Digital Communications and Networks 10 (2024) 75–82

Contents lists available at ScienceDirect

Digital Communications and Networks

journal homepage: www.keaipublishing.com/dcan

An artificial intelligence diabetes management architecture based on 5G

Ruochen Huang [b], Wei Feng [c], Shan Lu [d], Tao shan [d], Changwei Zhang [e], Yun Liu [a,c,*]

[a] Department of Geriatrics, The First Affiliated Hospital of Nanjing Medical University, Nanjing, 210096, China
[b] Department of Information, The First Affiliated Hospital of Nanjing Medical University, Nanjing, 210096, China
[c] Department of Medical Informatics, School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, 211166, China
[d] Department of Outpatient, The First Affiliated Hospital of Nanjing Medical University, Nanjing, 210096, China
[e] Jiangsu Key Laboratory of Wireless Communications, Nanjing University of Posts and Telecommunications, Nanjing, 210003, China

ARTICLE INFO

Keywords:
Diabetes
5G
Artificial intelligence
Deep forest
Smart hospital ward

ABSTRACT

Along with the development of 5G network and Internet of Things technologies, there has been an explosion in personalized healthcare systems. When the 5G and Artificial Intelligence (AI) is introduced into diabetes management architecture, it can increase the efficiency of existing systems and complications of diabetes can be handled more effectively by taking advantage of 5G. In this article, we propose a 5G-based Artificial Intelligence Diabetes Management architecture (AIDM), which can help physicians and patients to manage both acute complications and chronic complications. The AIDM contains five layers: the sensing layer, the transmission layer, the storage layer, the computing layer, and the application layer. We build a test bed for the transmission and application layers. Specifically, we apply a delay-aware RA optimization based on a double-queue model to improve access efficiency in smart hospital wards in the transmission layer. In application layer, we build a prediction model using a deep forest algorithm. Results on real-world data show that our AIDM can enhance the efficiency of diabetes management and improve the screening rate of diabetes as well.

1. Introduction

With the exponential growth of the Internet in recent years, network-based technologies such as 5G have begun to have an impact on eHealth. The ultra-low latency and ultra-high reliability of these technologies, as well as next-generation technologies such as cross-modal interaction, the healthcare industry has been transformed from traditional healthcare to smart healthcare [1]. According to Ericsson, healthcare has become decentralized, moving from hospitals towards homes while patient data has become centralized, turning hospitals into data centers [2]. Patients can get real-time healthcare responses from personalized healthcare systems [3], and physicians can get massive data from different kinds of sensors to evaluate patients and provide personalized therapeutic schedules. Due to its potential, many researchers expect 5G to continue to influence alter the existing healthcare industry profoundly in the future [4].

Diabetes is a common chronic disease whose global incidence has been rising in recent years. It may occur when the pancreas is no longer able to produce insulin or when the body cannot make good use of the insulin that it does produce. The International Diabetes Federation (IDF) published the 9th edition of the IDF Diabetes Atlas in 2019 wherein they estimated that 463 million people suffer from diabetes worldwide, and this number is projected to reach 578 million by 2030 and 700 million by 2045 [5]. The complications of diabetes can be divided into two types, acute complications such as Diabetic Ketoacidosis (DKA) and diabetic Hyperglycemic Hyperosmolar State (HHS), and long-term complications such as nephropathy, retinopathy, diabetic foot, cardiovascular disease, and stroke. Because diabetes is a chronic disease, patients with diabetes must perform many self-management tasks several times a day to control their Blood Glucose (BG) levels.

Effective management and care of diabetes are crucial to reduce the risk of complications [6]. As a result, several diabetes management systems have been created to help patients to optimize their lifestyles [4–8]. In Ref. [7],the authors built a system to improve cardiovascular risk of patients with diabetes by optimizing patient care workflows,and in Ref. [8], the authors proposed a diabetes management system to manage workflows for healthcare practitioners. The authors built a specialized platform for diabetic patients to prevent long-term complications by empowering their independence. In Ref. [9], an intelligent mobile diabetes management system was developed to control hemoglobin levels in patients and to improve them with disease management plans.

Despite the tremendous research efforts and existing systems for

* Corresponding author. Department of Geriatrics, The First Affiliated Hospital of Nanjing Medical University, Nanjing, 210096, China.
E-mail addresses: huangruochen@outlook.com (R. Huang), liuyun@njmu.edu.cn (Y. Liu).

https://doi.org/10.1016/j.dcan.2022.09.004
Received 18 August 2021; Received in revised form 27 July 2022; Accepted 5 September 2022
Available online 17 September 2022

diabetes management, most of the systems only consider long-term complications. However, the signs and symptoms of acute complications often develop quickly,so it is very important for a diabetes management system to monitor and make appropriate decisions automatically in real time.

Due to the defects of traditional diabetes management systems and the advantages of 5G technology, we think that it is necessary to introduce 5G into personalized healthcare systems for diabetes. With this goal, we propose an Artificial Intelligence Diabetes Management architecture (AIDM) that consists of five layers. In each layer, the AIDM supports low latency, high reliability, and a full range of medical services for patients using 5G technology. To the best of our knowledge, AIDM is the first diabetes management system for both acute complications and chronic complications that utilizes 5G. We summarize the contributions of this research below.

1) In order to realize real-time monitoring and provide early warnings of acute complications, AIDM uses edge computing in its computing layer to reduce the shortcomings of high delay and low reliability caused by previous early warning models in the cloud so as to help patients get timely attention and medical services when there are signs or symptoms of acute complications.
2) In AIDM's transmission layer, we use the Narrowband Internet of Things (NB-IoT) to allocate wireless spectrum resources and allow large-scale equipment access on the basis of ensuring diversified Quality of Experience (QoE) requirements. In our test bed, we apply a delay-aware RA optimization based on a double-queue model to improve the access efficiency of devices in smart hospital wards.
3) Existing diabetes management systems only collect and analyze data in a traditional way that is complex and prone to discrepancies. Therefore, AIDM uses AI-based applications in its application layer to help physicians quickly analyze data and make better decision for patients. In our test bed, we build a prediction model on real-word data to screen for early diabetes.

