Rathi, Vipin Kumar; Rajput, Nikhil Kumar; Mishra, Shubham; Grover, Bhavya Ahuja; Tiwari, Prayag; Jaiswal, Amit Kumar; Hossain, M. Shamim

An edge AI-enabled IoT healthcare monitoring system for smart cities

Published in:
Computers and Electrical Engineering

DOI:
10.1016/j.compeleceng.2021.107524

Published: 01/12/2021

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Rathi, V. K., Rajput, N. K., Mishra, S., Grover, B. A., Tiwari, P., Jaiswal, A. K., & Hossain, M. S. (2021). An edge AI-enabled IoT healthcare monitoring system for smart cities. *Computers and Electrical Engineering*, *96*, Article 107524. https://doi.org/10.1016/j.compeleceng.2021.107524

An Edge AI-enabled IoT Healthcare Monitoring System for Smart Cities

Vipin Kumar Rathi[a], Nikhil Kumar Rajput[a], Shubham Mishra[a], Bhavya Ahuja Grover[a], Prayag Tiwari[b], Amit Kumar Jaiswal[c], M. Shamim Hossain[d,e,*]

[a] *Ramanujan College, University of Delhi New Delhi, India*
[b] *Department of Computer Science, Aalto University, Finland*
[c] *School of Mathematics, University of Leeds, United Kingdom*
[d] *Research Chair of Pervasive and Mobile Computing, King Saud University, Riyadh 11543, Saudi Arabia*
[e] *Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia*

Abstract

Healthcare systems have significantly benefited from Artificial Intelligence (AI) and the Internet of Things (IoT). The vital signs of patients can be continuously monitored using the technologies mentioned above, and timely treatment can be provided. To this end, this paper proposes a scalable, responsive, and reliable AI-enabled IoT and edge computing-based healthcare solution with low latency when serving patients. The system comprises the collection of health-related data, data processing and analysis at edge nodes, and permanent storage and sharing at edge data centers. The edge nodes and edge controller schedule patients and provide resources in real time. Simulations were conducted to test system performance. The results for end-to-end time, computing, optimization, and transmission latency prove to be very promising. To determine system performance in a real-world scenario, a neural network was used to model transmission latency. The system is extremely useful for those who are disabled or elderly, as well as in pandemic situations.

Keywords: AI-enabled IoT, Smart Healthcare, IoT Edge, Multi-access Edge Computing, Smart City

1. Introduction

The emergence of IoT has revolutionized the way society operates. It opens up a plethora of opportunities to improve one's quality of life in all areas, ranging from education to healthcare, industrial operations to banking, increased business productivity, and finally automation and control. The healthcare sector can greatly benefit from IoT, such as improved monitoring with smart sensors, real-time access to vital signs, remote health monitoring, and efficient disease detection, to name a few.

Considering pandemic situations like COVID-19 [1], it is evident that the timely detection and treatment of disease can help save lives and control the pandemic. In such a situation, if people are equipped with

*Corresponding author

Email addresses: vipin.rathi@ramanujan.du.ac.in (Vipin Kumar Rathi), n.rajput@ramanujan.du.ac.in (Nikhil Kumar Rajput), shubham.mishra@ramanujan.du.ac.in (Shubham Mishra), b.ahuja@ramanujan.du.ac.in (Bhavya Ahuja Grover), prayag.tiwari@aalto.fi (Prayag Tiwari), a.k.jaiswal@leeds.ac.uk (Amit Kumar Jaiswal), mshossain@ksu.edu.sa (M. Shamim Hossain)

health monitoring devices like temperature sensors, anomalies can be detected at an earlier stage. The same can be communicated to a connected hospital, which can immediately initiate responsive action for treatment. This would be highly beneficial in preventing damage caused by the disease and its further spread.

This has become possible with an Artificial Intelligence (AI) enabled Internet of Things (IoT). AI-enabled IoT is a widely acclaimed technological tool that builds a provision for integrating various electronic devices. The devices can collect and transmit data and also trigger appropriate actions on other devices. It has immense applications in the fields of agriculture, industry, and inventory management, to name a few [2, 3]. Say, the temperature of the machine area in a factory is being monitored continuously. As soon as the value crosses a set threshold, an alarm can be raised to notify the authorities for immediate responsive action. In the field of healthcare, there are several sensors [4, 5] that can be incorporated to collect data related to a patient's health (including patient's images), and an emergent response can be triggered when required. Utilizing these devices, it is extremely simple to monitor a person's health remotely and to provide assistance [6].

Smart systems built using AI and the IoT have proved to be immensely effective in healthcare. The healthcare sector may be revolutionized by edge-enabled IoT, where a massive number of embedded sensors and IoT devices interconnect to provide different services to the community for the well-being of their citizens. With the current COVID-19 situation, along with the presence of chronic diseases and an aging population, interconnected IoT devices generate a huge amount of IoT data [7, 8, 9]. The situation is changing so quickly that healthcare systems have not been able to keep up. It is essential to diagnose and detect diseases at an early stage to facilitate timely treatment. This also reduces the cost of healthcare. To this end, the convergence of edge AI and IoT has the potential to classify and cluster this massive amount of IoT data, make predictions, and deliver early insights. This could address the global challenge of the pandemic and other related problems.

1.1. Contribution

- The present work proposes an edge AI enabled IoT healthcare monitoring system for a smart city that can remarkably improve healthcare facilities and infrastructure and ensure the timely treatment of patients. This could prove highly significant in strengthening the health infrastructure, especially during pandemic situations.

- The concept of end-to-end network slice for healthcare services is also proposed, which ensures reduced latency and scalability.

- A scheme is proposed for the efficient scheduling of patients and resource provisioning.

1.2. Organization

This paper is organized as follows: Section 2 provides a review of the literature for the present work that was done to design smart healthcare systems. Section 3 describes the proposed architecture and the implementation details. Section 4 presents the results of the experimental simulation conducted to test the performance of the system. The final section concludes the paper.

2. Related Works

The idea of continuously monitoring a patient's health using wearable and implanted devices has recently been gaining widespread adoption. A detailed account of the same was presented in [10], which reports a survey of the advances in Wireless Body Area Network (WBAN) systems and integrated technologies. This paper discusses several low-power wireless technologies that can be deployed. These sensor nodes can continuously monitor patients' health and collect vital data. The authors have detailed the usage of Bluetooth Low Energy (BLE) IEEE 802.15.4 and ZigBee, classic Bluetooth, ANT, RuBee, Sensium, and Zarlink among other low power consumption technologies. The authors also talk about intra-body communication technologies that use wirelessly interconnected implanted devices. The importance of communication protocols is also highlighted. The utility of m-Health applications built on smartphones that help monitor patient parameters like temperature, pulse rate, and breathing rate is also shown.

