Journal of Cloud Computing:
Advances, Systems and Applications

RESEARCH Open Access

Private anomaly detection of student health conditions based on wearable sensors in mobile cloud computing

Yu Xie[1], Kuilin Zhang[2], Huaizhen Kou[3] and Mohammad Jafar Mokarram[4*]

Abstract

With the continuous spread of COVID-19 virus, how to guarantee the healthy living of people especially the students who are of relative weak physique is becoming a key research issue of significant values. Specifically, precise recognition of the anomaly in student health conditions is beneficial to the quick discovery of potential patients. However, there are so many students in each school that the education managers cannot know about the health conditions of students in a real-time manner and accurately recognize the possible anomaly among students quickly. Fortunately, the quick development of mobile cloud computing technologies and wearable sensors has provided a promising way to monitor the real-time health conditions of students and find out the anomalies timely. However, two challenges are present in the above anomaly detection issue. First, the health data monitored by massive wearable sensors are often massive and updated frequently, which probably leads to high sensor-cloud transmission cost for anomaly detection. Second, the health data of students are often sensitive enough, which probably impedes the integration of health data in cloud environment even renders the health data-based anomaly detection infeasible. In view of these challenges, we propose a time-efficient and privacy-aware anomaly detection solution for students with wearable sensors in mobile cloud computing environment. At last, we validate the effectiveness and efficiency of our work via a set of simulated experiments.

Keywords: Mobile cloud computing, Healthcare, Privacy, Wearable sensors, Anomaly detection

Introduction

Currently, the wide spread of COVID-19 virus is a huge threat for the healthy living and life of people all over the world. Affected by the pandemic, people are more caring about their personal health conditions than ever before [1–3]. Especially for the students who are often of relatively weaker physique compared with adults, more attentions should be paid to know about the health conditions of students effectively and timely. Generally, through observing the health data of students, we can mine, analyze and discover the possible exceptions or anomalies existing in candidate students. Such anomalies can help education managers or medical departments to quickly filter out the possible patients [4–6].

However, the above anomaly detection process is often non-trivial because there are so many students in each school and the physique of each student is often varied. Fortunately, the quick development of mobile cloud computing technologies and wearable sensors has provided a promising way for school managers to know about the health conditions of students and find out the prospective anomalies timely [7–9]. For example, popular mobile devices (e.g., mobile phones, smart watches, wearable sensors, etc) can monitor the health conditions (e.g., blood pressure, heart rate and so on) of students in a real-time way [10] and transmitted them to a remote cloud platform. Through analyze the mine such

*Correspondence: m.j.mokarram@sutech.ac.ir

[4] Department of Electrical and Electronics Engineering, Shiraz University of Technology, Shiraz, Iran
Full list of author information is available at the end of the article

monitored health data in cloud, we can get to know the real-time health conditions of students and filter out the possible patients from candidate students through anomaly detection technologies [11–13].

However, two challenges are often present in the above anomaly detection process. First, the health data monitored by massive wearable sensors of students are often massive and updated frequently with time elapsing [14–16], which probably brings high sensor-cloud data transmission cost for anomaly detection. Second, the health data of students are often sensitive enough, which probably impedes the integration of health data in cloud especially when there are no enough and attractive incentive mechanisms [17, 18]. In this situation, the data sparsity probably renders the health data-based anomaly detection infeasible. In view of these challenges, we propose a time-efficient and privacy-aware anomaly detection solution for students with wearable sensors in mobile cloud computing environment. Concretely, we adopt an effective privacy-preserving technique to guarantee the sensitive information of people is secure, which can minimize the privacy disclosure concerns of people when a cloud platform integrates the distributed data of people together for uniform data processing and mining. At last, we validate the effectiveness and efficiency of our work via a set of simulated experiments.

In summary, the major novelty or contributions in this work are as follows.

(1) We recognize the significance of anomaly detection in accurate recognition of potential patients, and realize the importance of time-aware health data monitored by wearable sensors in anomaly detection.
(2) We propose an anomaly detection method for student health conditions based on wearable sensors in mobile cloud computing, named Ano-Det. The proposal can achieve efficient and privacy-aware anomaly detection.
(3) Simulated experiments are enacted and deployed to prove the feasibility of the proposal in terms of anomaly detection performances including accuracy, privacy-preservation and computational time in cloud environment. In Ano-Det, we use a hash technique to achieve the two goals of data efficiency and data privacy simultaneously since (a) hash indexes can be created offline whose time complexity is approximately O(1) and (b) hash indexes can secure the sensitive information of students well.

