is performed. The simulated outcomes depicts that the total number of packets transmitted in round 500 is 11200, total number of dead nodes during round 500 is 58, and time consumed by nodes at 500 rounds is 0.3751s respectively. The proposed method shows better trade off in contrast to conventional techniques.

Keywords - IoT, sensor data, Adaptive Positioning of Cluster Head based Map reducing protocol, data traffic, energy consumption.

I. INTRODUCTION

Internet of Things (IoT) is considered as a dynamic network where virtual and physical objects are inter-connected to one another.

Revised Manuscript Received on October 30, 2019.

* Correspondence Author

Gameil S. H. Ali*, Research scholar, Department of Computer Science, Rathnavel Subramaniam College of Arts & Science, Coimbatore, (Tamil Nadu), India. E-mail: Gameilsaad01@gmail.com

Dr. A.NITHYA, Associate Professor, Department of Computer Science, Rathnavel Subramaniam College of Arts & Science, Coimbatore, (Tamil Nadu), India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

IoT encompassing advanced technologies like data mining, wireless sensor networks (WSN), cloud computing, artificial intelligence plays a significant role in numerous domains including logistics, robotics, transportation and health-care monitoring [1]. For example, IoT based systems for health care monitoring comprises of WSN, sensing, smart gateways and cloud for data storage which offers a significant way to sense and monitor real time e-health. Innovations and advancements in WSN have made an innovative platform for wellness and e-health application development [2]. Ambient intelligence, ambient assisted living and smart homes are now becoming an interesting area of research. These are merged with other health solutions like wellness and fitness, chronic disease management, nutrition or diet monitoring applications. The novel initiatives are liable to be integrated with patient information ecosystem indeed of separated into decision processes and monitoring. There are certain advancements in ageing population, where older people can be treated and monitored with comfort level even from their houses. Fully automated health care monitoring wireless system comprises of numerous useful applications. One amongst those applications is glucose level measurement for diabetic patients [3]. Diabetes tends to be a major health care concern for numerous individuals in recent times. In accordance to the report provided by WHO, number of people with diabetes has been raised to 422 million and in late 2016, about 1.5 million people died due to diabetes. WHO categorized diabetes as a top ten mortality causes. Diabetes leads to be serious effects on health care of individuals and society [4]. Regrettably, there is still no permanent cure for diabetes known. Moreover, one solution of this crisis is to measure blood glucose levels continuously and close the loop with appropriate delivery of insulin. Statistics provided by Prospective Diabetes Group illustrate that continuous glucose monitoring can diminish long term complications between 40% and 75 %. Therefore, continuous glucose monitoring equipped with alarm systems can assist patients to take corrective measures like diet decisions, physical exercise and when to consume medications. Energy harvesters integrated with wearable devices facilitate powering wireless sensor operated applications, which leads to operate it continuously [5]. This regime has numerous implications on health care providers and patients, specifically for sensors that are implanted, while battery changing process could leads to discomfort and pain. Cautious design of both efficient energy harvesting strategy and low-power electronic circuitry is pivotal to fully autonomous wearable systems.

Modelling an Adaptive Cluster Head Positioning Based Map Reducing Strategy for Data Transmission in Medical IoT

In this investigation, experimentation attempts to analyze feasibility of secure and invasive continuous glucose monitoring systems using IoT based system and data mining approaches. Initially, the work starts with the system framework of IoT based system architecture from sensor devices to back end systems for processing real time body temperature, glucose and contextual data (i.e. environmental temperature) in human readable text forms and graphical forms to end users like doctors and patients. However, an effectual routing protocol termed Adaptive Positioning of Cluster Head based Map reducing (APCH-MR) protocol for continuous monitoring glucose level using IoT and code book model is designed. Data received by cluster head will be disseminated using code book approach in order to reduce the enormous data flow amongst the network and to reduce energy consumption due to huge data transmission. Thereby, higher energy efficiency is achieved, as well, energy consumption of sensor devices and energy harvesting units are measured using this approach. At last, data transmission is made amongst the connected nodes in the network for generating notification to doctors and patients during abnormal circumstances like too high glucose level or too low glucose level. The significant contributions of this paper are given below:

- 1) To design an effectual Adaptive Positioning of Cluster Head based Map reducing (APCH-MR) protocol for continuous monitoring glucose level using IoT and data mining approaches.
- 2) Utilization of APCH-MR based code model for reduction of continuous data in order to enhance storage space and reduced energy consumption by sensor nodes.
- 3) Routing table based code book is generated for privacy concern, in which the process of mapping and reducing the data for dissemination is performed.
- 4) Performance monitoring of network nodes, in terms of network lifetime, throughput, propagation rate, traffic analysis and so on is anticipated.

The reminder of the work is structured as follows: in section II related works in association with IoT and data mining approaches are discussed. Section III presented the continuous monitoring of glucose level using code book model and IoT based system architecture is discussed in detail. In section IV, simulation and the experimental results based on proposed protocol design and code book model is evaluated and measured. In section V, conclusion and the future dimensions of this investigation is discussed.

II. RELATED WORKS

Many research applications in glucose monitoring are not supported by IoT-based models. Correspondingly, specialists or patients can't screen glucose level of patient remotely. Murakami et al. present a CGM framework for heart patients in the emergency unit. The framework is worked by disposable subcutaneous glucose sensor, glucose client, and server. The framework gathers glucose information four times each day and stores in an emergency clinic data framework. Specialists can utilize the bedside screen to screen the glucose data.

In [6], Deepti Sisodia et al, gives systematic efforts in planning a framework which results in disease prediction such as diabetes. During this work, three machine learning classification algorithms are considered and assessed on different measures. Investigations are performed on Pima Indians Diabetes Database. Test results decide the sufficiency of designed framework with an accomplished accuracy of 76.30 % utilizing Naive Bayes classification calculation. In future, the designed framework with machine learning classification algorithms can be utilized to anticipate or analyze different diseases.

In [7], Hsin-Yi Tsao et al, anticipated diabetic retinopathy, initially demographical factors, laboratory test results, family diabetes history, and exercise habits from patients. At that point, diverse machine learning algorithms are used to accomplish both accuracy prediction and distinguish novel risk factors. Experimental outcomes show that support vector machines accomplished the best execution with 79.5% and 0.839 in accuracy and AUC, correspondingly. Decision trees and logistic regressions both select utilization of insulin and diabetes duration as the most discriminative features to anticipate diabetic retinopathy. In [8], N. Vijayalakshmi, [2017] made an attempt to utilize data mining as a tool for investigating clinical information records of both diabetic and non-diabetic patients. Data mining is an integral tool which is utilized for extricating huge data from historical data. This data can be utilized for further prediction and decision making. WEKA was utilized for applying different data mining procedures like Statistical analysis, Associative rule mining, and Clustering, Classification and subset assessment. This has been useful in extracting key data with respect to diabetics. Two characterization techniques were utilized on a similar record set to deliver practically comparative outcomes at different dimensions of precision. Among them, J48 pruned tree has been observed to be progressively accurate. Clustering is likewise completed to check yield of previous strategies. From the gained information, an attempt is made to construct decision tree model for diabetic's prediction. Accuracy of prediction is 81%. In [9], Hoa Hong Nguyen, [2017] sketched out medical problems that conventional human services models are looking in our ageing society, including the increase in chronic diseases and rise in hospital and clinical services costs. To diminish weight on emergency clinic frameworks and medicinal services providers, enhance the nature of consideration and decrease social insurance costs, successful and effective therapeutic frameworks should be created. Remote healthcare monitoring frameworks dependent on IoT technology has huge potential. This paper reviewed about IoT and exhibited IoTTA architecture in detail. In [10], Ioannis Kavakiotis, [2017] deliberate systematic exertion was made to distinguish and review machine learning and data mining approaches connected on DM research. DM is quickly developed as one of greatest global health challenges of 21st century. To date, there is a critical work completed in practically all parts of DM research and particularly biomarker identification and prediction diagnosis.