In the paper [11], the authors described the utility of IoT architectures for enhanced living environments and healthcare systems. They discussed at length open-source platforms like Kaa and Thingsboard, protocols like MQTT, smartphones, wearables serving as sensors, healthcare applications like the SPHERE project, and Home Health Hub IoT, as well as issues like Quality of Service, security, availability, compatibility, reliability, and future prospects.

In the paper [12], the authors specifically addressed the issue of an enhanced monitoring scheme for patients admitted in critical condition. A framework using low-cost, low-power wearable sensors connected to the internet and using the open communication protocol oneM2M was outlined. Here, the wearables were used as Application Dedicated Nodes (ADN) that communicate with an infrastructure node containing the common services entity. Furthermore, openEHR has been used in the higher layers for providing functionality, namely data semantics, storage, and monitoring. The experiments report a latency of 20–50 ms, and a 30–50 h sensor autonomy. They also discussed the design of efficient M2M-capable sensors that could provide the benefit of low cost, and energy requirements through ESP8266 Wi-Fi modules.

A multilayer fusion of Convolutional Neural Networks (CNNs) was proposed to detect Electroencephalogram (EEG)-based pathology detection in [13]. The fusion was done using a multilayer perceptron or an autoencoder. The experiments were performed in a smart healthcare framework. Around 90% accuracy and 97% specificity were obtained by the proposed system.

A heart rate and SPO2 monitoring system was proposed, which was specifically designed for mother and child healthcare [14]. The pulse oximetry method has been used for measuring heart rate and oxygen saturation using Arduino-based sensors and Raspberry Pi. The data pertaining to heart rate, SPO2, and the photoplethysmogram signal collected by sensors is analyzed with Raspberry Pi, and any issues are identified. All the data, along with any plausible abnormalities, are communicated to a Thingsboard IoT platform using the MTTQ protocol. The interface at the Thingsboard dashboard accurately reports the data and issues like tachycardia. Uncertainty measurements were reported, such as the standard deviation of heart rate at 2.85 bpm and SPO2 at 0.28%.

A low-cost product, "Smart Chair," is described in [15], and it has sensors for collecting health-related data for regular monitoring. The data are transmitted to a web server where they can be referenced at any time by a doctor or the individual. The results can be used remotely by a doctor for quick feedback and assistance. A multi-platform communication architecture based on SOA was proposed for data management through HL7 standards. A software solution to detect vital signs from data for enforcing control is also provided with the same.

Several papers have talked about an important concern related to the advancement of WBAN technology, which is the energy consumption of wearable and implanted sensor nodes. In the paper [16], the authors proposed a solution to the joint scheduling and admission control problem to optimize the energy consumption by the gateway node and WBAN sensor node. The authors used the constrained Markov decision process in their approach and the Lagrange multiplier to arrive at a solution. Through simulation, the method gave rise to a 100% throughput improvement and reduced the power consumption by nearly 5.5-fold.

In [17], the authors proposed an architecture for a scalable system that comprised sensors, a smart gateway to set up the network, communication of data to a server, automatic data collection, and processing and analysis through feature extraction. The implementation of an ECG monitoring system was presented, and the preliminary results obtained for accuracy, power consumption, and cost that seemed promising were reported.

With an enormous amount of data being generated, it has become challenging to store, access, and analyze the data in a time-bound manner to stimulate prompt actions that can preserve patient health and also reduce the cost of treatment. It is known that IoT devices have limited storage and computational capabilities. For the same, cognitive computing [18] may prove to be of immense significance.

Multi-access Edge Computing (MEC) is a relatively newer area of research and is quickly gaining popularity with the introduction of 5G technology. MEC emphasizes an architecture that uses base stations or servers as processing terminals. These stations can be used to not only process the received data but also to trigger quick responses due to their close proximity. The data can be stored permanently at a central cloud server.

Shifting to edge technologies may help solve many of the challenges faced by healthcare systems. With this, time sensitive data can be quickly analyzed at the edge server. This would help in multiple ways. The latency would be reduced, and hence the data could be processed in a timely way. The patient wait time would be reduced. Resources like doctors, nurses, physical infrastructural facilities like ICUs, stretchers, and wheelchairs could be promptly provided on a priority basis. Remote monitoring of patients could be done, and direct communication could be established between the patient and healthcare provider. This would facilitate prompt action and better care and would also control costs [1].

Security is another crucial requirement that needs to be considered when healthcare-related data is stored on a third-party cloud server. An edge near the hospital can be a viable solution for this as well. The edge server can store all the data for the hospital and its patients locally in a protected manner. This would also assist in the early analysis of any critical issues and the direction of appropriate resources to address that issue. For instance, through analytical processing at the edge, if a vital parameter like breathing rate is going awry, signifying that a patient in the hospital is critical, appropriate resources like doctors, nurses, masks, etc., can be quickly provisioned. The doctor can continuously monitor their patients' vitals through the dashboards and can update the prescribed treatment as and when required. The relevant data for maintaining patient records as well as for assisting in future research and other healthcare developments could be transmitted for storage in the cloud.

A decoupled blockchain-based approach using nearby edge devices to create decoupled blocks in a blockchain was proposed in [19]. The system was designed to securely transmit healthcare data from sensors to edge nodes, which would then transmit the same to the cloud using an incremental tensor-based scheme. The scheme was designed to reduce duplication of the huge amount of healthcare data transmitted.
In [20], the authors proposed a system comprising hybrid routers that supports two wireless protocols, BLE and Long Range (LoRa), and an IoT gateway. The router is equipped with a solar energy harvester to extend the router's lifetime. The system supports advanced edge computing tasks, such as data storage, processing, user interface, and cloud connection. Experimental simulations were conducted with respect to the healthcare scenario. The results showed a minimal delay of only 11.5 ms and a network boundary ranging to 2.4 km with the hybrid LoRa network implementation.

This would help in optimizing available resources. Resources like doctors and biomedical [21, 22] equipment could be provided to those in immediate need, especially during strenuous situations like a pandemic. Also, this would increase the responsiveness of the healthcare system, as treatment would be provided in a time-bound manner. The overall cost per patient would also be reduced with benefits like early diagnosis of diseases and the prompt detection of critical situations.

[1] https://www.business.att.com/learn/tech-advice/how-healthcare-organizations-are-innovating-with-edge-to-edge-technologies.html

3. Proposed Edge AI Enabled IoT Healthcare Monitoring System

There has been significant advancement and innovation in the area of edge healthcare networks. From the literature, we have identified some essential requirements and issues that prevail in implementing such systems. Our solution intends to address the same.

