Energy Efficient Data Mining Approach for Estimating the Diabetes

Rajesh Sharma R

Computer Science and Engineering, School of Electrical Engineering and Computing, Adama Science and Technology University, Adama, Nazret, Ethiopia
**E-mail:** sharmaphd10@gmail.com

Abstract

Diabetes is a major cause of organ failure in the human body, and it is one of the leading causes of organ failure. As of now, there is no preventive medicine or vaccine for diabetes. As a result, people all around the world are accustomed to living with diabetes for the rest of their lives. Medical practitioners advise diabetic patients to have a healthy lifestyle that includes regular exercise and a well-balanced diet in order to prevent the effects of diabetes from spreading to other organs of the human body. In most cases, the diabetes is spreading like a heredity disease to the infected people and even to children and it can't be estimated priory. In recent days, the deep learning algorithms are widely used to estimate the forthcoming effects of several problems by using the data mining process. In the proposed work, the performance of deep ANN and back propagation ANN is considered for estimating diabetes from several primary data factors obtained from a publicly available dataset called Pima Indian diabetes dataset.

Keywords: Diabetes, healthcare big data, ANN, DNN, BPANN

1. Introduction

A typical human cannot perform big data analytics using his or her intellect. Most of the time, conducting big data analytics necessitates the application of a theorem and formula [1, 2]. Also, in most of the analytic problems it consumes non acceptable amount of time than the usual. Naturally, the accuracy of human analytic models will be higher. However,

there is a greater possibility of human error while implementing human analytics [3, 4]. This can be rectified by implementing a neural network algorithm for the analysis process. The neural network algorithms are not common, it will change with respect to their connection between each neurons. Henceforth, the performance of each neural network algorithm will not be equal in nature. This necessitates the need to validate any analytic problem with more number of neural network algorithms in order to identify the best one [5, 6].
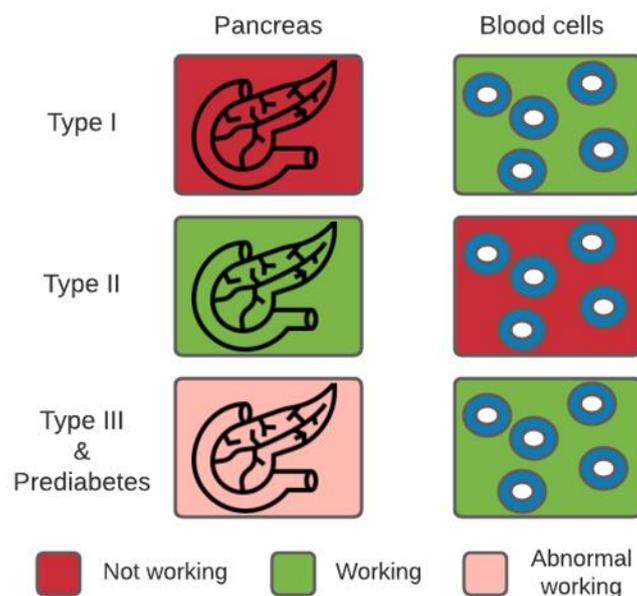
Figure 1. Types of Diabetes

The diabetes diseases are classified into three types; they are type 1, 2 and 3 with respect to their effects on human body. Type 1 diabetes is also called as insulin dependent diabetes, where the pancreas fails to generate the sufficient insulin required for a human body [7, 8]. The irregular level of insulin produced from human body results in the poor segregation of nutrients from the consumed food. A human action or a daily routine cannot be instantly identified in the scene as long as the human body try to manage with the available nutrients and produced insulin [9, 10]. The body reacts to such reduced insulin level only at some critical stages. The second type of diabetes is called as non-insulin dependent diabetes,

where the insulin level generated by the human body is normal but the human body cell loss its ability to accept the produced insulin for its function. As a result, the abnormal presence of insulin over a human body will result in some other problems rather than the type 1 diabetes [11]. There is also one more type called type 3, where the insulin level get increased for only few days or month due to pregnancy or other human disorders. Apart from this, there is prediabetes, in which the pancreas produces slightly more insulin to allow the abnormal blood cells to take the generated insulin. The abnormality is not termed type 2 diabetes in such circumstances. Figure 1 depicts the difference between various types of diabetes in a graphical format [12].