The coming of biotechnology, with the tremendous measure of data delivered, along with expanding measure of EHRs relied upon exploration to assist top to bottom investigation toward determination, etiopathophysiology and treatment of DM through work of machine learning and data mining systems in advanced datasets that incorporate clinical and biological data. In [11], Paul S. Fisher, [2017] posses a capacity to follow the outcomes after substation is made, or even better, capacity to track proposed changes to decrease the occurrence of false positives. Paul et al. trust that the model considers both past and future accomplishment of most trust in actual predictions. Paul et al. additionally think about different perspectives to broaden our examination, which will incorporate design of IoT-based healthcare frameworks, how to gather information, how to maximize network availability and security. In [12], Yukai Li, [2018] utilized diabetic patient health management follow-up information. *Yukai et al.* have joined feature selection and imbalanced processing procedures, and few researchers have used health management control satisfaction of patients with diabetes for classification predictions. In this work, *Yukai et al.* offered evidence that Adaboost calculation can be effectively utilized for health management control satisfaction of patients with diabetes. In [13], Ali et al. [2011] propose a Bluetooth low energy (BLE) implantable glucose monitoring framework. Glucose information gathered from the framework is transmitted through BLE to PDA (advanced mobile phone, or Ipad) which signifies to receive data in text forms for visualization. The framework demonstrates few accomplishments in lessening power utilization of an outer power unit and an implantable unit. In [14], Lucisano et al. [2016] present a glucose monitoring in people with diabetes utilizing long term embedded sensor framework and model. Glucose information is transmitted for every two minutes to outside collectors. The framework demonstrates its capacity of constant long term glucose monitoring. Likewise, the framework demonstrates that embedded sensors can be put inside human body for long period time (for example 180 days) for managing diabetes and different sicknesses. In [15], Menon et al. [2013] propose a non-invasive blood glucose monitoring framework utilizing near infrared (NIR). Glucose level in blood is monitored dependent on variety of received signal intensity attained from NIR sensor. The anticipated glucose information is sent remotely to remote computer for visualization. As of late, some IoT-based applications for glucose monitoring have been constructed. Moreover, those frameworks don't attentively consider energy efficiency of sensor nodes and communication between sensor devices and gateway. In [16], Rasyid et al. [2016] propose blood glucose level monitoring framework dependent on wireless body area network for distinguishing diabetes. The framework is worked by utilizing a glucometer sensor, Arduino Uno, and Zigbee module. Specialist and guardian can access to web page to screen glucose dimensions of patient remotely. Moreover, framework isn't energy effective because of high power utilization of Arduino Uno board and Zigbee module. In [17], Wang et al. [2012] present monitoring framework for types 2 diabetes mellitus. The framework can capable to make decision on the statuses of diabetes control and predict future glucose of an individual. Obtained glucose information can be observed remotely by medical staffs by means of wide area systems. In [23], FAlsubaei, A Abuhusseini, [2017] distinguish three

advancements that are probably going to shape future of data mining in IoT as follows.

1. Expanding computational power at edge: As cloud is a fundamental source of computation in most of framework designs, the regularly expanding power of small computational devices, for example, smart phones is set to adjust the dependence on cloud computing and edge computing in anticipated future. Therefore, edge investigation will flourish, and another type of distributed data stream mining algorithms will be produced to serve the assortment of IoT applications.
2. The expansion in communication abilities with 5G advancement: This would empower new framework models that bring both cloud and edge computing cooperating rather consistently. This, thusly, will empower data and models to move at incredibly high speeds among cloud and edge devices.
3. Enhancement in battery technologies: The improvement of enduring batteries of IoT devices along lightweight techniques will results in achieving equilibrium where these days power intense calculation turns out to be significantly more energy efficient procedures. This will thus improve quality of experience in utilizing IoT applications.

In spite of the fact, that these frameworks demonstrate their preferences in continuous glucose monitoring, there are numerous constraints. For instance, a few frameworks don't consider constant and remote monitoring while different frameworks don't focus on energy efficiency [18]. The primary motivation of the paper is to give progressed IoT-based framework for real time and remote continuous monitoring if glucose, logical information, and body temperature. The difference between this work and the existing work is the design of routing protocol using map reduction-based code book. The proposed method shows better trade off than the existing techniques.

III. PROPOSED SYSTEM

a. Sensor device structure

This section provides a detailed explanation of transmitting data which is attained from the diabetic patient to the remote location using sensors and data mining approaches known as Adaptive Positioning of Cluster Head based Map reducing (APCH-MR) based code book. Various parameters are computed in which transmission rate, data propagation, delays are the significant factors. IoT based sensors general structures are given in figure 1. It comprises of basic components such as microcontrollers, sensors, wireless communication blocks, management components and energy harvesting unit [26]. Initially, microcontrollers play significant tasks such as data transmission and acquisition. Henceforth, it utilizes huge amount of devices' total power consumption. Diminishing the power consumed by microcontrollers can save huge amount of power consumed by the device [27]. Ultra low power micro-controllers are competent of operating in sleep modes which is an appropriate candidate for target.

Modelling an Adaptive Cluster Head Positioning Based Map Reducing Strategy for Data Transmission in Medical IoT

With this device, micro-controller attains glucose data from an implantable glucose sensor through wireless inductive link receiver, where it body temperature and environmental data through data link wires like UART, SPI or I2C. Generally, SPI is preferable due to its lowest power consumption amongst these interfaces.