Security:. The data related to a healthcare scenario relates to the medical parameters of individuals. There is a great need for security for this data, as it is of utmost importance and is a privacy concern. Expecting the IoT devices to be trustworthy, transmission through the network is uncertain, as there is some chance of tapping or forging the data. Thus. the security of the system is a crucial issue to be resolved.

Ownership:. The data needs to come from the rightful person and not some adversary device. This poses a need for an authentication mechanism.

Low computational and storage capability of IoT devices:. These devices are embedded with sensors and do not have adequate processing power for data analysis, nor do they have enough memory to store the huge amounts of data generated. Such tasks need to be outsourced.

Responsiveness:. Health data needs to be continuously monitored, and any rainy day situation needs to be urgently identified and reported to a doctor for quick action. This is of particular relevance in emergency or critical circumstances, such as during a heart attack or a paralytic stroke. The doctor needs to be immediately notified so that he can swiftly issue a treatment. Hence, the entire time duration involving data transmission, analysis, notification, and redressal should be reduced to the minimum.

Data Analysis:. An intelligent system needs to be designed to analyze historical data to identify patterns, diagnose diseases, and prescribe treatments.

Data Sharing:. It is essential to communicate the data and any essential findings or models generated to other practitioners so that they can benefit from one system.

High-Quality Sensors:. There is a need for good quality sensors to collect data related to cardio, temperature, breathing, images, speech etc., efficiently.

Standardization of Data Collected from Different IoT Devices:. Every IoT device has different hardware, and the data format varies correspondingly. There is a need to understand and transmit the correct values of data redeemed from IoT devices. Standardization in the data format in this direction is required.

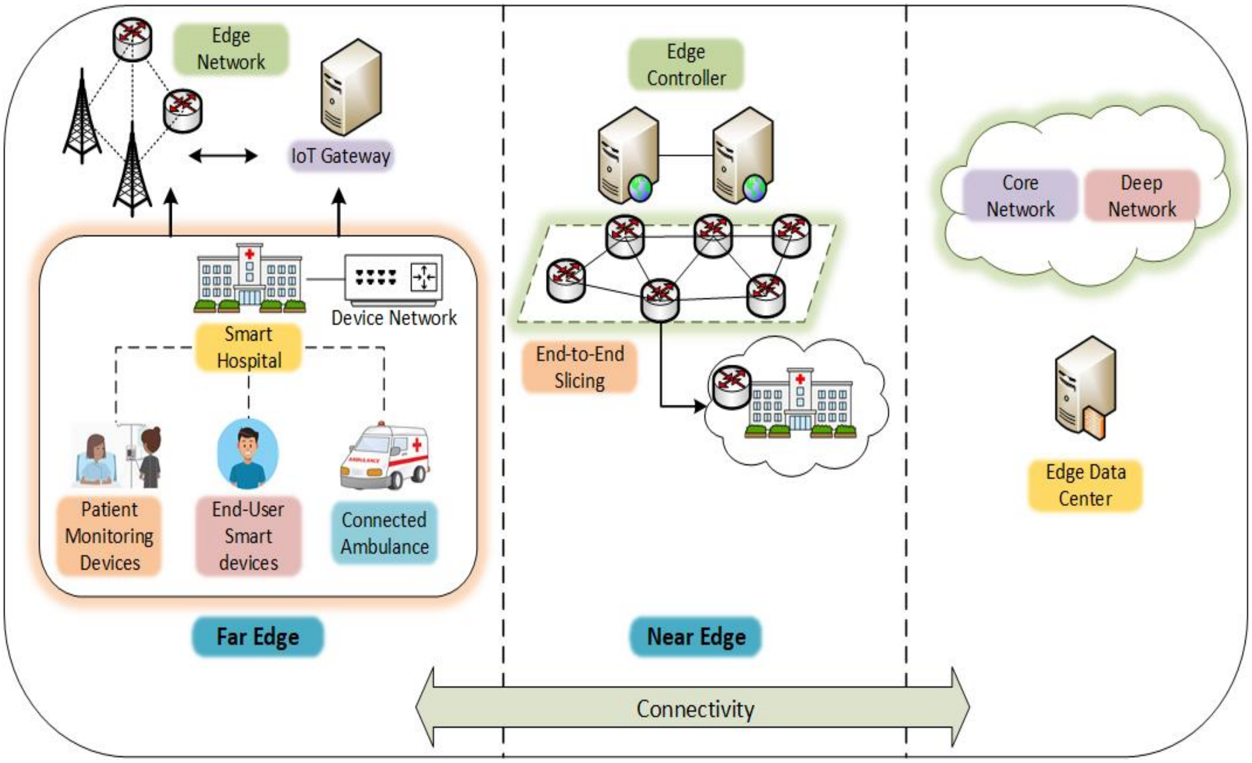

Figure 1: Architecture of proposed edge AI enabled IoT healthcare monitoring system

Appropriate notification mechanisms:. Notification services like short messaging service or emails need to be integrated with the system to inform the doctors or relatives of any urgencies.

Fig. 1 presents the proposed architecture for the IoT healthcare-based network designed to provide efficient and speedy healthcare services. The architecture can be visualized in three segments: Far Edge, Near Edge, and the Data Center.

The corresponding components are described in the following subsections:

3.1. Far Edge

Far Edge comprises the multi-access edge computing infrastructural facilities that are deployed near the hospital but distant from the Data Center. The Far Edge computing infrastructure hosts the IoT edge running IoT applications, including the IoT gateway function.

The hospitals are termed smart hospitals here, and the patients are connected to the Far Edge node using IoT devices. Each patient is equipped with health monitoring devices or sensors that continuously read the patient's vitals and transmit the same information to the Far Edge node. The patients and their relatives also have smart devices like smartphones with applications that are constantly communicating health-related data to the edge node. An ambulance is also available and connected as an essential resource to be deployed in case of emergency. The patient, hospital staff, and the ambulance service are constantly connected through the edge server. The IoT gateway is the device that establishes communication between the IoT sensors and the edge servers. It also ensures data security by providing encryption services.

The scheduler is embedded into the edge node near the hospital. This function, elaborated in subsection 2.4, bears the responsibility for scheduling a patient and providing the required resources to him. The edge node keeps receiving data from the IoT gateway. It monitors whether the received values are within a threshold. Once it is detected that the values corresponding to a particular patient have crossed a threshold, the patient's category and priority are determined. The scheduler then begins allotting resources to the patient with the highest priority. Hence, the scheduler begins to allocate resources in terms of doctors, medicines, etc., to the patient who should be attended to with the maximum priority. Now, if the edge node determines that sufficient resources are not available at the hospital, it forwards the request to the edge controller at the Near Edge. It is now up to the edge controller to ensure service to the patient.