The rest of this paper is organized as follows. Section 2 provides a review of the key features of 5G technology for diabetes management. Section 3 describes the details about AIDM. Section 4 introduces delay-aware RA optimization for devices in smarts hospital wards in the transmission layer. Section 5 builds a prediction model for screening patients with diabetes in the application layer. Section 6 builds a test bed to evaluate the algorithm in Section 4 and prediction model in Section 5. Finally, Section 7 concludes.

2. Key features of 5G technology for diabetes management

We expect the QoE of diabetes management systems to play a vital role in diagnosis and treatment in future diabetes management systems. Traditional systems focus on wireless solutions for monitoring Fasting Plasma Glucose (FPG) or 2-h plasma glucose with 4G or WiFi, which have limitations to meeting the diverse bandwidth needs of diabetes management. However, the key features of 5G technology such as ultra-low latency, high bandwidth, and ultra-high reliability can improve the QoE of diabetes management systems in the future. We briefly summarize these features below.

Feature 1:Ultra-Reliable Low-Latency Communication (URLLC). According to 3GPP, URLLC is a set of features that provide low latency and ultra-high reliability for mission critical applications such as industrial internet, smart grids, remote surgery, and intelligent transportation systems [10], and URLLC is also critical for acute complications in diabetes management systems. For example, when a patient's blood sugar suddenly rises abnormally because of improper diet, the system can immediately detect this condition and feed it back to physicians. The physicians can immediately obtain the corresponding diet information and give corresponding remote guidance with URLLC.

Feature 2: enhanced Mobile Broadband (eMBB). Bandwidth is the maximum rate of network communications on a line per unit time. At present, the bandwidth for medical signal data transmission in 4G networks is often very limited, especially in real-time applications [11]. Diabetic patients often suffer from various complications such as diabetic retinopathy, diabetic neuropathy, and diabetic lower extremity vascular disease. A 5G network can allocate bandwidth flexibly to medical services and support higher transmission rates than a 4G network. In a 5G network, physicians can see pictures in more detail and patient-physician communication can be improved by Augmented Reality (AR) and Virtual Reality (VR) technology.

Feature 3: Massive Machine-Type Communication. In future diabetes management systems, a large number of medical sensors may be deployed to ensure the reliability of the entire system with low latency characteristics, especially for acute complications of diabetes [12]. However, the large-capacity transmission process that would result from this could exceed the capacity of the network, so a new access technology is needed to support massive sensor access and data capacity. The communication infrastructure in 5G is flexible and scalable for adapting to the diverse needs of medical sensors. It also has stronger device connection capabilities than 4G networks.

3. Artificial intelligence diabetes management architecture

In this section, we proposed the AIDM structure shown in Fig. 1 that consists of five layers: the sensing layer, the transmission layer, the storage layer, the computing layer and the application layer.

3.1. The sensing layer

The sensing layer is divided into real-time monitoring sensors and nonreal-time monitoring sensors. The main feature of the real-time monitoring sensors is that the data sheet is collected frequently and uploaded immediately once the collection is completed. However, the amount of data collected is generally small, and it primarily formatted as text to include information about blood glucose, blood pressure, body temperature, electrocardiography, and blood oxygen. The main characteristics of the nonreal-time monitoring sensors are low data acquisition frequency and large data units that are insensitive to delay.

The central function of this layer is to collect and manage massive amounts of data through various distributed sensors. Considering the variety of IoT devices used in the process of collecting patient data, including those based on 4G and WiFi, we construct the sensing layer based on the current home IoT network structure and 5G networks.

3.2. The transmission layer

With 5G, a huge amount of medical IoT devices can connect to a base station. Hence, AIDM must use reliable technology. Especially for clinical care and nursing homes, broadband spectrum resources are scarce when there are numerous patients and medical sensors. The communication layer of AIDM therefore needs a technology to support massive sensors at low cost with low power consumption in order for it to provide enhanced coverage.

The NB-IoT is a 3GPP-licensed Low Power Wide Area (LPWA) technology that has attracted extensive research recently. In the NB-IoT system, the complexity of devices are low, reducing costs. In addition, many batteries' capacities for wearable sensors are limited to be lightweight and miniature. This is convenient because low power consumption is necessary for medical systems. The overarching design objective of NB-IoT is to provide prolonged battery life. For network coverage, NB-IoT is compatible with existing cellular networks, which is also necessary to effect widespread and consistent key parameter monitoring. Furthermore, NB-IoT is slated to meet the different Quality of Service (QoS) requirements from different users in its next-step evolution. In AIDM, NB-IoT is a key resource to boost capacity and performance for a large number of medical IoT devices.

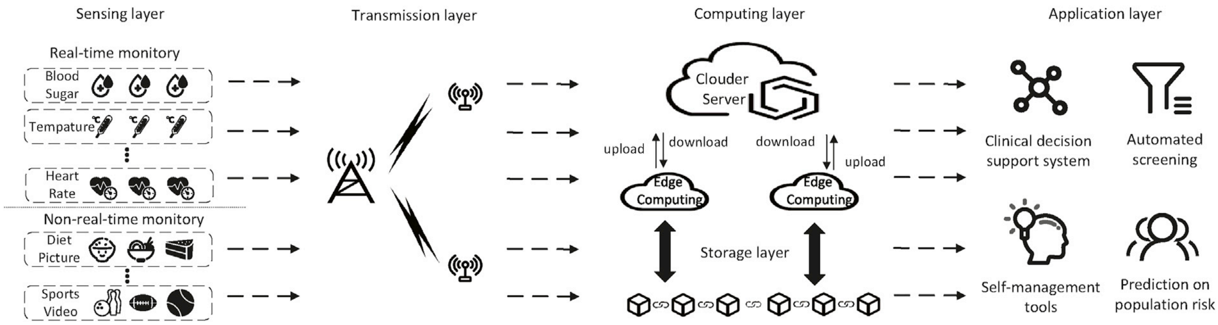


Fig. 1. An artificial intelligence diabetes management architecture based on 5G.