We summarize the rest of paper as follows. Section 2 reviews the state-of-the-art literature in the field. A motivating example is constructed in Section 3 to clarify the research background and challenges focused in this paper. Concrete steps of our proposed Ano-Det method are described in Section 4. Evaluation is made in Section 5. At last, conclusions are drawn in Section 6.

Related work

We investigate the current research literature about anomaly detection in health domain as follows.

Anomaly detection in big data

One of the major characteristics of big data is that the data are coarse [19–22] and therefore, pre-processing treasure before transmitting the data to a cloud platform is necessary to cope with the probably existed anomaly points. As an important part of civil infrastructure, health monitoring system generates a large amount of data; but there are many noises in the data, which is very time-consuming to detect. To tackle the above problem, the authors in [23] propose a data anomaly detection method based on computer vision and deep learning. Overall, the method can be divided into two stages. Firstly, data is converted through visualization, and the construction process of abnormal classification neural network is carried out at the same time. In the second stage, randomly selected training data are input into the neural network for training. After in-depth training, anomalies in a big volume of data can be detected. In [24], anomaly detection in data pre-processing stage before sensor-cloud data transmission is firstly studied, and a new data anomaly detection method based on convolutional neural network is proposed, which mainly imitates human visual and decision-making behavior. Experiments show that the proposed method can detect the abnormal condition of structural health data accurately and efficiently. In [25], an anomaly detection method is proposed combining CNN and GRU. In this method, the stacked convolutional neural network layer is used to capture the input data and extract the features, and then the stacked gated cyclic unit is used to learn the time features. Finally, anomaly detection is performed in the regression layer.

In [26], the authors present a new anomaly detection method which combines the Bayesian dynamic linear model with the switched Kalman filter theory. This method is based on the prior probability of abnormal state and the transition probability between normal state and abnormal state. Importantly, this approach operates under semi-supervised conditions, where normal and abnormal state labels do not require training models. As an unsupervised learning method for anomaly detection, Markov square distance (MSD) has some limitations. Inspired by this observation, work [27] proposes an anomaly detection method for structured health data (SHM) based on adaptive Mahalanobis distance and

KNN rules, i.e., AMSD-kNN. The AMSD-kNN method is mainly used to find the nearest neighbor of training set and test set in two steps, to eliminate the estimation of environment change and local covariance matrix. This method provides an unsupervised learning method for SHM through a new distance measure and kNN rules in cloud. In [28], a kind of hash-based anomaly detection method is introduced to recognize the abnormal points in time-related data stream in Internet of Things. Since hash is very efficient in time cost, the recognition speed of anomaly detection is improved significantly especially in the big data context. Other related literature includes [29] where transfer learning technique is adopted to achieve anomaly detection in big data system and [30] where the anomaly detection effect and performances are validated by mathematical way.

However, the above anomaly detection solutions often fall short in protecting the sensitive information of users, which constrains the wide applications of anomaly detection in various big data applications.

Privacy-aware data utilization

How to secure the sensitive information contained in big data is a key to make full use of the value hidden in big data [31–33]. Hash technique is recruited in [34, 35] to secure the personal information contained in user big data. Concretely, user data are modeled into less-sensitive user indexes and then the user indexes stored in cloud platforms are used in user clustering and missing value prediction. Since user indexes are built offline beforehand, the indexes-based user clustering is very efficient. However, anomaly detection issue is not considered. Blockchain is adopted in [36, 37] to realize the protection of user privacy during the cross-platform data sharing and integration process. This way, the sensitive data stored in different cloud platform can be fully shared for value-added smart applications while guaranteeing not to disclose much private information of users. The advantage of blockchain-based privacy protection solutions is that they are with strong mathematical foundation and confidential degree.