The Far Edge node keeps complete track of all the resources available at the hospital connected to it and also their present status of availability.

3.2. Near Edge

This is the edge computing infrastructure that is deployed between the Far Edge and the cloud Data Centers. The most important component here is the edge controller. The edge controller keeps updated information on the resources available at each hospital connected to it.

The Near Edge bears the responsibility of scheduling the patient and ensuring that appropriate resources are provided to him in as little time as possible. The edge controller monitors different Far Edge nodes that are connected to various hospitals. The edge controller receives a request from a Far Edge node to provide services to a patient, as sufficient resources were not available at that node. The controller then determines which connected node would be able to provide the resources necessary to the patient. Once it finds such a node, the request to service the patient is routed to that particular peer node and the resources from that node are diverted to the patient.

With the edge node and edge controller, all this happens in real time. The service latency is highly reduced with this architecture, with the edge controller bears the responsibility of providing resources. The architecture is also scalable, as several edge nodes can be connected to one edge controller. The edge node would ensure service to the patient through any of these nodes, making the system reliable and responsive.

End-to-end network slicing enables the provision of multiple services on the shared infrastructure at the Near Edge. The different functions at the edge node are implemented in different slices. This helps in delivering Quality of Service slices that bring together different core functions.

3.3. Data Center

The last segment is the edge Data Center that holds the data received from the edge controller. The Data Center contains the core network and the deep network. The core network comprises the communication between the edge controllers and the Data Center. The deep network is the AI layer that can be deployed to apply machine learning models [23] to the data received from different hospitals.

3.4. Proposed algorithm for implementation of the network functions

Algorithm 1: InitializeResourcePool()

```
resourcePool :=arrays of PriorityQueue
for resource Ri do
    M = nextAvailableTime[Ri]
    C = categories_Ri_can_serve[Ri]
    for category Cj do
        INSERT (M,Ri) to resourcePool[Cj]
    end
end
```

Algorithm 2: Categorise(Patient P)

```
Sensors_value_array := read_sensors(P)
for every sensor_value in Sensor_value_array: do
    if sensor_value > Threshold then
        sensor_weight = 1
    else
        sensor_weight = 0
    end
end
for i in range (0, no. of sensors): do
    C = C∨(sensor_weight_array[i] * 2)
end
return C
```

Here, we present the algorithms envisaged behind the edge node functions and edge controller. The algorithms are designed to schedule patients and provision resources efficiently.

Consider the following terminologies:

Algorithm 3: Priority(Patient P)

```
Sensors_value_array := read_sensors(P)
Priority := 0
for every sensor_value in Sensor_value_array: do
    if sensor_value > Threshold then
        Priority+=1
    else
        Priority+= 0
    end
end
return Priority
```

Algorithm 4: Scheduler(Patient P)

```
while patientPool.size() is not 0: do
    P = patientPool.top()
    C = categorise(P)
    R = resourcePool[C].top()
    if R.Time > P.Time then
        P.assignedResource = R.id
        R.Time = R.Time + P.TreatTime
        Alerter(R)
        resourcePool[C].push(R)
    else
        Request_To_Peer(P)
    end
end
```

P represents the set of patients in the hospitals.

S represents the set of sensors attached to the patient.

n is the number of sensors attached to a patient.

J is the set of jobs.

T is the set of threshold values assigned for each sensor.

R is the set of hospital resources that may include doctors, medicines, masks, etc.

C is the set of patient categories that decide which resource needs to be allocated to an individual patient. All patients will be categorized in one of these categories.

Now we present in detail the functions embedded in the edge server that collaboratively assign resources to patients depending on patient priorities and resource availability. We consider N number of hospital nodes that can provide service to patients. The edge server schedules the patients depending on their priority; resources can also be allotted from a peer hospital node if required.

First, the resource pool is initialized by the function InitializeResourcePool. Each resource R_i has an associated time at which the resource will be available (say, the time when the doctor will be free after an operation), termed M, and the set of patient categories that the resource can service (As each doctor can serve patients of a specific category, not others). Here, category refers to the different specialized requirements of a patient. Consider resourcePool to be an array of minimum heap priority queues, where the array index represents a category. For each category, the pair (M, R_i) is inserted into the corresponding resourcePool entry. Hence, the entry resourcePool[C_j] comprises the resources that will be used for that category and the time when each of those resources will be available. The min-heap priority queue is used to enable finding resources for a particular patient category that are available earlier.

$$C = \sum_{i=0}^{n-1} SW_i * (b+1)^i \tag{1}$$

$$C_p = \sum_{i=0}^{4} SW_i * 2^i \tag{2}$$

The function Scheduler() schedules the patients and assigns resources to them as and when required. This function is the most important entity. The Scheduler first selects a patient from the patient pool. The patient pool is stored as a max-heap-based priority queue that contains sets of (Priority, P) as its elements. The max-heap data structure allows treatment of higher priority patients earlier than others. For every patient P, the priority is determined using the Priority function. The patient priority is incremented by one for every sensor value that is greater than the threshold. The pair of Patient ID and that patient's priority is