3.3. The storage layer

To realize low delay and high reliability for the whole AIDM system, distributed storage is critically important. Distributed storage can store related information locally, which helps computing models in edge nodes access the data in a timely manner and monitor patient conditions. However, eHealth data contain personal and sensitive information that is important for both patients and physicians. Distributed storage can meet the high requirements of eHealth data security, especially privacy and integrity requirements [13].

A blockchain is a distributed database that has a lot of unique advantages such as absence of centralized control, an ultra-high degree of anonymity, and a consensus that is reached and distributed without centralized authority. Blockchain ensures the integrity, anti-tampering, and traceability of eHealth data. Moreover, eHealth data stored in different systems become information islands. A diabetic may have many healthcare service providers including primary care physicians, specialists, and therapists, and blockchain technology can be used to enable effective data sharing between these healthcare service providers. For example, a diabetic with acute complications may be taken to the hospital in comatose state, and it may be difficult for a physician to make a diagnosis immediately without related health information. Therefore, blockchain can play an important role for ADIM in allowing the secure sharing of eHealth information.

3.4. The computing layer

The computing layer is the core layer of AIDM. This layer provides core computing resources and AI models for AI-based applications in the application layer using training data from the storage layer. In the traditional cloud-centric approach, data collected by IoT devices is uploaded and processed centrally in a group of cloud servers or data centers [14]. This approach involves long propagation delays and sometimes incurs unacceptable latency. There are huge amount of data collected by IoT devices, and the transfer of data to the cloud for training and analysis stresses the backbone of the networks. In edge computing, the computing tasks are deployed in the edge nodes, which are close to where the data is produced. Edge computing can reduce the communication delay and the computing load of cloud data centers. However, in the scenario of diabetes management, privacy-sensitive patients may be unwilling to share privacy data to edge nodes. Therefore, existing edge computing strategies may not be available for diabetes management.

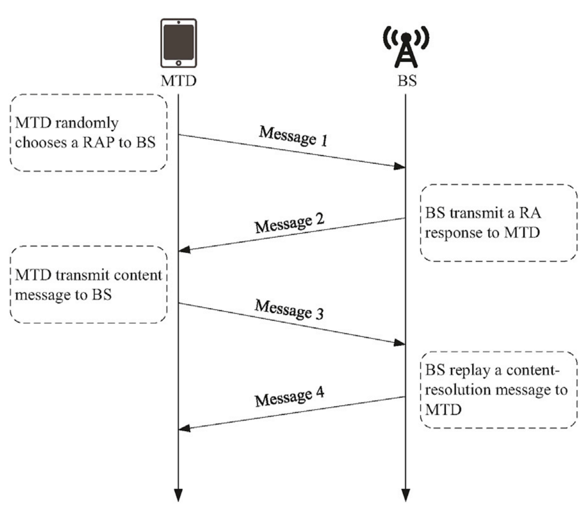


Fig. 2. The random access procedure in the cellular network.

Federated Learning (FL) can be an enabling technology for Machine Learning (ML) model training at mobile edge networks [15].In FL, mobile devices use their local data to train an ML model required by an FL server cooperatively and then update the model to the FL server.These steps are repeated several times until the model training is finished so that ML models can be trained using a mobile protocol with FL.

In summary, there are several advantages when using FL for AIDM.-First, it can reduce data communication costs.Local data can be processed in edge nodes and only some data must be transmitted to the cloud. Second, it enhances patient privacy and reduces the probability of data leakage.In fact, with enhanced privacy, more users may elect to take part in diabetic management. Third, FL is deployed in edge nodes and real-time decisions can be made locally. Therefore, the system can achieve lower overall latency.This is vital for patients in the early stages of acute complications.

3.5. The application layer

The Application layer is a container for all applications. Many AI-based applications are applied in this layer such as automated screening, clinical decision support systems, and self-management tools for patients. Patients, physicians, and research institutions can access AI-based applications with user interfaces, and consistent interactions across this layer can improve the performance of applications, which can provide more personalized service.

For example, patients can input their diet and exercise not only using structured data but also using unstructured data such as photos, videos, and voice. With pattern recognition, AI-based applications can transform unstructured data into structured personalized health profiles such as knowledge graphs. Different applications can share the knowledge and render appropriate decisions for patients. Structured personalized health profiles can also be easily transformed into scientific data and formats for both physicians and researchers.

The AI in AIDM is based on 5G features, which gives it inherent advantages over existing 4G and WiFi systems. With a sensing layer and a transmission layer, our system can collect large-scale data timely with massive amounts of IoT devices, which can improve the accuracy of AI models in the computing layer. Due to the use of a blockchain in the storage layer and FL in the computing layer, our system provides a safe environment for patients to share their data.Furthermore, patients and

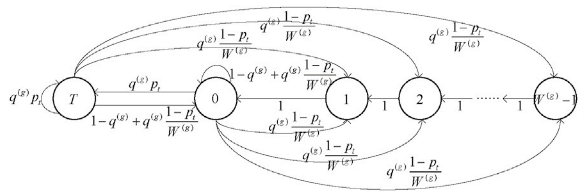


Fig. 3. State transition diagram of each access request in group g.

physicians can achieve better user experiences with 5G features. Routine screening and detection are important for complications of diabetes. The transmission layer can support massive IoT devices which offer more information than ever before. In the future, more and more medical IoT devices are set to be deployed in nursing homes and hospitals for screening and detection purposes. For acute complications, the computing layer of AIDM provides ultra-high reliability and low latency. When a patient is in a life-threatening medical emergency, a physician can obtain information and results from edge nodes immediately.For chronic complications, lifestyle monitoring is very important. With AIDM patients can update and share their pictures and videos about their life, allowing physicians to evaluate patients' relationships between lifestyle and chronic complications.