An edge computing-based data sharing and integration method is proposed in [38] to overcome the shortcoming of traditional central big data integration manner according to which user data are transmitted to a remote cloud platform and user information is probably disclosed during data transmission. Differential Privacy technique has been proven an effective privacy protection solution in big data application systems. For example, in [39], the authors use differential privacy to realize the secure sharing of graph data owned by different stakeholders. This way, data owners are willing to publish their respective graph data without the concerns of possible privacy disclosure. Since not all data are useful for creating a big data system, it is not necessary to open all the data owned by users to the public in data sharing. Inspired by this motivation, the authors in [40] put forward a sampling-based data sharing and publishing method. In concrete, only a small portion of user-related data are selected through a sampling process and released to the third party to realize effective data sharing and utilization. This way, most data that are not selected through the sampling process are secured well. In addition, [41] proposed a federated learning method for protecting user sensitive data in the Internet of Things scenario, in which the POI recommendation problem is solved in parallel with the federated learning framework in distributed systems.

However, the above privacy protection solutions do not consider the anomaly points probably existed in big data applications. Therefore, they cannot address the anomaly detection issue in mobile cloud environment well.

Motivation

To better ease the understanding of our motivation, a concrete example is shown in Fig. 1. Here, the health conditions of students are monitored by wearable sensors embedded in various mobile devices (e.g., smart watches, mobile phones, etc.) [42–44]. Thus, we can obtain real-time monitoring data (e.g., electrocardiogram) which can be analyzed and clustered to discover the possible patients from all candidate students. Generally, the monitored health data by wearable sensors need to be transmitted to a remote cloud platform for uniform data processing, during which several challenges are often present. First, the health data monitored by wearable sensors of students are of massive and therefore, the response time of anomaly detection is often long. Second, the students' health conditions are often varied or updated with time, which requires additional time cost to achieve scalable anomaly detection. Third, the monitored health data by wearable sensors need to be first transmitted to the central cloud platform, during which transmission process user privacy is probably disclosed to malicious third parties. Therefore, user privacy is often at risk in centralized anomaly detection process.

Considering the above three challenges, it is necessary to develop a novel anomaly detection approach for people's health conditions based on the monitored health data by wearable sensors in mobile cloud computing environment. Therefore, we propose a new Ano-Det method in the following sections to guarantee efficient, scalable and privacy-preserving anomaly detection in big data context.

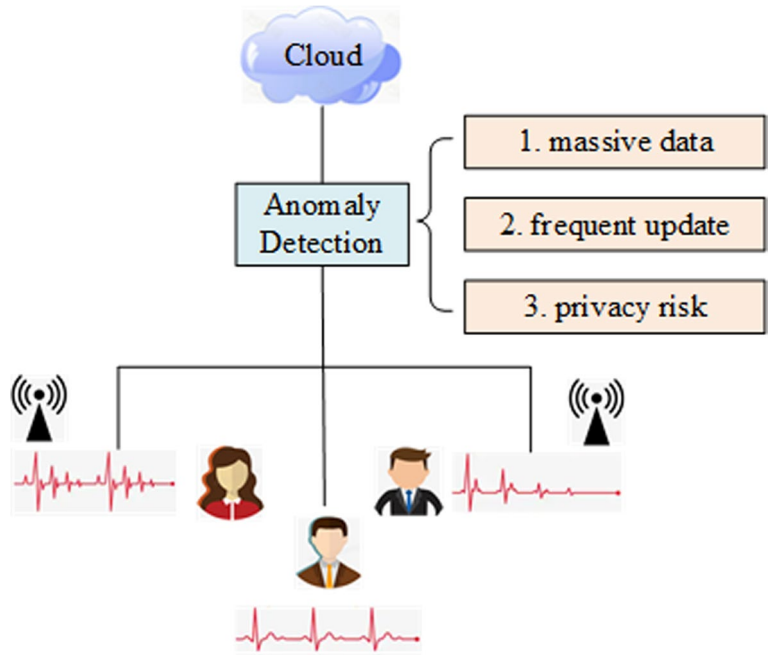

Fig. 1 Anomaly detection of health data with wearable sensors in mobile cloud computing: challenges

Anomaly detection method: Ano-Det

The basic idea of the proposed Ano-Det method is introduced as follows: we firstly convert the health data of students into lightweight health indexes and stored them in the cloud platform; next, we calculate the similarity between each pair of the health conditions of students based on the health indexes; finally, we cluster the students based on their health indexes and discover the possible anomalies based on the clustering results. The concrete details of Ano-Det method is described as follows.