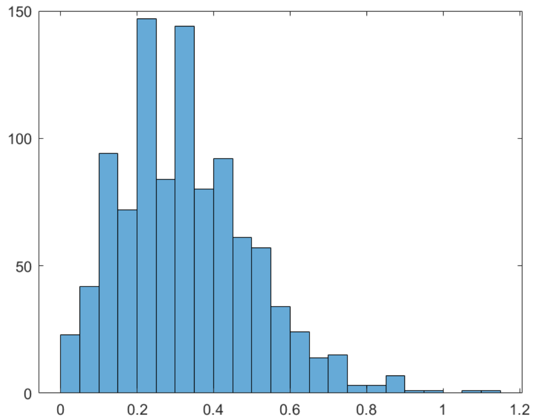

Figure 2: End-to-end time distribution

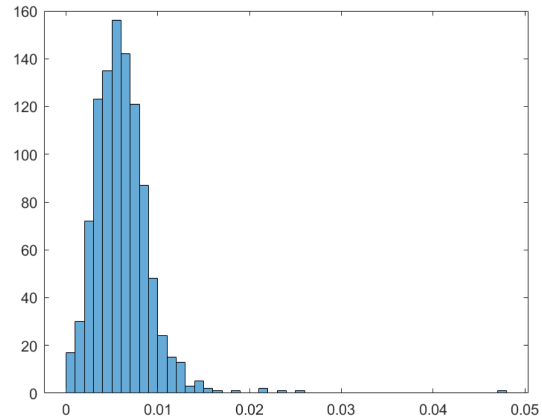

Figure 3: Computing time distribution

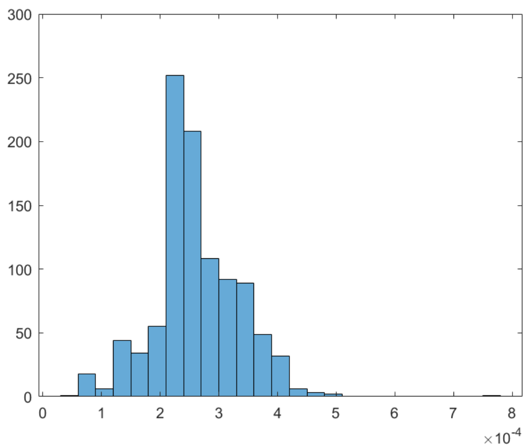

Figure 4: Priority latency time distribution

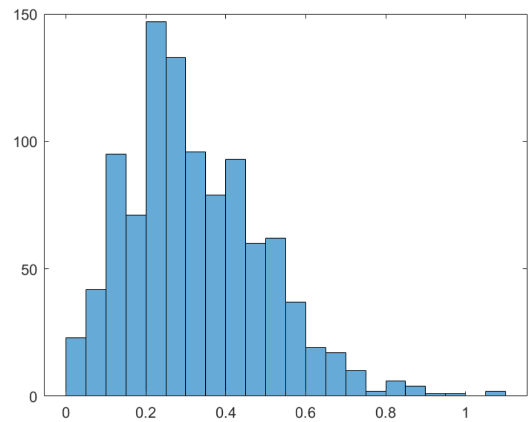

Figure 5: Transmission time distribution

inserted into the pool. The patient with the maximum priority is at the root node and is removed first by the Scheduler. Now the Scheduler determines the category for the patient. For this, the Categorise function is called.

The patient category is found using the formula in equation 1, which gives the patient category for n sensors and b threshold values. This is based on the concept of positional notation [2]. For instance, for one threshold value and five sensors, equation 2 can be used to calculate the patient category.

Each patient is assigned a category depending on the values reported by the sensors $S_1, S_2, ..., S_n$. Let

[2]https://en.wikipedia.org/wiki/Positional_notation

S_{W_i} denote the weight of the i^{th} sensor. This will be set to one if the value is greater than the threshold value for that sensor. For example, a sensor S_i, monitoring a patient's temperature, reports a value of 102 degrees Fahrenheit, and the corresponding threshold was set as 98. The value of S_{W_i} will hence be set to 1. If the temperature is normal, S_{W_i} will be set to 0.

The greater the number of sensors reporting higher values than the threshold, the higher the value C_p for that patient. Now the scheduler consults the resource pool to determine if the necessary resources are available to service the patient. Suppose the value corresponding to resourcePool[C_p] denotes that the necessary resources to service category C_p patient are indeed present and will be available at the needed time.

In that case, the resources are assigned to the patient P. The availability time for the resource R is changed to reflect that the resources are no longer available. The Alerter function is called to send an alert to the corresponding doctors or people who can start the process of resource provisioning.

Alternatively, if the needed resources are not available, then the scheduler sends a request to a peer hospital to provide resources if available. This process may continue until the time the required resources are made available to the patient.

This architecture may prove quite useful in pandemic situations like that of COVID-19. Consider a smart city where people are equipped with sensors that monitor parameters like temperature and SPO2 [24, 25]. The data read from these sensors is continuously transmitted to the edge node that corresponds to the nearest hospital. If a person catches the infection, he will begin to show changes in these vitals. The corresponding values will be analyzed at the edge node, and abnormalities will be detected as thresholds are met or exceeded. The person will hence be categorized, depending on their priority, and will be appropriately scheduled by the Scheduler at an early stage, and required resources like doctors or medicines can be reserved for them. This may help in provisioning treatment early and also cater to a higher number of individuals. Emergency situations or highly critical patients would also be identified early and could be given timely treatment.

4. Results

We performed an experimental simulation to test the performance of our system. We used different sensors to compute the following vital parameters:

- Temperature
- Diabetes Sugar Level
- Heart Rate
- SPO2 (in %)
- SYSTOLIC (mm HG)

- DIASTOLIC (mm HG)

We connected these sensors to Raspberry Pi and Arduino Yun boards, and the data collected were sent to an edge node. For the edge controller, we used a system with 16 GB RAM and the Ubuntu 18.04 operating system. For the cloud, we used Amazon Web Services (AWS). We used three edge nodes representing the hospitals.