The diabetes comes under the category of non-communicable disease. Though it can make a genetic disorder and spread the abnormality to the next-generation of diabetes affected people. Therefore, with the help of parental historic statements along with the human's body mass index and several health conditions, healthcare practioners can predict a diabetes condition little earlier [13, 14]. It will be easy to control the abnormality of a pancreas, when it is found at the earlier stage. The control techniques include diet management, adequate sleep, drugs and physical activity. The challenging task in predicting the diabetes is its accuracy. Due to the poor prediction accuracy, there is also a chance for observing false positive to diabetes. Treating a false positive human with the diabetes drug can make the person health worse than the actual diabetes patient. In order to minimize the false positive rate with better accuracy rate, a list of confirmed diabetes positive and negative samples with different attributes are considered for training the neural network model [15]. The following section 2 discusses about the achievements and limitations of the previous diabetic prediction models. Similarly, section 3 will provide the description and workflow of the proposed model. The experimental analysis carried out in the work is presented in section 4 and the performed research study is concluded at section 5.

2. Related Work

In the field of machine learning and data mining, different methods are available to collect information from a patient. Basically, there are two types of data transmission namely wireless and wired transmission. In both cases, the security is playing a major role in delivering the accurate data. Due to interruption in received data, there are chances for an algorithm to work improperly and it can lead to poor prediction [16]. The quality of wired transmission system is always reliable in such cases than the wireless transmission. However, for long distance communication the wireless transmissions are preferably used all over the world. To protect the data transmitted through wireless communication and IoT modules, the blockchain technology is developed before few years. Apart from IoT and WSNs unmanned aerial vehicles are also employed in several fields for transferring the measured healthcare data from one place to another place [17].

Data forwarding through wireless body area networks are applied to identify several critical and sudden illness from any human. Small self-battery operated devices are widely used in such cases to measure certain medical parameters from a human body at certain time interval. In some cases, the devices are connected to observe the patients in a continuous manner. In such continuous monitoring devices, the limitation is about its continuous power backup. Therefore, most of the devices are connected to monitor the required parameters at certain time intervals. The observed data will also be forwarded to the base stating when the measured reading crosses the given threshold value [18]. These kinds of devices are widely used for patients, who have undergone critical surgery and having the chances for getting sudden heart arrest. Some of the recent devices are having the ability to send the location of patent data along their observed parameters. In such cases, locating the medical emergency patient becomes easy for a medical practitioner. This eliminates the need to make phone calls to hospital, when the patient is connected to a GPS tracking device on the health monitoring device [19].

Electric
Engineering and Automation

The energy wastage on such healthcare data transmission devices can be minimized by employing an efficient data mining algorithms. An algorithm with simple computational steps can minimize and save huge amount of energy available in a battery source. Developing an efficient computational algorithm is not easy [20]. However, due to the advancement of neural network algorithms, this is now possible to a certain extent. All neural network algorithms operate in terms of their internal neural connections, which might result in varying computing time and efficiency across different algorithms. Therefore, it is always necessary to verify the developed prototype with multiple neural network algorithms before its final implementation [21].

Several routing algorithms are also being developed to reduce the energy consumption of healthcare devices. When many data sets are sent at the same time, the data waiting time at the source is reduced [22]. The routing process can also be changed and modified with respect to the hardware modules implemented in the healthcare device. Some of the IoT devices are directly connected to the server as a mediator for making a reliable transmission between the source and destination. In such cases, the data loss will be minimized and the length of data transmission can also be improved. In some other cases, the data from source place is transmitted to the destination via several sub source stations available near the actual devices [23]. However, these types of sub source station devices are only suitable for very short range data transmission. To improve the quality of utilizing health care device on a human body, it should satisfy the patient's comfort level. As a result, the incorporation of light-weight components is always necessary to be addressed while developing such devices [24, 25].

3. Proposed Method

The proposed work has two major modules in its architecture. The primary module is a healthcare device connected with very few sensors and input devices for including the observed and measured data from a human body. The secondary module is a computer