4. Delay-aware RA optimization in the transmission layer

With the development of the IoT, more and more medical sensors are beginning to be applied in hospitals and nursing homes. There are many studies on improving the access efficiency of such sensors in massive access scenarios. In Ref. [16], the authors propose an Access Class Barring (ACB) factor determination method to improve the success rate when Base Stations (BS) become backlogged with Machine-type Devices (MTDs). Specifically, with their ACB algorithm, before performing the RA procedure, each MTD first randomly generates a value between 0 and 1. If this value is smaller than the ACB factor broadcast by the BS, the MTD can perform the RA procedure immediately. Otherwise, the RA gets temporarily postponed. The authors do not consider the priority of MTDs. Another way to improve the access efficiency of MTDs is to divide them into several groups and each with a different QoS requirement [17–19]. In Ref. [20], the authors propose a double-queue model for the RA procedure and they model the request queue of each MTD with a Geo/G/1/1 queue. However, they pay little attention to satisfying the delay requirements of delay-sensitive MTDs in a real-world scenario. In this subsection, we use a multigroup analytical framework proposed in Ref. [21] to satisfy the delay requirements of delay-sensitive applications while optimizing the throughput of delay-tolerant applications in a smart hospital ward.

Fig. 2 shows the logical sequence of the RA procedure. In message 1, an MTD randomly chooses a Random Access Preamble (RAP) and transmits the request to the BS. If multiple MTDs choose the same RAP, a collision occurs and the BS sends the same RA response to the MTDs. In message 2, the BS transmits an RA response to the MTDs that contains the information on the uplink channel for the third step of the RA procedure. In message 3, after receiving the RA response, the MTDs transmit the connection message. All MTDs that select the same RAP in the first step transmit the same connection message over the same resources to the BS, and this leads to failure of the RA procedure. In message 4, the connection requests from the MTDs are received by the BS and it replies with a contention-resolution message to acknowledge the success of the RA procedure. From 2, we see that the first step is very important for the RA procedure.

We now discuss the collision in the first step. Consider n MTDs in a cell and that all MTDs are divided into M groups for different QoS. Let $n^{(g)}$ mean the number of MTDs in the gth group. $q^{(g)}$ means ACB factor, and $W^{(g)}$ mean the Uniform Backoff (UB) window size of the MTDs in the gth group such that $g = \{1, \ldots, M\}$, where M denotes the number of groups. Fig. 3 shows the state transition diagram of each access request in the gth group. From this diagram, we can formulate the steady-state probability distribution of the access requests in the gth group:

$$\begin{cases} \pi_T^{(g)} = \left(\dfrac{1}{q^{(g)}p} + \dfrac{(1-p)(W^{(g)}-1)}{2p}\right)^{-1} \\ \pi_0^{(g)} = \dfrac{1-q^{(g)}p}{q^{(g)}p}\pi_T^{(g)} \\ \pi_j^{(g)} = \dfrac{(1-p)(W^{(g)}-j)}{pW^{(g)}}\pi_T^{(g)}, j = 1, 2, \ldots, W^{(g)}-1 \end{cases} \tag{1}$$

The throughput of the gth group i, $\hat{\lambda}_{\text{out}}^{(i)}$ and the network throughput $\hat{\lambda}_{\text{out}}$ using the Geo/G/1/1 model are as follows:

$$\hat{\lambda}_{\text{out}}^{(i)} = \frac{\hat{\lambda}^{(i)}}{\frac{\hat{\lambda}^{(i)}}{n^{(i)}}\left(\frac{1}{q^{(i)}p} + \frac{(1-p)(W^{(i)}-1)}{2p}\right) + 1} \tag{2}$$

and

$$\hat{\lambda}_{\text{out}} = \sum_{i=1}^{M} \frac{\hat{\lambda}^{(i)}}{\frac{\hat{\lambda}^{(i)}}{n^{(i)}}\left(\frac{1}{q^{(i)}p} + \frac{(1-p)(W^{(i)}-1)}{2p}\right) + 1} \tag{3}$$

From (2) and (3), we can see that the success probability of access requests p is important to $\hat{\lambda}_{\text{out}}^{(i)}$ and $\hat{\lambda}_{\text{out}}$. As shown in Fig. 2, an MTD builds a connection with BS in the RA procedure as long as all other MTDs have the following.

1. failure of ACB check
2. empty request queue
3. backoff procedure

The success probability of access request p can be found by solving the following equation (4):

$$\begin{aligned} p &= \left(1-\rho^{(g)}+\rho^{(g)}\left(\sum_{i=1}^{W^{(g)}-1}\pi_i^{(g)} + \left(1-q^{(g)}\right)\left(\pi_0^{(g)}+\pi_T^{(g)}\right)\right)\right)^{n^{(g)}-1} \cdot \\ &\prod_{j=1, j\neq g}^{M}\left(1-\rho^{(j)}+\rho^{(j)}\left(\sum_{i=1}^{W^{(j)}-1}\pi_i^{(j)} + \left(1-q^{(j)}\right)\left(\pi_0^{(j)}+\pi_T^{(j)}\right)\right)\right)^{n^{(j)}} \\ &\approx \exp\left(-\frac{n^{(g)}\rho^{(g)}\pi_T^{(g)}}{p}\right)\cdot\prod_{j=1, j\neq g}^{M}\exp\left(-\frac{n^{(j)}\rho^{(j)}\pi_T^{(j)}}{p}\right) \\ &= \exp\left(-\sum_{i=1}^{M}\frac{\hat{\lambda}^{(i)}}{\frac{\hat{\lambda}^{(i)}}{n^{(i)}}\left(\frac{1}{q^{(i)}}+\frac{W^{(i)}-1}{2}\right)+p\left(1-\frac{\hat{\lambda}^{(i)}(W^{(i)}-1)}{2n^{(i)}}\right)}\right) \end{aligned} \tag{4}$$