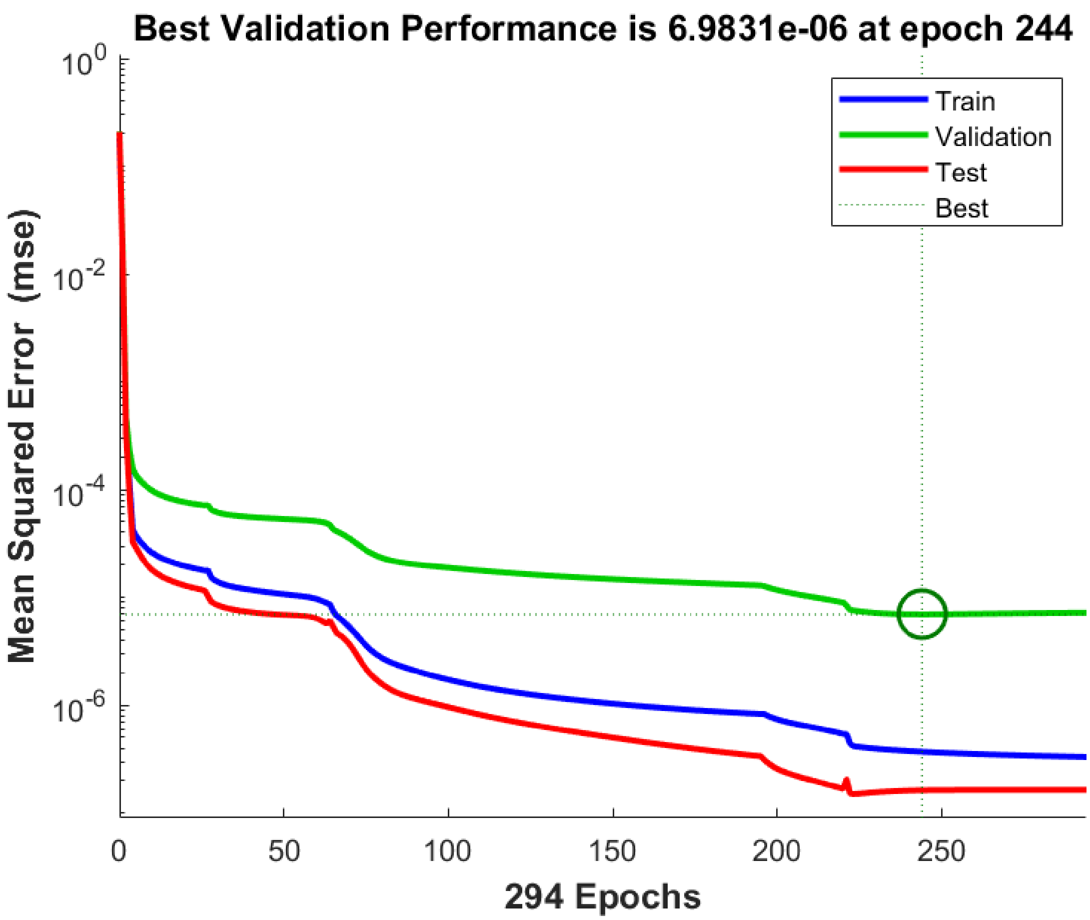

Figure 6: Mean square error for training, validation, and test sets

The following times were computed to analyze the performance of the system.

The end-to-end time corresponds to the end-to-end network slice time—This is the combined sum of the compute latency, prioritizing latency, and transmission latency that is the overall time it takes to service a patient.

The compute time is the total time required for computation at each node, excluding the transmission time, if more than one edge node is being used to service the patient.

The prioritizing time is the time for assigning priority to a patient, depending on his category calculated

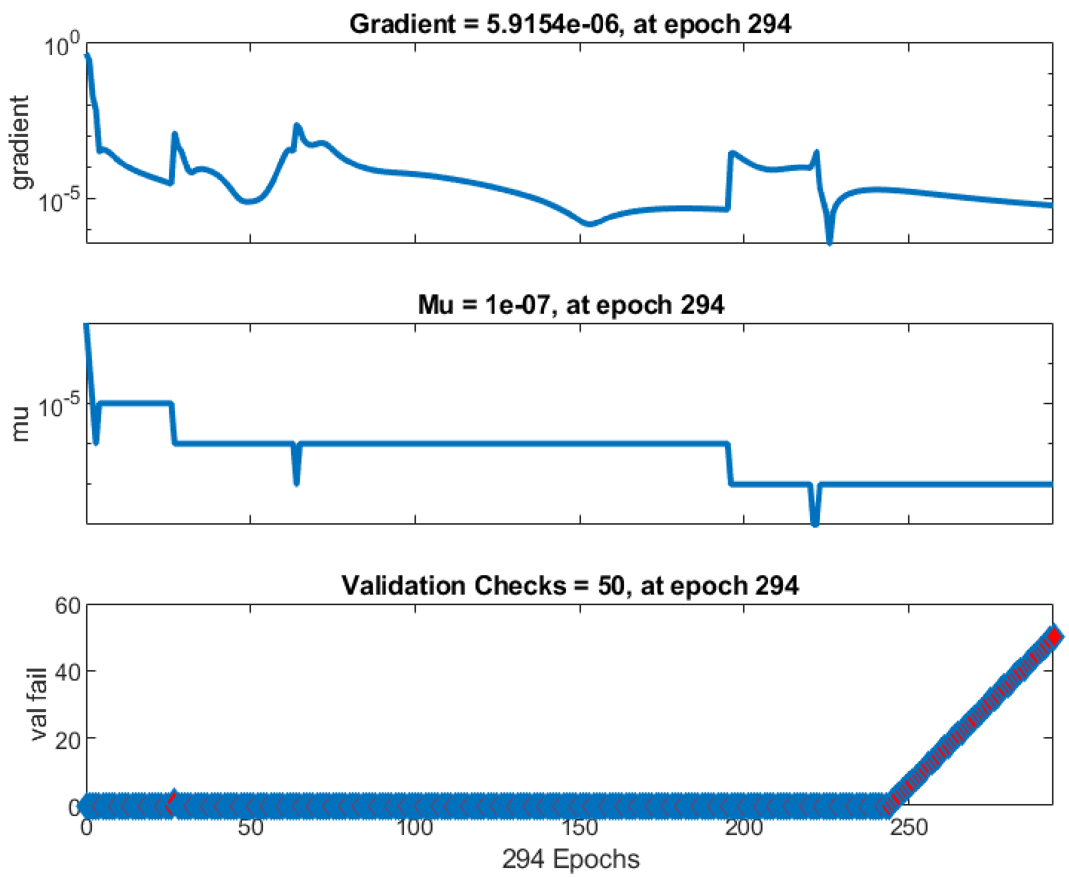

Figure 7: μ, gradient, validation check for failure

using the sensor values.

The transmission time is the time required for the transmission of data between different edge nodes. Figures 2–5 present the distributions for end-to-end time, compute time, prioritizing time, and transmission time. These histograms portray the time distributions obtained from the simulated data.

The end-to-end time was primarily obtained in the range of $0.2 - -0.4$ time units. The compute time was found to be within $0 - -0.05$ time units. The time needed to prioritize a patient was mostly recorded to be $2X10^{-4}$–$3X10^{-4}$. The transmission time between nodes was observed to be within $0 - -1.2$ time units.

The simulated data were generated using three nodes representing three hospitals and a total of a hundred patients. The architecture is scalable and would need to cater to thousands of patients and a large number of hospitals. To replicate a real scenario with many hospitals and patients, we used neural network to model the transmission latency, which further affected the end-to-end time and compute latency.

The neural network was generated using the neural network modeling tool in MATLAB. The technique used was feed-forward backpropagation with ten hidden layers.

Fig. 6 shows that the mean square error is in the range of 10^{-6}. Fig. 7 further provides the validation accuracy of the model.

Fig. 8 demonstrates the accuracy received by fitting a linear model on the data corresponding to the transmission latency. It can be seen that the line fits the data with high accuracy for all the training, validation, and test data sets. The regression value is close to one in all cases.