system with neural network frameworks for classifying the patients based on the observed readings. Figure 2 indicates the architecture of the primary module, which consists of sensors and keyboards connected to a microcontroller along with an IoT peripheral module. The IoT module is attached here to send the sensor's observed values from the base station to the destination for graphing in an online data streamer. The suggested approach makes use of a Microsoft data streamer to store the data. The device also has a keyboard for sending health care data to the online streamer. The energy requirement of the utilized healthcare device is limited by changing the interval time of measurements. The microcontroller algorithm is designed to follow its internal clock cycle for enabling the sensor pins at certain regular period. The energy is also saved by enabling the sensors one by one. There are various procedures to follow in medical aspects, such as if some 'abc' readings are normal, there is no need to measure the 'xyz' readings. For monitoring sensor readings, the suggested model employs such technique. As a result, the suggested technique avoids the undesired measurement of several sensor values.
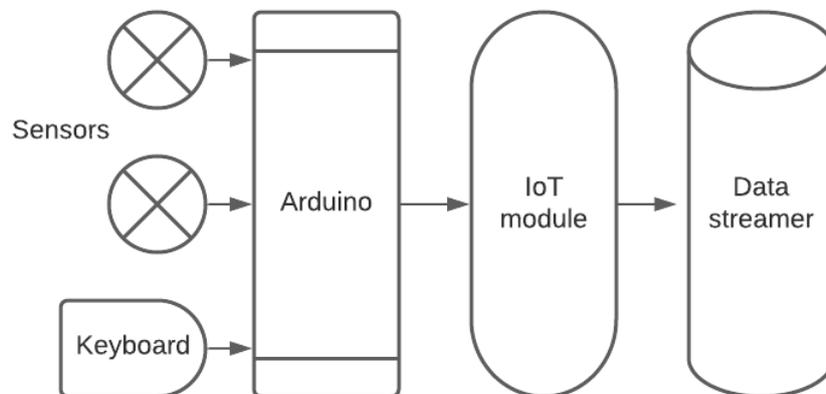
Figure 2. Architecture of the Data Collection Module

The architecture of the secondary module is shown in figure 3, where the data streamer is specified as a self-made dataset. The architecture of the data collection module is also specified in the diagram as an energy efficient healthcare device. The self-made dataset

is connected with an unbalanced dataset to make an efficient database in terms of equality in all the classes. Following to the database, there is a pre-processing step employed in the work to verify the abnormality in the observed healthcare data. The pre-processing step is included in the proposed model to remove and restructure the abnormal data available in the database due to human error and machine error.
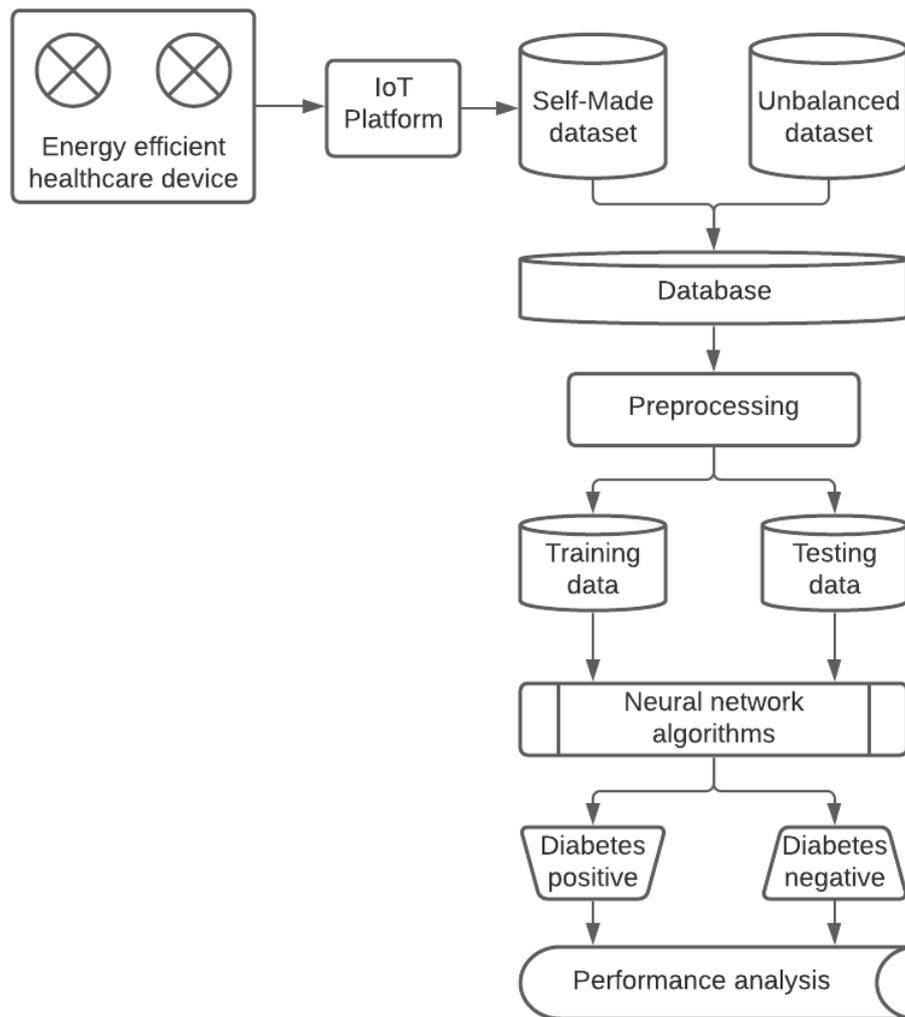
Figure 3. Architecture of the Proposed Classification Model