Step 1: Generate each student's health index As indicated in the example in Fig. 1, the students' health data monitored by wearable sensors are often expressed with a curve which fluctuates with time. Therefore, we first model the students' health data with a multi-dimensional matrix κ depicted in Eqs. (1)-(2). Here, we assume that there are N students, i.e., s_1,..., s_N and M health criteria (e.g., heart rate, blood pressure, etc), i.e., c_1,..., c_M. Moreover, each entry in matrix κ, i.e., $A_{i,j}$ ($i = 1, 2,..., N$; $j = 1, 2,..., M$) represents the student s_i's health data over criterion c_j. Furthermore, as described in Fig. 1, each entry $A_{i,j}$ is a time-aware fluctuant curve; therefore, we formulate $A_{i,j}$ with a vector in Eq. (2) where K denotes the number of time points at which wearable sensors monitor and record the health conditions of students. For example, $K = 3$ means that three pieces of health data are monitored by wearable sensors. From certain points of view, parameter K describes the health data monitoring frequency.

$$\kappa = \begin{array}{c} \\ s_1 \\ \vdots \\ s_N \end{array} \begin{array}{c} \begin{array}{ccc} c_1 & \cdots & c_M \end{array} \\ \begin{bmatrix} A_{1,1} & \cdots & A_{1,M} \\ \vdots & \ddots & \vdots \\ A_{N,1} & \cdots & A_{N,M} \end{bmatrix} \end{array} \tag{1}$$

$$A_{i,j} = (a_{i,j,1}, \ldots, a_{i,j,K}) \tag{2}$$

As Eqs. (1)-(2) shows, κ is an $N * M * K$ tensor. To ease the following calculations, we need to convert the $N * M * K$ tensor κ into a multi-dimensional vector. To achieve this goal, we first convert the K-dimensional vector $A_{i,j}$ into a concrete value. Concretely, we first produce a K-dimensional vector B presented in Eq. (3). Here, each entry in vector B is generated by Eq. (4) where function $\Gamma(-1, 1)$ is responsible for producing a random data belonging to [-1, 1]. Thus, with the K-dimensional vector $A_{i,j}$ and the K-dimensional vector B, we compute their inner product according to Eq. (5) and the final result is denoted by $\Omega_{i,j}$.

$$B = (b_1, \ldots, b_K) \tag{3}$$

$$b_k = \Gamma(-1, 1)(k = 1, 2, \ldots, K) \tag{4}$$

$$\Omega_{i,j} = A_{i,j} * B \tag{5}$$

According to Eq. (5), $\Omega_{i,j}$ is a concrete value belonging to $(-inf, +inf)$. Next, to ease the following calculations, we convert the real-value $\Omega_{i,j}$ into a Boolean-value $\Psi_{i,j}$, which is formulated by Eq. (6). In Eq. (6), $\Psi_{i,j}$ value is mapped to be 1 or 0, whose rationale is explained as follows: let us consider a data point D and a hyperplane H; if point D is above the hyperplane H, then the $\Psi_{i,j}$ value corresponding to D is equal to 1; otherwise, if point D is below the hyperplane H, then the $\Psi_{i,j}$ value corresponding to D is equal to 0. This way, we can use such a kind of position relationship between point D and hyperplane H to evaluate whether two points are close or not. This is the theoretical basis behind the hash mapping operation adopted in Eq. (6).

This way, we convert the K-dimensional vector $A_{i,j}$ in Eq. (2) into a Boolean-value $\Psi_{i,j}$. Correspondingly, the $N * M * K$ tensor κ in Eq. (1) can be simplified to be the $N * M$ matrix κ in Eq. (7). Next, we continue to simplify the $N * M$ matrix κ into an N-dimensional vector, which could be finished by the transformation in Eq. (8). Here, π_i is the decimal value corresponding to the Boolean vector $(\Psi_{i,1},..., \Psi_{i,M})$. For example, if $(\Psi_{i,1},..., \Psi_{i,M}) = (1, 1, 1)$, then $\pi_i = 7$. This way, we successfully convert the $N * M$ matrix κ in Eq. (7) into an N-dimensional vector κ in Eq. (8). In other words, each student s_i is corresponding to a concrete decimal value π_i. According to the index theory, decimal value π_i can be considered as the health index of student s_i.