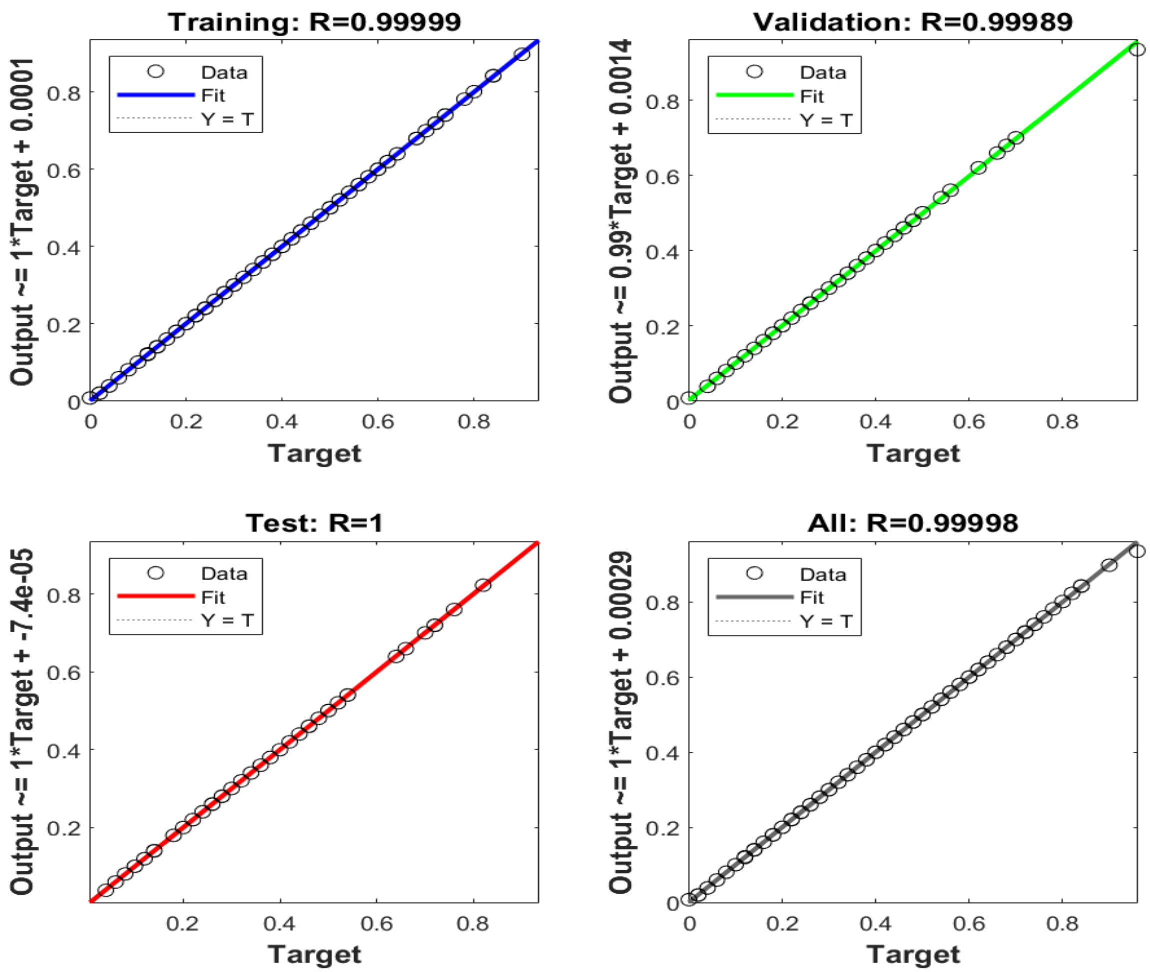

Figure 8: Linear regression fit to the data and corresponding regression values

5. Conclusion

An edge AI enabled IoT healthcare monitoring system for smart cities was proposed in the current work. The system is mainly intended for the real-time scheduling of patients. It provides resources like doctors, medicines, ambulance services, blood, ICUs, and medical equipment like oxygen masks on a priority basis, and reserves resources based on a person's condition or category.

First, the vitals of the patient are collected using AI-enabled IoT devices attached to the patient. This information is stored at an edge server that contains the processing capability and memory requirements for analyzing and storing it. The edge node monitors the data and immediately and runs a scheduler to provision resources from the hospital when required. If the resources are not available, the request is forwarded to an edge controller that provisions resources using other peer nodes connected to other hospitals. In this way, the resources are made available through the architecture in real time. The data is then further communicated to an edge data server where it is stored for later access anytime, anywhere. Machine learning algorithms can be deployed in the deep network to analyze past data, deduce symptoms for diseases, and suggest appropriate treatment in conjugation with an expert specialist. Experimental simulations were also performed, and the time distributions for the same were presented. A neural network was used to model the transmission latency, and a linear fit to the data was reported. This could be used to analyze system performance in a real-world scenario. The security, responsiveness, and analytic capability of the architecture make it an empowered healthcare system design. The system could prove immensely useful for catering to the needs of those who are elderly or disabled, or during pandemic situations when timely help is of utmost importance.

Acknowledgement

This work was supported by the Deanship of Scientific Research at King Saud University, Riyadh, Saudi Arabia, through the Vice Deanship of Scientific Research Chairs: Research Chair of Pervasive and Mobile Computing.

References

[1] H. Lin, S. Garg, J. Hu, X. Wang, M. J. Piran, M. S. Hossain, Privacy-enhanced data fusion for covid-19 applications in intelligent internet of medical things, IEEE Internet of Things Journal (2020) 1–1doi:10.1109/JIOT.2020.3033129.

[2] Y. Shifeng, F. Chungui, H. Yuanyuan, Z. Shiping, Application of iot in agriculture, Journal of Agricultural Mechanization Research 7 (2011) 190–193.

[3] M. S. Hossain, G. Muhammad, Emotion-aware connected healthcare big data towards 5g, IEEE Internet of Things Journal 5 (4) (2018) 2399–2406.

[4] P. Gope, T. Hwang, Bsn-care: A secure iot-based modern healthcare system using body sensor network, IEEE sensors journal 16 (5) (2015) 1368–1376.

[5] M. S. Hossain, Cloud-supported cyber–physical localization framework for patients monitoring, IEEE Systems Journal 11 (1) (2017) 118–127.

[6] S. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, K.-S. Kwak, The internet of things for health care: a comprehensive survey, IEEE access 3 (2015) 678–708.

[7] Y. Hao, et al., Smart-edge-cocaco: Ai-enabled smart edge with joint computation, caching, and communication in heterogeneous iot, IEEE Network 33 (2) (2019) 58–64.

[8] Y. Abdulsalam, M. S. Hossain, Covid-19 networking demand: An auction-based mechanism for automated selection of edge computing services, IEEE Transactions on Network Science and Engineering (2020) 1–1doi:10.1109/TNSE.2020.3026637.