In order to identify an efficient neural network algorithm, the proposed model is verified with a set of three most popular neural network algorithms namely ANN, DNN and BPANN. All these three neural network algorithms are belong to one family, where there is only a very few changes on its neurological connections. The artificial neural network (ANN) is connected here with a single hidden layer neuron connections between input and output layers. The network layers are only connected in a forward direction, from input to output layers. The deep neural network (DNN) operates on the same forward direction concept as the neural network, but with multiple hidden layer connections between the input and output layers. There, the information collected in DNN are usually better than the actual ANN. BPANN is also a kind of ANN model having a single hidden layer connection with a backward propagation between the output layer and hidden layer. Therefore, the learning ability of hidden layers will be improved based upon its recent accuracy. The complexity of BPANN is limited by utilizing it with a single hidden layer. Apart from their accuracy attainments, the computational speed is also considered in most of the neural network models during their implementation stage.

4. Experimental Analysis

Pima Indians diabetes database is employed in the work for analysis and it is mentioned as unbalanced dataset in the architecture image. The dataset is specified as unbalanced one because the class distribution of the specified dataset is not even. The dataset consists of 500 diabetes negative instances and 268 diabetes positive instances. Therefore, the dataset is made balanced in the proposed work without any artificial data augmentation process. A set of 232 instances are measured with diabetes patients through a simple health care device and stored in a data streamer for making a balanced dataset. The complete database is separated as like 75% and 25% for the training and testing process. Therefore, a total of 375 samples from both the classes are taken in to consideration for the training and 125 samples are considered for the testing process. The dataset contains eight attributes, and

in the proposed study, all eight attributes are evaluated for assessment. Table 1 compares the performance of ANN, DNN, and BPANN algorithms.

Table 1. Performance Analysis of the Proposed Work

Algorithm	ANN	DNN	BPANN
Accuracy	92.00	92.40	95.60
Precision	93.39	92.74	95.24
Recall	90.40	92.00	96.00
F1 Score	91.87	92.37	95.62
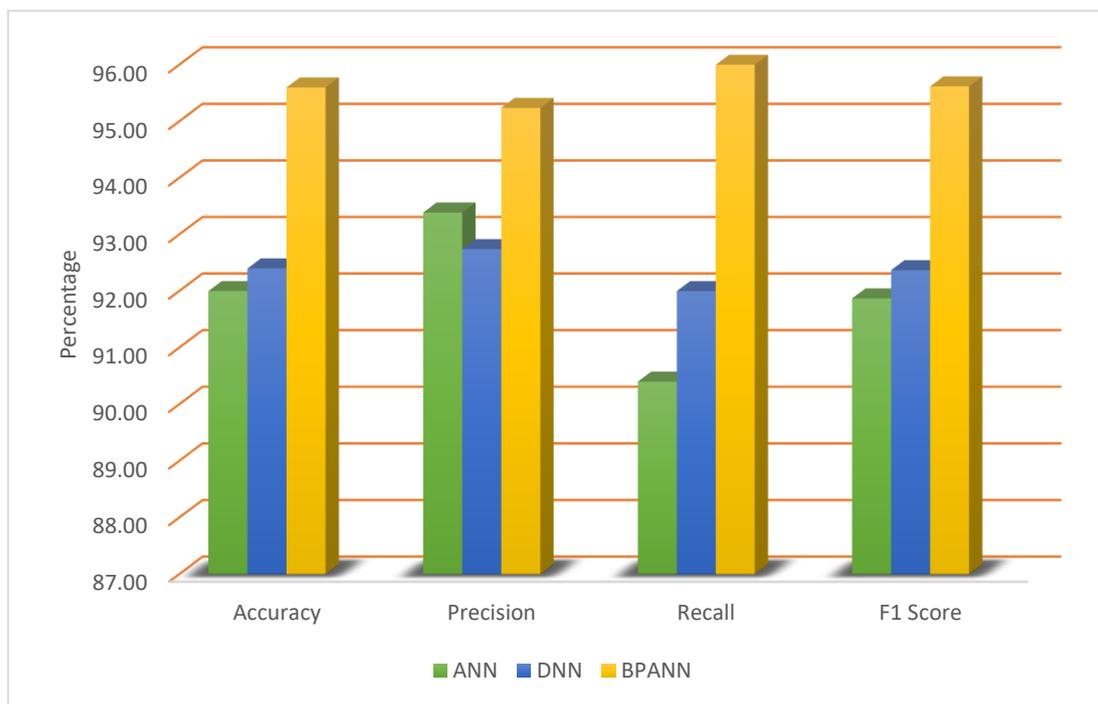
Figure 4. Performance Comparison among the Different Algorithms