$$\Psi_{i,j} = \begin{cases} 1 & \text{when } \Omega_{i,j} > 0 \\ 0 & \text{when } \Omega_{i,j} < 0 \end{cases} \tag{6}$$

$$\kappa = \begin{array}{c} \\ s_1 \\ \vdots \\ s_N \end{array} \begin{array}{c} \begin{array}{ccc} c_1 & \cdots & c_M \end{array} \\ \begin{bmatrix} \Psi_{1,1} & \cdots & \Psi_{1,M} \\ \vdots & \ddots & \vdots \\ \Psi_{N,1} & \cdots & \Psi_{N,M} \end{bmatrix} \end{array} \tag{7}$$

$$\kappa = \begin{array}{c} \\ s_1 \\ \vdots \\ s_N \end{array} \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_N \end{bmatrix} \tag{8}$$

The advantages of health index here are three-fold: first, health index contains little privacy of students and hence can be transmitted or released to the cloud platform with less privacy risks, which can minimize the privacy disclosure concerns of people when a cloud platform integrates the distributed data of people together for uniform data processing and mining; second, health index-based similar student retrieval is rather quick; third, health index-based similar student retrieval results are rather close to the similar student retrieval results based on original health data that are sensitive to students. Therefore, we use the health indexes of students to take part in the subsequent distance calculation (Step 2) and anomaly detection (Step 3). This way, we can guarantee that the distance calculation and anomaly detection process is time-efficient and privacy-guaranteed.

Step 2: Calculate the similarity between each pair of students based on their health indexes As discussed in Step 1, each student s_i is corresponding to a concrete decimal value π_i. Here, π_i is obtained from the random vector B in Eq. (3) which bring additional uncertainty in creating the accurate health indexes of students. To minimize the uncertainty, q (q is an integer larger than 1) decimal values are necessary to be obtained for each student s_i. In concrete, for each s_i, we repeat the operations in Eqs. (3)-(8) q times to generate $\pi_{i,1},..., \pi_{i,q}$. After that, we get a new matrix κ as specified in Eq. (9). According to Eq. (9), each student s_i is corresponding to a q-dimensional vector $(\pi_{i,1},..., \pi_{i,q})$. Then vector $(\pi_{i,1},..., \pi_{i,q})$ can be regarded as the health index of student s_i.

$$\kappa = \begin{array}{c} \\ s_1 \\ \vdots \\ s_N \end{array} \begin{bmatrix} (\pi_{1,1} \cdots \pi_{1,q}) \\ \vdots \\ (\pi_{N,1} \cdots \pi_{N,q}) \end{bmatrix} \tag{9}$$

With the health indexes of two students s_i and s_j, i.e., $(\pi_{i,1},..., \pi_{i,q})$ and $(\pi_{j,1},..., \pi_{j,q})$, we can compute the similarity between s_i and s_j (denoted by $Sim(s_i, s_j)$) based on the formula in Eqs. (10)-(11). Here, $Sim(s_i, s_j)$ represents the number of dimensions whose values of s_i and s_j are equal. For example, let us consider two students s_1 and s_2 whose health indexes are (1, 2, 3, 4, 5) and (1, 2, 3, 6, 7), respectively. Then their similarity $Sim(s_1, s_2) = 3$ according to Eqs. (10)-(11). Furthermore, to loosen the judgement condition in Eq. (11), we create p (p is an integer larger than 1) hash tables, i.e., we generate κ_1 ,..., κ_p by Eq. (9). Next, we update Eq. (11) to be Eq. (12) where the similarity judgement condition is loosened considerably.

$$Sim(s_i, s_j) = \sum_{z=1}^{q} Sim_{i,j,z} \tag{10}$$

$$Sim_{i,j,z} = 1, \text{iff } \pi_{i,z} = \pi_{j,z} (z = 1, 2, \ldots, q) \tag{11}$$

$$Sim_{i,j,z} = 1, \text{iff } \pi_{i,z} = \pi_{j,z} (z = 1, 2, \ldots, q) \text{ holds in any } \kappa_1, \ldots, \kappa_p \tag{12}$$