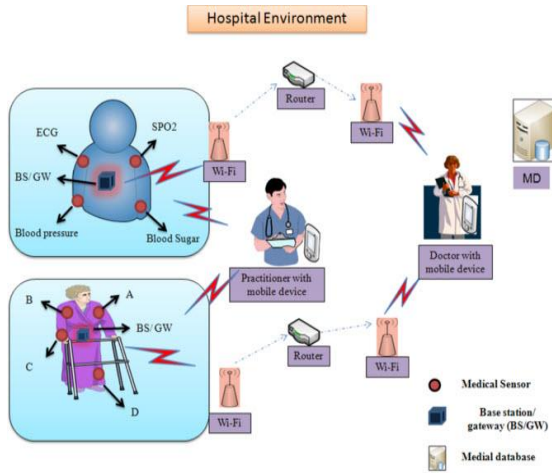


Figure 1: General IoT framework

In general, wireless communication block is responsible for broadcasting data from micro-controllers to gateway which is equipped with transceivers [19]. The block comprises of RF transceiver IC for 2.4 GHz ISM band and an embedded antenna. Based on the transmission of 2 Mbps data, radio frequency completely fulfils the need of transmission data rates in CGM system. Transmission data rates can be configured for attaining certain levels of energy efficiency. For instance, instead of utilizing 2Mbps, data rate of 256 kbps can be utilized for saving power when transmitting temperature, glucose and contextual data [20]. As well, radio frequency is competent of both long and short transmission range from few centi-meters to hundreds of meters. Based on significant applications, transmission power and transmission range can be configured [21]. For short communication range, radio frequency consumes very lesser energy. In sensor node, power management node and energy harvesting unit described are the two foremost significant components as they have direct impact on energy consumption and operating functionality of sensor node duration [22]. Nodes have to transmit the data to the remote location effectively. In order to transmit the data nodes have to be clustered to eliminate the unwanted data from source to destination. Thus, nodes are placed in a proper location and a cluster head is to be elected for optimal utilization of clusters.

b. Adaptive positioning of optimal cluster head

In this segment, optimal number of cluster estimation is performed in such a way that it has to diminish energy consumption while providing higher degree of connectivity for transmitting the glucose level to doctors [24]. Optimal amount of clustering is based on overall nodes in the network, network area and transmission range. During the process of clustering, there are certain nodes that have not joined in any clusters or groups. Those nodes are termed as slack nodes. This may occur due to elected cluster head (CH) or exposure of employment nodes [25].

Therefore, probabilities of occurrence of those nodes have to be attained initially. For this purpose, mapping of sensor field to non-overlapping cluster circle with the radius 'r' and consider an assumption that nodes outside the cluster circle boundary are secondary nodes, whereas the remaining nodes in the cluster are cluster members. Secondary node communicates or transmits data to cluster head with the assistance of cluster member inside the cluster and turns as multi-hop member of cluster head. Square field $\frac{K^2}{(2r)^2}$ is packed with non-overlapping circle of radius r. Therefore, probability of multi-hop member (MHM) is provided in Equation 1:

$$MHM = \frac{K^2}{(2r)^2} * \frac{(2r)^2 - \pi r^2}{\pi^2} \approx 0.2 \quad [1]$$

Energy consumption of every node within the cluster member is provided in Equation given below [2] & [3]:

$$Cluster_{member} = S_{dp} (cluster_{elected} + \epsilon_{amplifier} r^2 (1 + MHM_p)) \quad [2]$$

Where;

$cluster_{elected}$ Specifies energy utilized by electronic circuitry in modulation, coding and spreading signal.

$$Cluster_{member} = S_{dp} (cluster_{elected} + \epsilon_{amplifier} \frac{K^2 (1 + S_{dp})}{2\pi E}) \quad [3]$$

Where;

$\epsilon_{amplifier} r^2$ represents energy consumed for amplification of signal over short distance r. S_{dp} represents data packet size The equation provided above can be modified by measuring the circle area [4]:

Consider that, a sensor field is measured by circle radius R and the sink nodes will be placed in the middle of circle. Based on this assumption, the packets are transmitted from cluster head to sink nodes. As cluster head will not broaden its transmission range, the radius 'r' will be similar as that of the member node. This model assists to route packets from cluster head to sink node.

In the proposed positioning of cluster head, the cluster distance is partitioned into numerous concentric rings with distance of radius. Hence, energy consumed for relaying packets from outside ring to inside ring is provided below:

$$Energy_{inout} = S_{dp} (2cluster_{elected} + \epsilon_{amplifier} r^2) \quad [4]$$

The total number of hops needed to transmit packets from cluster head to sink node is computed using the equation given below [5]:

$$TNH = \frac{R}{r} (1 - P_{hops}) \quad [5]$$

Where;

P_{hops} specifies probability of distance based on hops to sink. It is computed in accordance to equation of node distribution in cluster ring [6].

$$P_{hops} = \frac{r}{R} \sum_{i=1}^{R/r} \frac{R^2 - (ir)^2}{K^2} \quad [6]$$

Packets which are placed far away from the sink nodes are relayed with the assistance of intermediate nodes. Assume N_s as number of neighbours of source nodes S. Therefore, amount of energy consumed by neighborhood electrical circuitry during transmission of data packets by source is given as $N_s * C_{elected}$ as in Equation [7]:

$$N_s = n \frac{\pi r^2}{k^2} \quad [7]$$

Energy utilization while forwarding data from cluster head to sink node is given below [8]:

$$Cluster_{CH_{sink}} = S_{dp}(N_s C_{elected} + C_{elected} + 2(C_{elected} + \epsilon_{amplifier} r^2 + N_s C_{elected}) N_{hop} \quad [8]$$

Now, total energy utilized by network is provided in equation below [9]:

$$C_{totalenergy} = ((n + nN_s) Cluster_{elected} + E(2cluster_{elected} + \epsilon_{amplifier} r^2 + N_s Cluster_{elected}) N_{hop} + n \epsilon_{amplifier} \frac{k^2(1+S_{dp})}{2\pi k} \quad [9]$$

if $r < R$, the optimal value of E can be found by taking derivative of equation 10 with respect to E and then equalizing to zero

In case, if $r < R$, optimal value of energy can be attained by considering derivation and equalizing to zero:

$$E_{opt} \approx \sqrt{\frac{n(1+S_{dp})}{(2\pi(1 + \frac{2cluster_{elected}}{\epsilon_{amplifier} r^2} + \frac{N_s C_{elected}}{\epsilon_{amplifier} r^2}) N_{hop}}} * \frac{k}{r} \quad [10]$$

Optimization value significantly depends on transmission range of data packets. Henceforth, for longer range transmission E_{opt} is lesser. Energy utilization for optimal cluster selection is to compute the energy level. Initially yje node is provided with energy C_0 which is considered as Energy constrained device and the remaining cluster member nodes' energy is considered as higher. Therefore, it is called as energy sourced device as in Equation [11].

Consider,

Total amount of nodes $\rightarrow n$;

Total amount of energy sourced nodes $\rightarrow m$;

$\gamma, \gamma_1, \gamma_2 \dots$ as extra energies of 'm' nodes

$$m < \frac{n}{2} \text{ or } m > \frac{n}{2} \text{ for } m \leq n \quad [11]$$

Then, the nodes which makes the network heterogeneous can be provided as follows [12] & [13]:

$$m_\gamma = 1 - \left| \frac{n-2m}{n} \right| \quad [12]$$

$$\text{if } \gamma_i \neq \gamma_j; \text{ where } i \neq j \quad [13]$$

The value of 'γ' increases by increasing the value of 'm', therefore, when γ is a variable, the fraction is equal to m/n.