118

The performance of different algorithms is tabulated from the different neural network algorithms are represented in a graphical format in figure 4. It clearly indicates that, the performance of BPANN is performing in a better way than ANN and DNN. The overall performance deviation is not gradual in all the parameters. Among the others, the accuracy of DNN deviates slightly more than that of ANN. On the other hand, BPANN requires a longer processing time than the other two algorithms. However, in medical and healthcare applications, accuracy is more important than computational time. The following table 2 specifies the energy requirement of the proposed healthcare device model during its enabling time and sleeping time.

Table 2. Performance of the Healthcare Device by the Proposed Method

Modes	Power consumption
Active Mode	12mA
Sleep Mode	32µA

5. Conclusion

Healthcare devices are primarily used to monitor several heath issues at an instant manner. In the proposed work, the role of the health care devices is extended to store the observed reading for further data mining process. An experimental analysis is performed in the proposed work to verify the accuracy and several classification parameters. It shows that, the back propagation artificial neural network algorithm performs better than the ANN and DNN in terms of different parameters. The accuracy level of BPANN reaches to 95.6 with an acceptable computational time. The energy consumption of the utilized healthcare module is

also observed in the proposed work for proving an efficient device. The developed module is designed to operate with a simple battery source for longer hours by using a simple on-off control algorithm and the clock cycle of the employed microcontroller.

References

[1] Haoxiang, Wang, and S. Smys. "Big Data Analysis and Perturbation using Data Mining Algorithm." Journal of Soft Computing Paradigm (JSCP) 3, no. 01 (2021): 19-28.

[2] Reddy, G. Thippa, M. Praveen Kumar Reddy, Kuruva Lakshmanna, Rajesh Kaluri, Dharmendra Singh Rajput, Gautam Srivastava, and Thar Baker. "Analysis of dimensionality reduction techniques on big data." IEEE Access 8 (2020): 54776-54788.

[3] Nayak, Lipsa, and V. Jayalakshmi. "A Survey on Privacy Preserving Approaches on Health Care Big Data in Cloud." In International conference on Computer Networks, Big data and IoT, pp. 32-38. Springer, Cham, 2019.

[4] Sungheetha, Akey. "COVID-19 Risk Minimization Decision Making Strategy Using Data-Driven Model." Journal of Information Technology 3, no. 01 (2021): 57-66.

[5] Hezekiah, James DK, and Shanty Chacko. "A Review on cornea imaging and processing techniques." Current medical imaging 16, no. 3 (2020): 181-192.

[6] Chitra, S., and V. Jayalakshmi. "A Review of Healthcare Applications on Internet of Things." Computer Networks, Big Data and IoT (2021): 227-237.

[7] Sungheetha, Akey, and Rajesh Sharma. "Design an Early Detection and Classification for Diabetic Retinopathy by Deep Feature Extraction based Convolution Neural Network." Journal of Trends in Computer Science and Smart technology (TCSST) 3, no. 02 (2021): 81-94.

[8] Wu, Zeng-hong, Yun Tang, and Qing Cheng. "Diabetes increases the mortality of patients with COVID-19: a meta-analysis." Acta diabetologica 58, no. 2 (2021): 139-144.

[9] Yashodhara, P. H. A. H. K., and D. D. M. Ranasinghe. "Early Detection of Diabetes by Iris Image Analysis." In Inventive Computation and Information Technologies, pp. 615-632. Springer, Singapore, 2021.

[10] Chen, Joy Iong Zong, and P. Hengjinda. "Early Prediction of Coronary Artery Disease (CAD) by Machine Learning Method-A Comparative Study." Journal of Artificial Intelligence 3, no. 01 (2021): 17-33.