Step 3: Student health condition clustering and anomaly detection According to the similarity between different students calculated in Step 2, we can cluster the students into different groups. In general, the students whose

similarity with each other is large belong to an identical group. For example, if two students whose similarity is q, then they would be put into an identical group. Here, for discovering the most similar students, we set a threshold $T(T \leq q)$ for $Sim(s_i, s_j)$. More specifically, only the students s_i and s_j whose $Sim(s_i, s_j)$ is not smaller than T are deemed as similar. Following such a clustering rule, we can divide all the students into different groups. Furthermore, the students who have no similar students could be regarded as anomaly. This way, we can recognize the anomaly students accurately and meanwhile the sensitive information contained in health data transmitted to the cloud platform can be protected very well.

Next, we use the following algorithm to better ease the understanding of our Ano-Det method.

Algorithm 1 Ano-Det

Require: κ: $N * M$-dimensional health matrix of students
Ensure: Ano_{Set}: anomaly student set
for $i = 1$ to M **do**
  for $j = 1$ to N **do**
    $A_{i,j}$=($a_{i,j,1}$, ..., $a_{i,j,k}$)
  end for
end for
for $k = 1$ to K **do**
  $b_k = \Gamma(-1, 1)$
end for
$B = (b_1, ..., b_K)$
for $i = 1$ to M **do**
  for $j = 1$ to N **do** $\Omega_{i,j} = A_{i,j} * B$
    if $\Omega_{i,j} > 0$ **then**
      $\Psi_{i,j} = 1$
    else
      $\Psi_{i,j} = 0$
    end if
  end for
end for
Update κ by (7)
Update κ by (8)
for x=1 to q **do**
  Repeat lines 1∼20
end for
Update κ by (9)
for $j = 1$ to p **do**
  Repeat lines 21∼24
end for
Update $\kappa_1, ..., \kappa_p$
for $i = 1$ to M **do**
  $mark_i$=0
  for $j = 1$ to M **do**
    calculate $Sim(s_i, s_j)$ by (9)-(11)
    if $Sim(s_i, s_j) \geq T$ **then**
      put s_i and s_j into a group
      $mark_i = 1$
    end if
  end for
  if $mark_i$=0 **then**
    put s_i into Ano_Set
  end if
end for