To analyze the delay performance, a request in the backoff procedure can be expressed as State B which is a combination of State i, $i \in \{1, \ldots, W^{(g)}-1\}$. Simplifying all states into three types of states in group g: State T, State B and State 0, we define $D_j^{(i)}, j = T, 0, B$, as the whole time from State j to State T in the ith group and define $D_T^{(i)}$ as the access delay of the access requests in the ith group. Thus, the mean access delay of the access requests in the ith group can be obtained as

$$E\left[D_T^{(i)}\right] = q^{(i)}p + (1+D_0^{(i)})\left(1-q^{(i)}+q^{(i)}\frac{1-p}{W^{(i)}}\right) + (1+D_B^{(i)})q^{(i)}\frac{(1-p)(W^{(i)}-1)}{W^{(i)}} = \frac{1}{q^{(i)}p} + \frac{\left(W^{(i)}-1\right)(1-p)}{2p} \tag{5}$$

and the mean access delay of all MTDs is given by

$$E[D_T] = \frac{\sum_{i=1}^{M} E\left[D_T^{(i)}\right]\cdot n^{(i)}}{\sum_{i=1}^{M} n^{(i)}} \tag{6}$$

Finally, for each group, the mean access delay of access requests in the ith group is given by

$$E\left[D_T^{(i)}\right] = \frac{n^{(i)}}{\hat{\lambda}_{\text{out}}^{(i)}} - \frac{1}{\lambda^{(i)}} \tag{7}$$

In this paper, we split all MTDs into two groups: a medical group for delay-sensitive MTDs and an environment group for delay-tolerant MTDs in order to optimize the access efficiency of the MTDs in environment group under the delay constraints of MTDs in the medical group. For MTDs in the medical group, we assume the delay constraints to be $E\left[D_T^{(j)}\right] \le D^{(j)}$, $j \in \{1,2,\dots,M-1\}$. For the MTDs in the environment group, the throughput is $\hat{\lambda}_{\text{out}}^{(M)}$. This leads to the following optimization problem:

$$\max_{\{q^{(i)},W^{(i)}\}} \quad \hat{\lambda}_{\text{out}}^{(M)} s.t. \quad E\left[D_T^{(j)}\right] \le D^{(j)}, \quad j \in \{1,2,\dots,M-1\} \tag{8}$$

In [21], we find that to achieve the optimal network throughput for the environment group under the delay constrains of the medical group, the backoff parameters of each MTD ($q^{*,(i)}$, $W^{*,(i)}$) should be chosen carefully according to

$$\begin{cases} \dfrac{e}{q^{*,(i)}} + \dfrac{(W^{*,(i)}-1)(e-1)}{2} = D^{(i)}, \ i = 1,\dots,M-1 \\[2ex] \dfrac{\hat{\lambda}^{(M)}}{\dfrac{\hat{\lambda}^{(M)}}{n^{(M)}}\left(\dfrac{1}{q^{*,(M)}} + \dfrac{(1-e^{-1})\left(W^{*,(M)}-1\right)}{2}\right) + e^{-1}} = 1 - \sum_{j=1}^{M-1} \dfrac{\hat{\lambda}^{(j)}e}{\dfrac{\hat{\lambda}^{(j)}}{n^{(j)}}D^{(j)} + 1} \end{cases} \tag{9}$$

In multi-RAP scenarios, we define $n^{(i),(j)}$ as the number of MTDs in the ith group that choose the jth RAP. Therefore, the optimal backoff parameters of each MTD are

$$\begin{cases} \dfrac{e}{q^{*,(i)}} + \dfrac{(W^{*,(i)}-1)(e-1)}{2} = D^{(i)}, \ i = 1,\dots,M-1 \\[2ex] \dfrac{n^{(M)}\lambda^{(M)}/N}{\lambda^{(M)}\left(\dfrac{1}{q^{*,(M)}} + \dfrac{(1-e^{-1})\left(W^{*,(M)}-1\right)}{2}\right) + e^{-1}} = 1 - \sum_{j=1}^{M-1} \dfrac{n^{(j)}\lambda^{(j)}e/N}{\lambda^{(j)}D^{(j)} + 1} \end{cases} \tag{10}$$

5. The application layer prediction model for screening patients with diabetes

In a personalized diabetes management system, it is very important to design an efficient prediction algorithm for a patient's diabetes status. At present, neural networks have been widely used in various fields, and many industrial frameworks, such as tensorflow, keras, and pytorch, have been applied in the screening and prediction of diabetes and its complications.

However, neural networks usually need strong cloud computing platforms for training. Considering that AIDM is primarily distributed among edge nodes, we implement a deep forest algorithm as shown in Fig. 4 to predict the status of a patient with diabetes.

The Multi-grained Cascade Forest (gcForest) uses a cascade forest structure and multi-grained scanning that can work well with small-scale training data and has achieved highly competitive results on both classification and sequence tasks [22,23]. Specifically, a gcForest consists of two modules, a fine-grained module and a cascading module.

First, raw training data are transferred into the process of the multi-grained scanning. In fine-the grained module raw features are scanned by sliding windows. If there are sequential feature relationships in the raw features, the multi-grained scanning can extract information with different window sizes. Let there be k dimensional features in the raw data and a window size of w dimensional features. If the step of the window is S, then the number of instances I_{num} is

$$I_{num} = \frac{k+w}{S} + 1 \tag{11}$$

After this, each instance is applied in training a random forest and a completely-random forest, and then each forest discriminates a result as the class vectors. The transformed features as the output of the fine-grained module are then concatenated by the class vectors.