Consider, β as energy factor in which energy sourced nodes have greater nodes than remaining nodes in network. The value can be provided as in Equation given below [14]:

$$\gamma = \frac{1}{mC_0} \sum_{i=1}^m (C_i - C_0) \quad [14]$$

The simplified form of equation for constant can be given as below. For constant computation, the Equation can be simplified as in Equation [15];

$$\gamma = \frac{m(C_i - C_0)}{C_0} = \frac{C_i}{C_0} - 1 \quad [15]$$

Hence, the heterogeneity attained due to optimal energy consumption or energy gain by the nodes using energy sourced nodes in the network is computed effectively.

c. Design modelling

Main Objective is the transformation of sensitive data and is achieved by quantizing the data. Vector Quantization [23] is performed based on the following steps:

1. Generate codebook from the data to be transmitted.
2. Encode the sensitive data by mapping code book to the nearest node in network.
3. Using routing table, data transmitted to destination will be reconstructed by map reducing the codebook index.

The significant objective of this section is transformation of the data from IoT devices by quantizing the sensitive diabetic data of individuals. Quantization of sensitive data can be performed with certain steps. The construction of codebook is nothing but the functionality of routing table based on the proposed techniques. The construction and reconstruction of node placement and termination has to be described in the code book. Based on this cluster head selection and node establishment and node termination is identified.

Privacy preserving in medical data mining significantly needs two primary steps, thus this work anticipates to reconstruct data back, therefore the reconstructed data should look alike the original data. But it should not reveal the original one. Transformed data set specifies approximating data but not exactly as the original data. Henceforth, privacy is preserved.

Modelling an Adaptive Cluster Head Positioning Based Map Reducing Strategy for Data Transmission in Medical IoT

Table I: Original Data (Glucose level, Temperature)

| | | | | |
|-----|-----|-----|-----|-----|
| 70 | 100 | 70 | 90 | 140 |
| 60 | 150 | 90 | 100 | 370 |
| 270 | 550 | 610 | 200 | 150 |
| 75 | 60 | 45 | 56 | 76 |

Table II: Routing Index table

| | |
|----|----|
| 7 | 9 |
| 11 | 12 |

Table III: Code Book generation based on routing index table

| Indices | Code book | | | | |
|---------|-----------|-----|-----|-----|-----|
| 1 | 20 | 30 | 590 | 190 | 50 |
| 2 | 800 | 270 | 110 | 340 | 450 |
| 3 | 40 | 110 | 100 | 20 | 80 |
| 4 | 650 | 230 | 700 | 160 | 150 |
| 5 | 880 | 120 | 200 | 180 | 400 |
| 6 | 30 | 220 | 150 | 200 | 240 |
| 7 | 90 | 60 | 90 | 90 | 90 |
| 8 | 70 | 70 | 20 | 30 | 50 |
| 9 | 60 | 170 | 90 | 100 | 390 |
| 10 | 170 | 390 | 500 | 190 | 450 |
| 11 | 250 | 500 | 610 | 190 | 120 |
| 12 | 750 | 560 | 430 | 540 | 740 |

The above given table I shows the original data from the patient, and it is given in the matrix format, i.e. 4 *5 matrix. The table contains original data, i.e. 4 column and 5 rows. With original data, codebook will be developed. Codebook is utilized to carry out encoding and decoding, that is, reconstruction process of original data. Consider the first row (70, 100, 70, 90 and 140) which original data of an individual, now quantizer will look into its best

approximating row in codebook (routing table) and found that approximating row in codebook (routing table) is present in index 7. Index table specifies original data indices of original data and not the original data values. Subsequently, reconstruction is performed using index table, at last (90, 60, 90, 90, 90) is produced in the place of (70, 100, 70, 90,140)

Construction of routing table based on codebook is done with vector quantization. This can be applied in any medical datasets. This work significantly concentrates on overall quality of clusters generated after transforming data from dataset and to evaluate the distortion produced in the dataset.

Experimentation is carried out using the following steps:

- 1) Medical dataset is pre-processed with missing data during transmission process.
- 2) So, missing value has to be replaced with average value over the entire dataset.
- 3) Vector quantization is applied to transform the original data.
- 4) Map reducing is introduced in the generation of codebook. Mapping the index is performed and reducing the data set is performed next.

In order to propose that how the clusters are close to each other in transmitting data, that matches with the cluster in original dataset. Quality of placing the node as cluster member is computed with the distortion of original data.

d. Quantizing original data

In this work, routing table-based codebook is constructed from training data and then reconstructed data is generated by approximating every data point to nearest value. This is due to the fact that original data should not be revealed once transmission is performed; transformed data is only the approximated data and not the original data. Therefore, privacy is preserved.

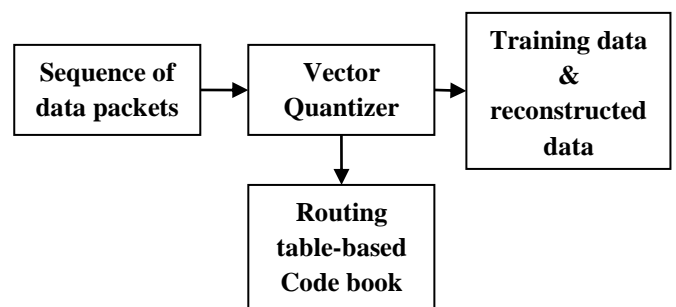


Figure 2: Internal processing of proposed work

Algorithm 1:

Step 1: Consider training sequences from real world data, i.e., data from diabetic individual from UCI repository data set have been considered. Thus, the experiment is conducted on data from diabetic patients.

Step 2: Generate a codebook based on routing table by considering the data from training sequence.

Step 3: The codebook comprises of code words. Code book data is reduced form of original training sequence.

Step 4: Mapping is performed on both the original data sequence and the code word data sequence.

Consider the first row, i.e. (70, 100, 70,90, and 1400) is original data, quantizer looks for best sequence in codebook and found that approximating row possesses an index 7. Routing table indices specifies indices of original data and not direct code book values. This is followed by, reconstruction process with index table, at last (80, 90, 80, 100, 130) is produced instead of (70, 100, 70, 90, 140).

This investigation considers only sensitive attributes (glucose level, body temperature and so on) and replaces the original value with new values with quantization approach. Once transformed data set is produced, data owner establishes the transformed data to data miners for getting the hidden knowledge from data store.