[9] M. S. Hossain, G. Muhammad, N. Guizani, Explainable ai and mass surveillance system-based healthcare framework to combat covid-i9 like pandemics, IEEE Network 34 (4) (2020) 126–132.

[10] M. Ghamari, B. Janko, R. S. Sherratt, W. Harwin, R. Piechockic, C. Soltanpur, A survey on wireless body area networks for ehealthcare systems in residential environments, Sensors 16 (6) (2016) 831.

[11] G. Marques, R. Pitarma, N. M Garcia, N. Pombo, Internet of things architectures, technologies, applications, challenges, and future directions for enhanced living environments and healthcare systems: a review, Electronics 8 (10) (2019) 1081.

[12] C. Pereira, J. Mesquita, D. Guimarães, F. Santos, L. Almeida, A. Aguiar, Open iot architecture for continuous patient monitoring in emergency wards, Electronics 8 (10) (2019) 1074.

[13] G. Muhammad, M. S. Hossain, N. Kumar, Eeg-based pathology detection for home health monitoring, IEEE Journal on Selected Areas in Communications 39 (2) (2021) 603–610.

[14] T. Kadarina, R. Priambodo, Monitoring heart rate and spo2 using thingsboard iot platform for mother and child preventive healthcare, in: IOP Conference Series: Materials Science and Engineering, Vol. 453, IOP Publishing, 2018, p. 012028.

[15] G. Ganesh, K. Jaidurgamohan, V. Srinu, C. R. Kancharla, S. V. Suresh, Design of a low cost smart chair for telemedicine and iot based health monitoring: An open source technology to facilitate better healthcare, in: 2016 11th International Conference on Industrial and Information Systems (ICIIS), IEEE, 2016, pp. 89–94.

[16] W. Zang, F. Miao, R. Gravina, F. Sun, G. Fortino, Y. Li, Cmdp-based intelligent transmission for wireless body area network in remote health monitoring, Neural computing and applications 32 (3) (2020) 829–837.

[17] A. Saeed, M. Faezipour, M. Nourani, S. Banerjee, G. Lee, G. Gupta, L. Tamil, A scalable wireless body area network for bio-telemetry., JIPS 5 (2) (2009) 77–86.

[18] S. U. Amin, et al., Cognitive smart healthcare for pathology detection and monitoring, IEEE Access 7 (2019) 10745–10753.

[19] G. S. Aujla, A. Jindal, A decoupled blockchain approach for edge-envisioned iot-based healthcare monitoring, IEEE Journal on Selected Areas in Communications.

[20] F. Wu, C. Qiu, T. Wu, M. R. Yuce, Edge-based hybrid system implementation for long-range safety and healthcare iot applications, IEEE Internet of Things Journal (2021) 1–1doi:10.1109/JIOT.2021.3050445.

[21] W. Tan, P. Tiwari, H. M. Pandey, C. Moreira, A. K. Jaiswal, Multimodal medical image fusion algorithm in the era of big data, Neural Computing and Applications (2020) 1–21.

[22] P. Tiwari, S. Uprety, S. Dehdashti, M. S. Hossain, Terminformer: unsupervised term mining and analysis in biomedical literature, Neural Computing and Applications (2020) 1–14.

[23] W. Min, B.-K. Bao, C. Xu, M. S. Hossain, Cross-platform multi-modal topic modeling for personalized inter-platform recommendation, IEEE Transactions on Multimedia 17 (10) (2015) 1787–1801.

[24] A. Yassine, et al., Iot big data analytics for smart homes with fog and cloud computing, Future Generation Computer Systems 91 (2019) 563 – 573.

[25] M. S. Hossain, G. Muhammad, A. Almari, Smart healthcare monitoring: a voice pathology detection paradigm for smart cities, Multimedia Systems 25 (5) (2019) 565–575.

Author biographies

Vipin Kumar Rathi is an assistant professor at Ramanujan College, University of Delhi. He received his M.Tech. degree in computer science from Jawaharlal Nehru University, New Delhi, India . Currently he is chairperson of the Hyperledger Telecom Special Interest Group. His research interests include edge computing, Blockchain, cloud computing, 5G, IoT , Post Quantum Cryptography and machine learning techniques.

Nikhil Kumar Rajput received his PhD degree in 2015 in computer science from Jawaharlal Nehru University, New Delhi. He is presently working as an Assistant Professor at Ramanujan College (University of Delhi). His area of interest includes blockchain, computational linguistics, immersive technology, IoT, machine learning and information theory.

Shubham Mishra is currently an SDE intern at Postman while Pursuing his bachelors of Computer Science from Ramanujan College, Delhi University. He is interested in 5g, NFV and cloud native. He is also involved in open source projects and Spend majority of his bachelors time in contributing to organisations - Anuket, KDE.

Bhavya Ahuja Bhavya Ahuja Grover pursued MSc in Computer Science from University of Delhi in 2011. She is currently working as an Assistant Professor in Department of Computer Science, Ramanujan College (University of Delhi) Her areas of interest include cryptography, blockchain, IoT, computational linguistics and machine learning.

Prayag Tiwari is working as a Postdoctoral Researcher at the Aalto University. Previously, he was working as a Marie Curie Researcher and Ph.D. student at the University of Padova, Italy. He received his MS Degree from NUST MISIS, Moscow. He also worked as a Research Assistant at NUST MISIS, and he has had Teaching and Industrial work experience. He has several publications in top journals and conferences. His research interests include Machine Learning, Deep Learning, Quantum-Inspired Machine Learning, Information Retrieval, Health Informatics, and IoT.

Amit Kumar Jaiswal is a final year PhD student as well as a Marie Curie Researcher at the University of Bedfordshire, working on Quantum Information Retrieval. He is also a Research Software Engineer in School of Mathematics at the University of Leeds. His research spans a wide range of topics in areas, including existing Quantum-IR frameworks, applied machine learning, and their applications in Web search, medical imaging and financial domain in general.

M. Shamim Hossain is currently a Professor in the Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia. He is also an adjunct professor with the School of Electrical Engineering and Computer Science, University of Ottawa, Canada. He received his Ph.D. in Electrical and Computer Engineering from the University of Ottawa, Canada in 2009. He is the chair of the IEEE Special Interest Group on Artificial Intelligence (AI) for Health with IEEE ComSoc eHealth Technical Committee. He is on the editorial board of the IEEE Transactions on Multimedia, IEEE Multimedia, IEEE Network, IEEE Wireless Communications, IEEE Access, Journal of Network and Computer Applications (Elsevier), and the International Journal of Multimedia Tools and Applications (Springer). He is a senior member of both the IEEE and ACM. He is also an IEEE ComSoc Distinguished Lecturer (DL).