Second, in the cascading module, a multi-layer structure is built. The structure helps gcForest to learn from representation data using a model built with layers. Each layer consists of several ensemble learners, and each ensemble learner is a branch of a decision tree. For diversity, the cascading module uses completely-random forests or random forests, which can alleviate overfitting. The forest in each layer can get feature information from previous layers and use them as input features, and the output of all these forests in this layer are concatenated with the original features to be input for the next layer. Then, cross-validation is used in order to avoid overfitting. When there are no more new layers, the whole structure is tested in the validation data. When the model learns nothing from the validation set, the training process is finished, and the number of cascade layers is determined.

Compared to traditional deep learning networks, gcForest has three advantages. First, the gcForest can work well without a GPU, which is necessary for a deep learning network so that forests in the same layer can be trained in parallel. Second, gcForest is a kind of tree model that retains good interpretability and understandability. This is very important for both patients and physicians. Third, deep learning networks need large-scale training data that make fine tuning complicated and time consuming.

With the help of a cascading structure and the characteristics of tree model, the number of layers is automatically determined by gcForest. This mechanism effectively enables low computation complexity and reduces training time. Therefore, the accuracy and efficiency of gcForest are superior in small-scale training data, which makes it suitable for edge computing. In summary, gcForest is a strong framework for AIDM's application layer.

6. Simulations results

We now discuss our building of a test bed to validate AIDM. In our test bed, we use a multigroup analytical framework to optimize the access efficiency of medical sensors and environmental sensors in a real-world scenario in the transmission layer. We then build a prediction model with a real-world health examination dataset to improve the screening rate of diabetes patients and deploy the model in the application layer.

6.1. A real-world hospital ward

We present the simulation results in this section. All the simulation settings are the same as the above RA procedure. We simulate the de-

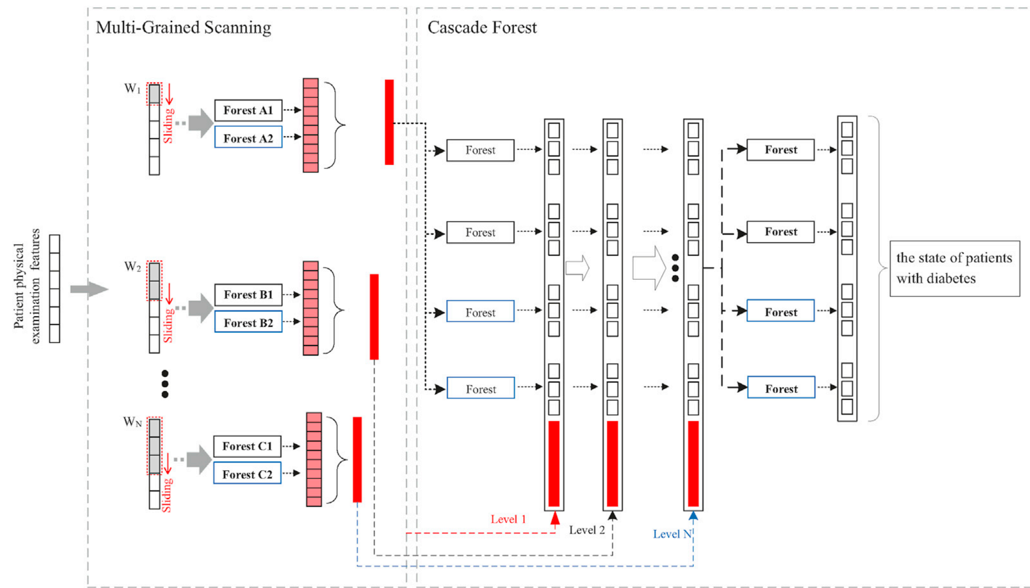


Fig. 4. The application of a deep forest in predicting the status of a patient with diabetes.

mands of a massive access smart hospital ward for diabetes in a Jiangsu province hospital. As Section 4 illustrates, we divide all sensors and smart devices into two groups: the medical group and the environment group. The medical group contains sensors for collecting blood glucose, blood pressure, body temperature, and electrocardiography information. The environmental group contains sensors for measuring ambient temperature, humidity, and noise. We define one time slot as 5 ms. The delay constraint of the medical group is 1s, and the sample interval is 0.5s. If an MTD chooses a unique RAP, the RA procedure succeeds, and the duration of each simulation run is 10^6 time slots. The success probability of access requests (P_{sar}) is calculated as follows:

$$P_{sar} = \frac{the\ number\ of\ successful\ access\ requests}{the\ total\ access\ requests} \tag{12}$$

The throughput is given by

$$throughput = \frac{the\ number\ of\ successful\ access\ requests}{the\ duration\ of\ the\ simulation} \tag{13}$$

We define t_{succ} as the time slot that during which the access request is transmitted successfully and t_{gene} as the time slot that when the request is generated. Hence, the access delay is given by

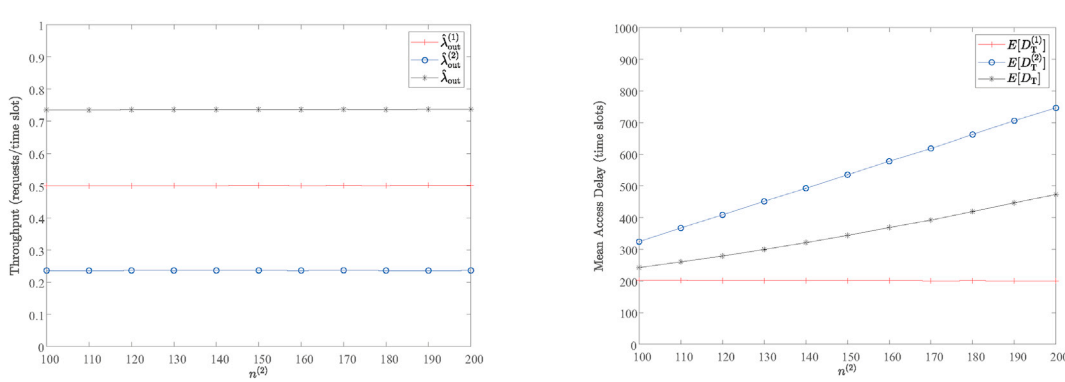
$$access\ delay = t_{succ} - t_{gene} \tag{14}$$

Fig. 5 presents the curves of the throughput and the mean access delay of both the environment group and the medical group versus the number of MTDs in the environment when the backoff parameters of each group are tuned optimally. With the optimal tuning of the backoff parameters, the delay performance of the medical group can always satisfy the delay requirements, and the network throughput can always achieve the optimal value irrespective of the number of MTDs in the environment group. However, with an increase of MTDs in the networks, the delay performance of the environment group increases due to more collisions.