Table IV: Original Data (Glucose level, Temperature)

| | | | | |
|-----|-----|-----|-----|-----|
| 70 | 100 | 70 | 90 | 140 |
| 60 | 150 | 90 | 100 | 370 |
| 220 | 550 | 610 | 200 | 150 |
| 75 | 60 | 45 | 56 | 76 |

Table V: Routing Index table

| | |
|----|----|
| 7 | 9 |
| 11 | 12 |

Table VI: Code Book generation based on Quantization table

| Indices | Code book | | | | |
|---------|-----------|-----|-----|-----|-----|
| 1 | 20 | 30 | 590 | 190 | 50 |
| 2 | 800 | 270 | 110 | 340 | 450 |
| 3 | 40 | 110 | 100 | 20 | 80 |
| 4 | 650 | 230 | 700 | 160 | 150 |
| 5 | 880 | 120 | 200 | 180 | 400 |
| 6 | 30 | 220 | 150 | 200 | 240 |
| 7 | 90 | 60 | 90 | 90 | 90 |
| 8 | 70 | 70 | 20 | 30 | 50 |

| | | | | | |
|----|-----|-----|-----|-----|-----|
| 9 | 60 | 170 | 90 | 100 | 390 |
| 10 | 170 | 390 | 500 | 190 | 450 |
| 11 | 250 | 500 | 610 | 190 | 120 |
| 12 | 750 | 560 | 430 | 540 | 740 |

Table VII: Reconstructed codebook table

| | | | | |
|-----|-----|-----|-----|-----|
| 80 | 90 | 80 | 100 | 130 |
| 70 | 170 | 100 | 140 | 350 |
| 200 | 500 | 650 | 240 | 120 |
| 700 | 550 | 400 | 550 | 730 |

e. Data transformation using K-means clustering

Grouping of objects of similar kind is termed as clustering. Clustering offers a significant role in numerous data mining like: scientific data exploration, Information retrieval and text mining, Spatial database applications, Web analysis, CRM and marketing, Medical diagnostics, Computational biology and many others.

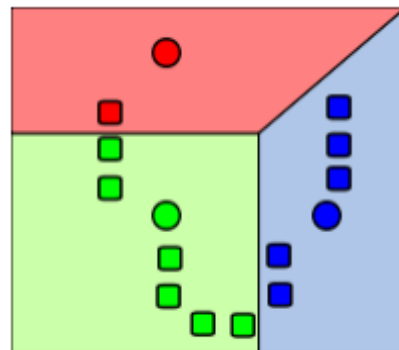


Figure 3: Clustering nodes

Figure 3 shows the samples of clustering. Here, we have considered k-means clustering for analyzing the data and broadcasting the data using it:

Algorithm:

Step1: Select the value of K, which is nothing but number of clusters.

1. Choose Value of 'K', which is number of clusters.
2. Select 'k' object in a random manner. This is turned out to be initial centroids.
3. Objects which are nearer to clusters will be allocated or forwarded to certain group (known as clusters) in which objects are closer to centroid of network cluster.
4. Re-evaluate centroid for freshly attained 'k' clusters.

Modelling an Adaptive Cluster Head Positioning Based Map Reducing Strategy for Data Transmission in Medical IoT

- Repeat step 3 & 4 till entire objects has been removed. The objective of k-means clustering is to locate objects to closer to nearest cluster with assistance of centroid condition and nearest neighbour search condition.

It is anticipated that grouping of original data and transformed data is significantly to evaluate the performance of proposed technique. The clusters generated will be analyzed and compared. The anticipated methods are flexible; any data mining approach can be utilized over transformed data and can be utilize accurate secure outcomes.

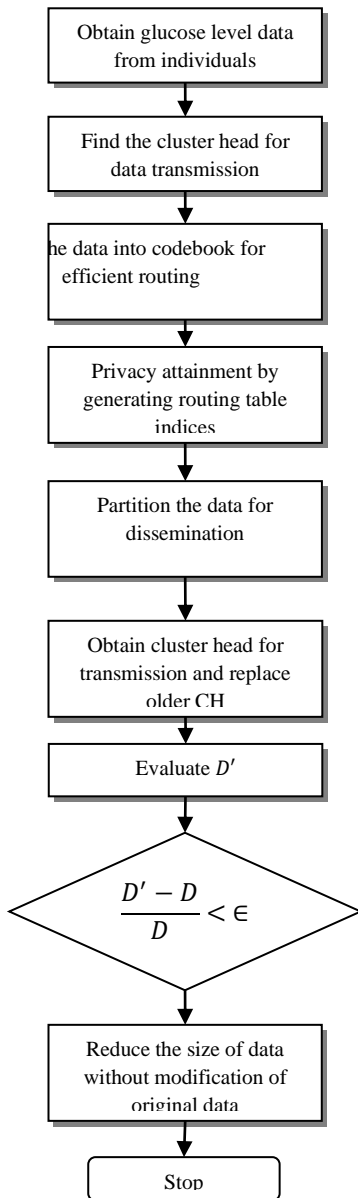


Figure 4: Flow diagram of proposed method

Construction of codebook is a significant step in Quantization process. Essential things are needed for codebook generation in the training sequence. Training sequence is attained from UCI data repositories for medical dataset.

Algorithm:

Input: Region of training data $\rightarrow R$, Training vectors $\rightarrow 'T'$

- Develop a new codebook for routing table which is considered as 'C', it is considered as the mean or centroid of training dataset.
- Partition the routing table codebook as C_n^+ and C_n^- is computed with $C_n^+ = C(1+\epsilon)$ and $C_n^- = C(1-\epsilon)$ in which $\epsilon = 0.01$ which is minimum distortion between old routing table and new routing table.
- Partition the received data from the individual diabetic partition into R1 and R2 based on the difference between training data and testing data. The training vectors which is generally closer to C_n^+ falls in routing code book R1 and training vectors that are closer to C_n^- falls under region R2.
- Routing R1 is considered as Training vector 1, i.e. TV1 and the Routing table R2 is called as training vector TV2.
- Now generate the cluster head for Training vectors TV1 and TV2. Let the new cluster head will be CHR1 and CHR2.
- Replace older cluster head by new cluster head.
- Evaluate the data transmission from the cluster head CHR1 and CHR2, and notice the difference between them.
- Repeat the process until $\frac{D' - D}{D} < \epsilon$
- Repeat step 2 to 8 till new codebook for propagating the medical data is attained. Let $N = 2^b$ specifies the code words in code book and b represents number of bits utilized for code book generation, D is difference between training sequence and old code word, D' is difference between training sequence and new code words.

In the process of vector quantization, code book plays a significant role in broadcasting data and accurate results are mined depending on the code book design, as quantization is carried out based on the assistance of code book index. The values attained are exactly the same value as that of the original value and not the original data from the individual. This is mainly performed to attain privacy of preserving the individual persons' detail.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, simulation of the proposed Adaptive Positioning of Cluster Head based Map reducing (APCH-MR) protocol using codebook generation is performed using MATLAB simulation environment. Once, the cluster head is elected, it starts receiving the data, i.e. individual diabetic data from different gateway, while the access point generates an routing table based indices with code book for maintaining the sensitive information such as glucose level, body temperature and so on.