[11] Williams, Rhys, Suvi Karuranga, Belma Malanda, Pouya Saeedi, Abdul Basit, Stéphane Besançon, Christian Bommer et al. "Global and regional estimates and projections of diabetes-related health expenditure: Results from the International Diabetes Federation Diabetes Atlas." Diabetes research and clinical practice 162 (2020): 108072.

[12] Shibly, Mir Moynuddin Ahmed, Tahmina Akter Tisha, and Md Mahfuzul Islam Mazumder. "Predicting Early Readmission of Diabetic Patients: Toward Interpretable Models." In International Conference on Communication, Computing and Electronics Systems: Proceedings of ICCCES 2020, vol. 733, p. 185. Springer Nature, 2021.

[13] Manoharan, Samuel. "Patient diet recommendation system using K clique and deep learning classifiers." Journal of Artificial Intelligence 2, no. 02 (2020): 121-130.

[14] Apoorva, S., P. Snigdha, P. Darshini, and H. A. Sanjay. "Prediction of diabetes mellitus type-2 using machine learning." In International Conference On Computational Vision and Bio Inspired Computing, pp. 364-370. Springer, Cham, 2019.

[15] Koresh, H. James Deva, Shanty Chacko, and M. Periyanayagi. "A modified capsule network algorithm for oct corneal image segmentation." Pattern Recognition Letters 143 (2021): 104-112.

[16] Raj, Jennifer S. "Security Enhanced Blockchain based Unmanned Aerial Vehicle Health Monitoring System." Journal of ISMAC 3, no. 02 (2021): 121-131.

[17] Aggarwal, Shubhani, Neeraj Kumar, Musaed Alhussein, and Ghulam Muhammad. "Blockchain-Based UAV Path Planning for Healthcare 4.0: Current Challenges and the Way Ahead." IEEE Network 35, no. 1 (2021): 20-29.

[18]   Chen, J. I. Z., & Yeh, L. T. (2020). Data Forwarding in Wireless Body Area Networks. Journal of Electronics, 2(02), 80-87.

[19]   Mbizvo, Gashirai K., Christian Schnier, Colin R. Simpson, Susan E. Duncan, and Richard FM Chin. "Validating the accuracy of administrative healthcare data identifying epilepsy in deceased adults: A Scottish data linkage study." Epilepsy research 167 (2020): 106462.

[20]   Shakya, Subarna, and Lalitpur Nepal. "Computational Enhancements of Wearable Healthcare Devices on Pervasive Computing System." Journal of Ubiquitous Computing and Communication Technologies (UCCT) 2, no. 02 (2020): 98-108.

[21]   Li, Shuisheng, Baoxin Zhang, Panfeng Fei, P. Mohamed Shakeel, and R. Dinesh Jackson Samuel. "Computational efficient wearable sensor network health monitoring system for sports athletics using IoT." Aggression and Violent Behavior (2020): 101541.

[22]   Suma, V. "Wearable IoT based Distributed Framework for Ubiquitous Computing." Journal of Ubiquitous Computing and Communication Technologies (UCCT) 3, no. 01 (2021): 23-32.

[23]   Ullah, Ata, Ghawar Said, Muhammad Sher, and Huansheng Ning. "Fog-assisted secure healthcare data aggregation scheme in IoT-enabled WSN." Peer-to-Peer Networking and Applications 13, no. 1 (2020): 163-174.

[24]   Hariharakrishnan, Jayaram, and N. Bhalaji. "Adaptability Analysis of 6LoWPAN and RPL for Healthcare applications of Internet-of-Things." Journal of ISMAC 3, no. 02 (2021): 69-81.

[25]   Masud, Mehedi, Gurjot Singh Gaba, Karanjeet Choudhary, Roobaea Alroobaea, and M. Shamim Hossain. "A robust and lightweight secure access scheme for cloud based E-healthcare services." Peer-to-peer Networking and Applications (2021): 1-15.

Electric
Engineering and Automation

Author's biography

Rajesh Sharma R is a Computer Vision and Robotics SIG Coordinator, in the department of Computer Science and Engineering, in School of Electrical Engineering and Computing, in Adama Science and Technology University, Adama, Nazret, Ethiopia. He has more than 8 years of academic experience. His areas of research are networking, probabilistic computing, fuzzy, bio- inspired computing, data visualization, fault diagnosis, robotics, internet of things, neurocomputing, information retrieval, human-machine interface and network security. He has published in more than 20 international and national journals. He is a life member of International Association of Engineers and Indian Society of Technical Education, and member of IAENG, IACSIT and AACE.