Fig. 6 shows the curves of the throughput and the mean access delay of both the environment group and medical group versus the backoff parameters of the environment group. In this simulation, the ACB factor in each simulation run is fixed. Here we see that with the increment of $q^{(2)}$, there are more MTDs in each time slot, which leads to more severe congestion and the throughput and access delays of the MTDs in the medical group varies. For the environment group, when $q^{(2)} < 0.03$, the throughput increases because more MTDs can perform the RA procedure, but with continuous increases in $q^{(2)}$, more attempts bring more collisions. Thus, the throughput declines, and the access delay increases.

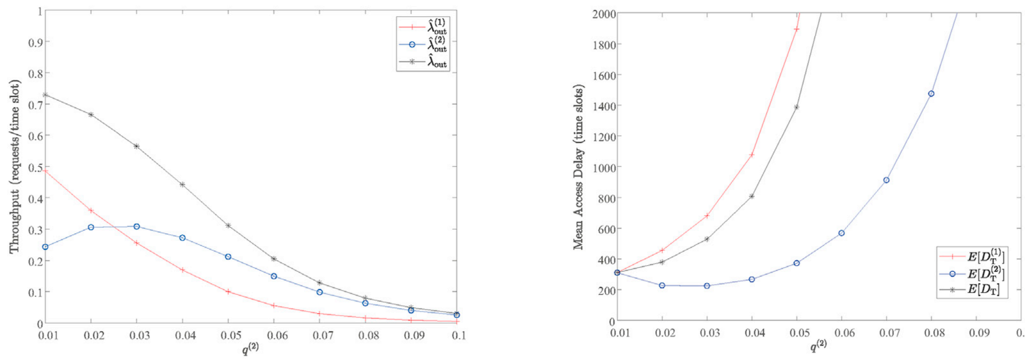
6.2. The health examination dataset

In this subsection, we provide the simulation results to evaluate the



(a) (b)

Fig. 5. Throughput and mean access delay (in time slots) versus the number of MTDs in environment group. $N = 2$. $\lambda^{(1)} = \lambda^{(2)} = 0.01$. $D^{(1)} = 200$. $n^{(1)} = 200$. $W^{(1)} = W^{(2)} = 1$. $q^{(i)} = q^{*,(i)}_{W^{(i)}=1}$.



(a) (b)

Fig. 6. Throughput and mean access delay (in time slots) versus the ACB factor of environment group. $N = 2$. $\lambda^{(1)} = \lambda^{(2)} = 0.01$. $n^{(1)} = 200$. $n^{(2)} = 100$. $W^{(1)} = W^{(2)} = 1$. $q^{(1)} = 0.01$.

Table 1
Accuracy and complexity comparison of SVM, DT, ANN and gcForest.

Model	Precision	F1	Training time(sec)
SVM	0.7781	0.8689	0.139
DT	0.8619	0.9234	0.1530
ANN	0.9263	0.9307	0.25
gcForest	0.9587	0.9688	0.172

performance of gcForest in predicting a patient's diabetes status and choose three other representative algorithms: Decision Tree (DT), Support Vector Machines (SVM), and Artificial Neural Networks (ANN) for comparison. We obtained a health examination dataset from a Jiangsu province hospital, which we then processed as follows. First, we removed irrelevant data such as cardiogram data and follow-up data to reduce the required computing resources. Then, we transformed all categorical features using one-hot encoding. Finally, we standardized the data.

We selected 43 features from a total of 61 by Principal Component Analysis (PCA) and removing multicollinearity. After data processing and feature selection, the dataset contained 1521 people comprising two groups, normal and diabetic.

We built our gcForest in Python and used training time to evaluate the complexity of the model's structure. The ratio of training dataset and testing dataset was 7:3, and we used F1 and precision to measure accuracy by 10-fold cross-validation.

The simulation results are shown in Table 1. The health examination dataset was a high dimensional dataset, and SVM is not appropriate for this type of data. Hence we see that SVM was the worst in all three metrics. Since it uses a cascading model, however, gcForest can indeed tackle the high dimensional work. The performance of gcForest is also better than ANN in both accuracy and complexity and is also more easily explained and understood than ANN, which itself is like DT. Compared to DT, gcForest improve F1 and precision by 0.1 at only a slight increase in training time. We therefore consider gcForest to be the best algorithm for AIDM in 5G. Furthermore, gcForest is suitable for the application layer and use with smart devices.

7. Conclusion

With the help of existing studies and an understanding of diabetes management in 5G and of AI, we proposed an artificial intelligence diabetes management system based on 5G called AIDM. The AIDM contains five layers: the sensing layer, the transmission layer, the storage layer, the computing layer, and the application layer. In the transmission layer, we apply a delay-aware RA optimization to improve the RA procedure of devices. In the application layer, we build a prediction model for screening early diabetes. Our experimental results show that AIDM can balance the requirements of medical sensors and environment sensors and achieve high accuracy with low latency in the prediction of diabetes status. In the future, AIDM can be applied in nursing homes and hospitals to help the elderly and patients of all sorts deal with diabetes. In addition, many research institutions can collect and share research data thought AIDM and diabetics can have more timely access to healthcare services.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by grants from the industry prospecting and common key technology key projects of Jiangsu Province Science and Technology Department (Grant no. BE2020721), the Special guidance funds for service industry of Jiangsu Province Development and Reform Commission (Grant no.(2019)1089), the big data industry development pilot demonstration project of Ministry of Industry and Information Technology of China (Grant no. (2019)243), (2020)84),the Industrial and Information Industry Transformation and Upgrading Guiding Fund of Jiangsu Economy and Information Technology Commission(Grant no. (2018)0419),the Research Project of Jiangsu Province Sciences(Grant no. 2019-2020ZZWKT15),the found of Jiangsu Engineering Research Center of Jiangsu Province Development and Reform Commission (Grant no. (2020)1460),the found of Jiangsu Digital Future Integration Innovation Center (Grant no. (2018)498).