The proposed method is effectual based on privacy concern; meanwhile the data essential for transmission is propagated to the health care person. So as to verify the quality of data broadcasted through clustered nodes, optimal positioning of nodes to receive the data is essential. Once the data obtained from sensor node to gateway, CHR1 and CHR2 is chosen based on training the data with vector quantization. The data which is quantized is propagated to remote location and were tested under diverse circumstances. In this experimentation, distance between sensor nodes and gateway is within a few meter ranges, in some cases the distance will be less than one meter. The results projects that even after mapping the data into CHR1 and CHR2, the data which is sent and data which is received is similar to one another, i.e. glucose level, environmental temperature, body temperature and so on. Also, there is no loss encountered during the process of transmission. In some scenario, radio signals are blocked, while sensor nodes attempt to transmit the data using higher power. However, average power consumption of those sensor nodes does not changes dramatically owing to long interval of transmission (transmission time exceeds ten mins). Power consumption of sensor components is shown in figure below. From the obtained output, sensor nodes consume lesser power than that of other existing techniques proposed by various investigators. Most of the prevailing energy efficient sensor nodes from continuous glucose monitoring utilize more power than 5mA, while the proposed method consumes lesser power in case of simulation level. Power consumption can be diminished when the prototype is changed for complete structure. For the purpose of testing the functionality of the designed protocol, glucose level includes high, medium and low glucose level along with temperature value which is sent from sensor node to remote location for health care personalities. The outcomes projects that data is represented and categorized accurately. As well, notification services are work more accurately in real time condition during abnormalities detection (i.e. too high glucose level or too low glucose level)

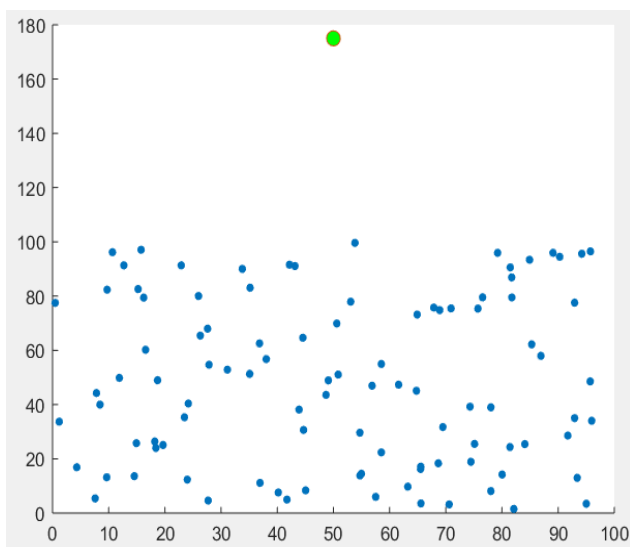


Figure 5: Node placement in network

Fig 4 illustrates the number no nodes active in the network connectivity. Nodes that are not active will be terminated from the network. The node placement will be random as the individual will show random mobility towards the region of connectivity. Node placement is described as an approach

to place nodes effectively in simulation region that successfully relates entire network that resolves the crisis of network connectivity management in nodes using code book mapping of routes. considers a cluster of nodes distribution in random -pattern in such a way that base station is located at centre surrounding with huge sensor nodes covering equal area on circle, determined by covering rest of the area on same pattern of deploying nodes on circular area so as to cover entire simulation area. Advantage of placement of node is that maximum area is covered; burden on nodes is diminished by increasing network lifetime.

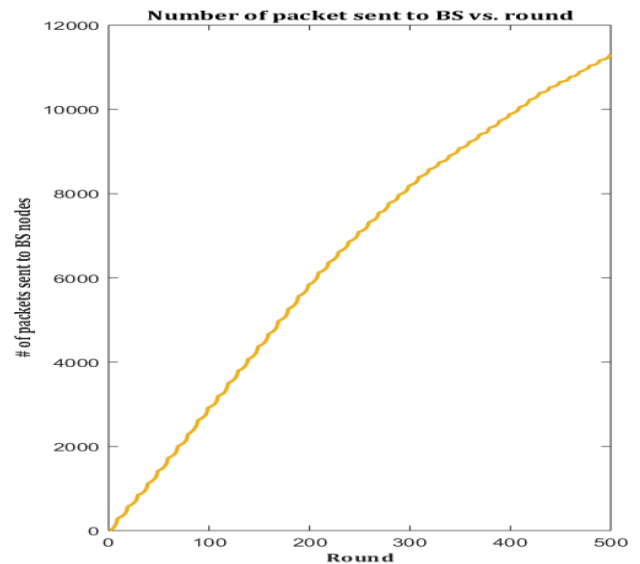


Figure 6: Number of packet sent to BS Vs number of round

Figure 5 shows the number of packets transmitted to base station and the number of rounds for transmitting the packets. The packet transmission will be higher when the round increases. Network based code booking increases transmission rate devoid of data distortion and decreases number of data transmission, and henceforth energy consumption throughout network. Moreover, based on random type of network topology, certain nodes have to deal with more traffic than others. As an outcome, their energy drains out faster during continuous data transmission, these nodes die and network become disconnected. Although this drawback is resolved by nodes connected network should increase their transmission power to preserve connectivity, this will increase overall energy consumption and decreases network lifetime.

Modelling an Adaptive Cluster Head Positioning Based Map Reducing Strategy for Data Transmission in Medical IoT

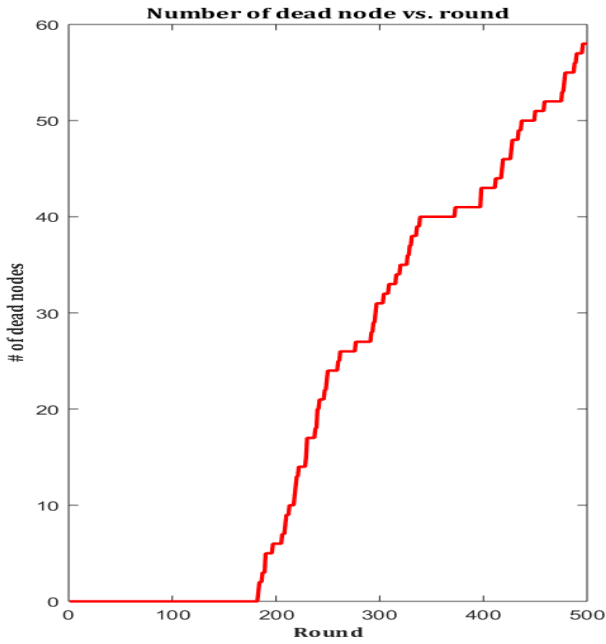


Figure 7: Number of Dead nodes Vs Number of rounds

Figure 6 shows the number of dead nodes in the network and the number of rounds for transmitting the packets. The dead nodes will be higher when the round increases. Nodes which remain idle for long time will be removed from the network as the Energy consumption will be reduced. Uneven data distribution problem and energy consumption creates discontinuities in network by optimizing positions of nodes placed in the network (i.e., distribution of first node on bridge). In this manner, total volume of data distribution is maximized that can be processed and maximize overall lifetime of network. In specific, numerical solutions calculates required number of nodes and their positioning over a network cluster given the length of the nodes connected to the network and required data throughput. The numerical and simulation results depicted that anticipated method can prolong network lifetime while at the same time it eliminates any discontinuities owing to nodes failure.