References

[1] L. Zhou, D. Wu, J. Chen, X. Wei, Cross-modal collaborative communications, IEEE Wireless Commun. 27 (2) (2019) 112–117.
[2] FROM HEALTHCARE TO HOMECARE the Critical Role of 5G in Healthcare Transformation, Ericsson, 2017.
[3] M. Chen, Y. Zhang, M. Qiu, N. Guizani, Y. Hao, Spha: smart personal health advisor based on deep analytics, IEEE Commun. Mag. 56 (3) (2018) 164–169.
[4] L. Zhou, D. Wu, X. Wei, J. Chen, Cross-modal stream scheduling for ehealth, IEEE J. Sel. Area. Commun. 39 (2) (2020) 426–437.
[5] R.L. Thomas, S. Halim, S. Gurudas, S. Sivaprasad, D.R. Owens, IDF Diabetes Atlas: a review of studies utilising retinal photography on the global prevalence of diabetes related retinopathy between 2015 and 2018, Diabetes Res. Clin. Pract. 157 (2019) 107840.
[6] S. El-Sappagh, F. Ali, S. El-Masri, K. Kim, A. Ali, K.-S. Kwak, Mobile health technologies for diabetes mellitus: current state and future challenges, IEEE Access 7 (2018) 21917–21947.
[7] M. Rigla, I. Martínez-Sarriegui, G. García-Sáez, B. Pons, M.E. Hernando, Gestational diabetes management using smart mobile telemedicine, J. Diabetes Sci. Technol. 12 (2) (2018) 260–264.
[8] I. Contreras, S. Kiefer, J. Vehi, Adaptive Workflows for Diabetes Management: Self-Management Assistant and Remote Treatment for Diabetes, IOS Press, 2017.

[9] M.M. Alotaibi, R. Istepanian, N. Philip, A Mobile Diabetes Management and Educational System for Type-2 Diabetics in saudi arabia (Saed), Mhealth 2 (2016) 33–39.
[10] M. Chen, J. Yang, J. Zhou, Y. Hao, J. Zhang, C.-H. Youn, 5g-smart diabetes: toward personalized diabetes diagnosis with healthcare big data clouds, IEEE Commun. Mag. 56 (4) (2018) 16–23.
[11] J. Lloret, L. Parra, M. Taha, J. Tomás, An architecture and protocol for smart continuous ehealth monitoring using 5g, Comput. Network. 129 (2017) 340–351.
[12] G. Gardašević, K. Katzis, D. Bajić, L. Berbakov, Emerging wireless sensor networks and internet of things technologies—foundations of smart healthcare, Sensors 20 (13) (2020) 3619.
[13] S. Cao, X. Zhang, R. Xu, Toward secure storage in cloud-based ehealth systems: a blockchain-assisted approach, IEEE Network 34 (2) (2020) 64–70.
[14] M. Chen, X. Wei, J. Chen, L. Wang, L. Zhou, Integration and provision for city public service in smart city cloud union: architecture and analysis, IEEE Wireless Commun. 27 (2) (2020) 148–154.
[15] W.Y.B. Lim, N.C. Luong, D.T. Hoang, Y. Jiao, Y.-C. Liang, Q. Yang, D. Niyato, C. Miao, Federated learning in mobile edge networks: a comprehensive survey, IEEE Commun. Surv. Tutorials 22 (3) (2020) 2031–2063.
[16] S. Duan, V. Shah-Mansouri, Z. Wang, V.W. Wong, D-acb, Adaptive congestion control algorithm for bursty m2m traffic in lte networks, IEEE Trans. Veh. Technol. 65 (12) (2016) 9847–9861.
[17] S.-Y. Lien, K.-C. Chen, Y. Lin, Toward ubiquitous massive accesses in 3gpp machine-to-machine communications, IEEE Commun. Mag. 49 (4) (2011) 66–74.
[18] S.-Y. Lien, K.-C. Chen, Massive access management for qos guarantees in 3gpp machine-to-machine communications, IEEE Commun. Lett. 15 (3) (2011) 311–313.
[19] P. Si, J. Yang, S. Chen, H. Xi, Adaptive massive access management for qos guarantees in m2m communications, IEEE Trans. Veh. Technol. 64 (7) (2014) 3152–3166.
[20] W. Zhan, L. Dai, Massive random access of machine-to-machine communications in lte networks: modeling and throughput optimization, IEEE Trans. Wireless Commun. 17 (4) (2018) 2771–2785.
[21] C. Zhang, X. Sun, J. Zhang, X. Wang, S. Jin, H. Zhu, Throughput optimization with delay guarantee for massive random access of m2m communications in industrial iot, IEEE Internet Things J. 6 (6) (2019) 10077–10092.
[22] Z.-H. Zhou, J. Feng, Deep forest, Natl. Sci. Rev. 6 (1) (2019) 74–86.
[23] Y.-L. Zhang, J. Zhou, W. Zheng, J. Feng, L. Li, Z. Liu, M. Li, Z. Zhang, C. Chen, X. Li, et al., Distributed deep forest and its application to automatic detection of cash-out fraud, ACM Trans. Intell. Syst. Technol. 10 (5) (2019) 1–19.