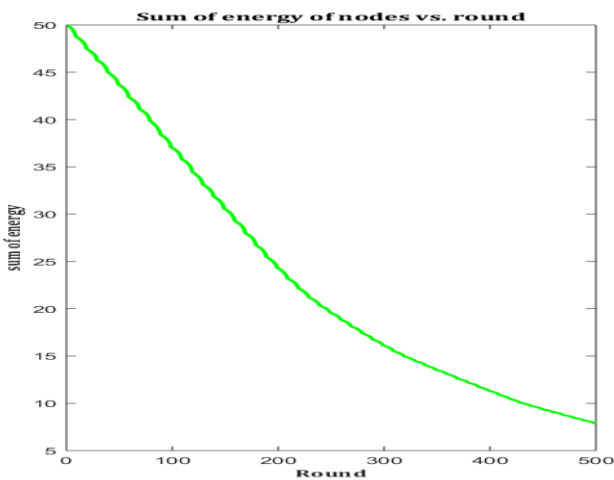


Figure 8: Sum of Energy consumption Vs Number of rounds

Figure 7 shows Energy consumed by nodes and the number of rounds for transmitting the packets. When the dead nodes are terminated from network connectivity, power consumption will be reduced. Energy consumption shows fluctuation based on number of iterations. To maximize operational life time of sensor nodes which have limited energy, nodes has to transmit their data at limited power and therefore form multi-hop network. Owing to the transmission power, consider different network connectivity cases: 1-hop, 2-hop and 3-hop connectivity and also assume omni-directional transmissions. In specific, with 1-hop case, every node sets its transmission power level, therefore only first neighbour can decode its corresponding packets.

Table VIII: Tabular representation of packet transmission in subsequent rounds

| Round | 100 | 200 | 300 | 400 | 500 |
|---------------------|------|------|------|-------|-------|
| No. of packets sent | 2200 | 6000 | 8200 | 10000 | 11200 |

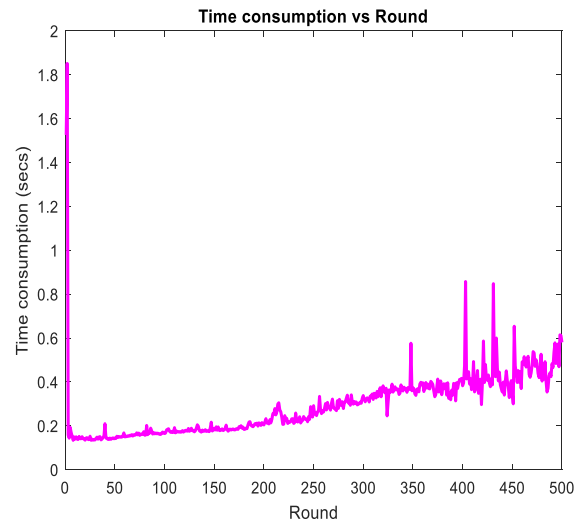


Figure 9: Time consumption Vs Number of rounds

Figure 8 shows Time consumed by nodes and the number of rounds for transmitting the packets. When the data is reduced using the code book generation with quantization technique, Energy is also reduced, whereas the rounds will be higher for continuous transmission. As nodes located at centre position of connectivity, it will relay more packets than those nodes located towards the edges, if all nodes transmit same power, nodes at centre will drain out of energy and dies rapidly and network will be disconnected. Henceforth, nodes at centre must transmit at lower power as compared to nodes towards edges. All nodes are required to transmit at same data rate; distance between nodes at centre must be smaller as compared to distance of nodes closer to network edges.

Table IX: Tabular representation of number of dead nodes in subsequent rounds

| Round | 100 | 200 | 300 | 400 | 500 |
|-------------------|-----|-----|-----|-----|-----|
| No. of dead nodes | 0 | 7 | 31 | 43 | 58 |

Table VIII depicts the Number of packets transmitted to the base station versus the number of rounds. The rounds increase by 100, 200, 300, 400, 500 respectively. Number of rounds is directly proportional to number of packets transmitted to base station with original glucose level of patients. The packet size increases by 2200, 6000, 8200, 10000, 11200 respectively. Delay model considers probability that channel is busy whenever node senses that channel is independent of back-off stage. Moreover, probability that transmission which yields busy channel when the channel is sensed for next time is not negligible. Influence of retransmissions on collision probability is neglected. Whenever frame generation rate increases, more collisions are expected which results in less accurate model. Table IX depicts the Number of dead nodes in the network versus the number of rounds. The rounds increase by 100, 200, 300, 400, and 500 respectively. The number of dead nodes in the network may vary based on number of rounds. A fixed number of nodes is distributed uniformly on a finite square area and all of them follow the same traffic model. All packets are proposed to have common sink node and routing tries to reduce hop count. To capture properties of those networks, a model is proposed for probability of simultaneous transmissions which serves as a baseline to calculate collision and busy channel probabilities.

Table X: Sum. of energy Vs number of rounds

| Round | 100 | 200 | 300 | 400 | 500 |
|----------------|-----|-----|-----|-----|-----|
| Sum. of energy | 40 | 25 | 16 | 13 | 8 |

Table X depicts the Energy consumed by nodes versus the number of rounds. The rounds increase by 100, 200, 300, 400, 500 respectively. Number of rounds is inversely proportional to number of packets transmitted to base station with original glucose level of patients. Energy consumption varies based on active nodes in the network. As well, enhancements of energy consumption are satisfied for both cluster head and cluster members sensors. This development is achieved owing to switch off and sleep mode allocated to each sensor node after transmission and appropriate CH selection that are determined before. At last, Switch off and sleep mode protect sensor node from inefficient transmission and cluster head from an idle listening stage.

Table XI: Tabular representation of time consumed by nodes in subsequent rounds

| Round | 100 | 200 | 300 | 400 | 500 |
|------------------|------------|------------|------------|------------|------------|
| Time consumption | 0.141 1 | 0.164 9 | 0.268 2 | 0.254 7 | 0.375 1 |

Table XI depicts the Energy consumed by nodes versus the number of rounds. The rounds increase by 100, 200, 300, 400, 500 respectively. The average energy consumed during each iteration is 0.1411, 0.1649, 0.2682, 0.2547 and 0.3751 correspondingly. Energy consumption determines the total energy consumed by network to carry out transmission, reception and data aggregation. Time consumption comparison is performed among different approaches based on energy consumption in both cluster head sensor nodes and cluster member sensor nodes.

V. CONCLUSION

In this investigation, a continuous monitoring of glucose level using an effectual protocol termed Adaptive Positioning of Cluster Head based Map reducing (APCH-MR) is anticipated. Here, initially cluster heads are elected for data transmission, when the data reaches the cluster head, mapping of data to codebook is performed and reduction of codebook with routing table indices is carried out. This makes the sensitive data to be more privacy concerned as some individuals needs their data to be confidential. The data that dissipates from cluster head reaches the remote location with reduced power consumption and reduction of unwanted data that causes congestion. Sensor nodes of system are capable to attain several data (i.e. glucose, body temperature, and environmental data) and transmit data wirelessly to gateway resourcefully in term of energy consumption. Therefore, the total number of packets transmitted in round 500 is 11200, total number of dead nodes during round 500 is 58, and time consumed by nodes at 500 rounds is 0.3751s respectively. The proposed method shows better trade off in contrast to conventional techniques. Finally, Doctors or patients' smart phone can also act as a gateway for data transmission and can identify the notification. The attained result is more feasible to monitor the glucose level remotely from any time and can also applicable in real time systems which causes energy efficiency as an effectual factor.

REFERENCES

1. S.A. Haque et al. Review of cyber-physical system in healthcare. *International Journal of Distributed Sensor Networks*, 2014, 2014.
2. A. Aragues et al. Trends and challenges of the emerging technologies toward interoperability and standardization in e-health communications. *IEEE Communications Magazine*, 2011.
3. WHO. Global report on diabetes. <http://apps.who.int/iris/bitstream/10665/204871/1/9789241565257eng.pdf> [accessed 2016 - 12 - 22].
4. P. King et al. The uk prospective diabetes study (ukpds): clinical and therapeutic implications for type 2 diabetes. *British Journal of Clinical Pharmacology*, 1999.



Modelling an Adaptive Cluster Head Positioning Based Map Reducing Strategy for Data Transmission in Medical IoT

5. A. Murakami et al. A continuous glucose monitoring system in critical cardiac patients in the intensive care unit. In *2006 Computers in Cardiology*, pages 233–236. IEEE, 2006.
6. Deepthi Sisodia, “Prediction of Diabetes using Classification Algorithms, International Conference on Computational Intelligence and Data Science (ICCIDIS 2018).
7. Hsin-Yi Tsao, “Predicting diabetic retinopathy and identifying interpretable biomedical features using machine learning algorithms”, *Bioinformatics* 2018, 19(Suppl 9):283
8. N. Vijayalakshmi, “An Analysis Of Risk Factors For Diabetes Using Data Mining Approach”, *International Journal of Computer Science and Mobile Computing*, Vol. 6, Issue. 7, July 2017, pg.166 – 172
9. Hoa Hong Nguyen, “A Review on IoT Healthcare Monitoring Applications and a Vision for Transforming Sensor Data into Real-time Clinical Feedback”, *International Conference on Computer Supported Cooperative Work in Design*, IEEE 2017.
10. Ioannis Kavakiotis, “Machine Learning and Data Mining Methods in Diabetes Research”, *Computational and Structural Biotechnology Journal* 15 (2017) 104–116
11. Paul S. Fisher, “Mining intelligent solution to compensate missing data context of medical IoT devices”, *Personal and Ubiquitous Computing* <https://doi.org/10.1007/s00779-017-1106-1>
12. Yukai Li, Huling Li, “Analysis and Study of Diabetes Follow-Up Data Using a Data-Mining-Based Approach in New Urban Area of Urumqi, Xinjiang, China, 2016-2017”, *Hindawi Computational and Mathematical Methods in Medicine* Volume 2018, Article ID 7207151, 8 pages <https://doi.org/10.1155/2018/7207151>
13. M. Ali et al. A bluetooth low energy implantable glucose monitoring system. In *EuMC 2011*, pages 1265–1268. IEEE, 2011.
14. J. Lucisano et al. Glucose monitoring in individuals with diabetes using a long-term implanted sensor/telemetry system and model. *IEEE Transactions on Biomedical Engineering*, 2016.
15. KAU. Menon et al. A survey on non-invasive blood glucose monitoring using nir. In *ICCSP 2013*, pages 1069–1072. IEEE, 2013.
16. MUH. Al Rasyid et al. Implementation of blood glucose levels monitoring system based on wireless body area network. In *Consumer Electronics-Taiwan (ICCE-TW), 2016 IEEE International Conference on*, pages 1–2. IEEE, 2016.
17. N. Wang and G. Kang. A monitoring system for type 2 diabetes mellitus. In *Healthcom 2012*, pages 62–67. IEEE, 2012
18. H. Park, S. N. Lee, M. Y. Baek et al., “The Well-Being and Treatment Satisfaction of Diabetic Patients in an Outpatient Setting at a General Hospital in Korea,” *The Journal of Korean Diabetes*, vol. 17, no. 2, p. 123, 2016.
19. A. M. Boels, R. C. Vos, T. G. Hermans et al., “What determines treatment satisfaction of patients with type 2 diabetes on insulin therapy? An observational study in eight European countries,” *BMJ Open*, vol. 7, no. 7, p. e016180, 2017.
20. S. Mani, Y. Chen, and T. Elasy, “Type 2 Diabetes Risk Forecasting from EMR Data using Machine Learning,” *AMIA Annual Symposium Proceedings*, pp. 606–615, 2012.
21. K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, “Machine learning applications in cancer prognosis and prediction,” *Computational and Structural Biotechnology Journal*, vol. 13, pp. 8–17, 2015.
22. F. L. Meng and S. H. JIN, “Investigation and Analysis of Satisfaction of Hypertension Patients in Community Health Service in Hangzhou,” *Health Research*, vol. 32, no. 2, pp. 132–134, 2012.
23. F. Alsubaei, A. Abuhusseini, and S. Shiva (2017) Security and Privacy in the Internet of Medical Things: Taxonomy and Risk Assessment, *Proc. of 2017 I.E. conference on local computer networks workshop*, pp 112–120
24. Chisci, H. ElSawy, A. Conti, M. S. Alouini, M. Win (2017) On the Scalability of Uncoordinated Multiple Access for the Internet of Things, *Proc. of 2017 International Symposium on Wireless Communications Systems*. IEEE, Bologna, pp 402–407
25. D. Masouros, I. Bakolas, V. Tsoutsouras, K. Siozior, and D. Soudris (2017) From edge to cloud: Design and Implementation of a Healthcare Internet of Things Infrastructure, *Proc. of 27th International Symposium on Power and Timing Modeling, Optimization and Simulation*, pp. 1–6. <https://doi.org/10.1109/PATMOS.2017.8106984>
26. P. Sherubha, “A detailed survey on security attacks in wireless sensor networks: *International Journal of Soft Computing* 11 (3), 221-226.
27. P. Sherubha, M. Banu chitra, “Multi class feature selection for breast cancer detection”, *International journal of pure and applied mathematics*, 201

AUTHORS PROFILE



GAMEIL S. H. ALI, Ph.D Research scholar, Department of Computer Science, Rathnavel Subramaniam College of Arts & Science, Coimbatore, Tamil Nadu, India, Gameilsaad01@gmail.com



Dr. A.NITHYA, Research Supervisor Associate Professor, Department of Computer Science, Rathnavel Subramaniam College of Arts & Science, Coimbatore, Tamil Nadu, India