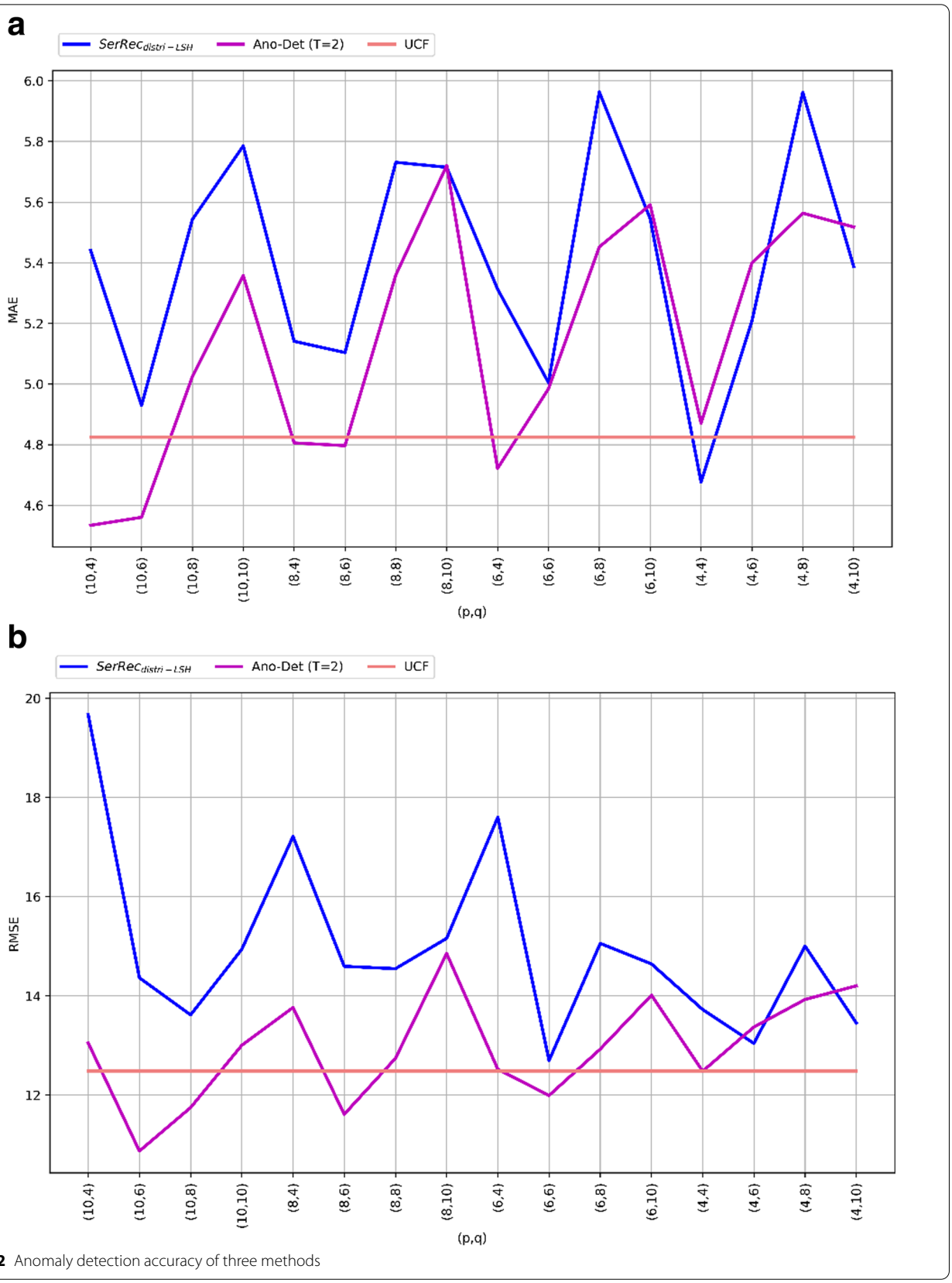

Fig. 2 Anomaly detection accuracy of three methods

Evaluation

We evaluate the feasibility of Ano-Det method via a set of simulate experiments which deployed on WS-DREAM dataset. In concrete, the users and services in the dataset are used to simulate the students and health criteria involved in Ano-Det method. Moreover, only one dimension of response time in the dataset is considered [8]. For comparisons, we also compare Ano-Det method with existing methods $\text{SerRec}_{\text{distri-LSH}}$ [45] and UCF (user-based collaborative filtering). The experiments are run on a computer with 3.20 GHz processor and 8.0 GB memory. The algorithm is developed by Windows 7 and Python 2.7. In concrete, the following three profiles are investigated to prove the algorithm performances.

Profile 1: detection accuracy of three methods Here, the anomaly detection accuracy of Ano-Det method is measured and compared to $\text{SerRec}_{\text{distri-LSH}}$ and UCF. Here, the accuracy is reflected by MAE and RMSE. In the parameter settings, student volume $N = 142$, health criteria volume $M = 4500$, time point volume $K = 64$, threshold $T = 2$, p and q are both varied from 4 to 10. Experimental results are presented in Fig. 2. Concretely, MAE comparison is presented in Fig. 2a where Ano-Det method performs better than SerRecdistri-LSH method (i.e., the MAE of Ano-Det is smaller than $\text{SerRec}_{\text{distri-LSH}}$) because time factor is considered in Ano-Det method and therefore, more accurate anomaly detection results are guaranteed. Although the accuracy of Ano-Det method is worse than UCF method (i.e., the MAE of Ano-Det is larger than UCF), Ano-Det method can secure user privacy well while UCF method cannot. In summary, the comparison results shown in Fig. 2 mean that our Ano-Det method can achieve a good balance between anomaly detection accuracy and privacy-preservation capability. Besides, RMSE comparison is presented in Fig. 2b where similar results are observed as in Fig. 2a. The reason is the same as that analyzed in Fig. 2a and will not be repeated again.

Profile 2: detection efficiency of three methods Here, the anomaly detection efficiency of Ano-Det method is compared to $\text{SerRec}_{distri-LSH}$ and UCF. Here, the efficiency is measured by time cost. In the parameter settings, student volume $N = 142$, health criteria volume $M = 4500$, time point volume $K = 64$, threshold $T = 2$, p

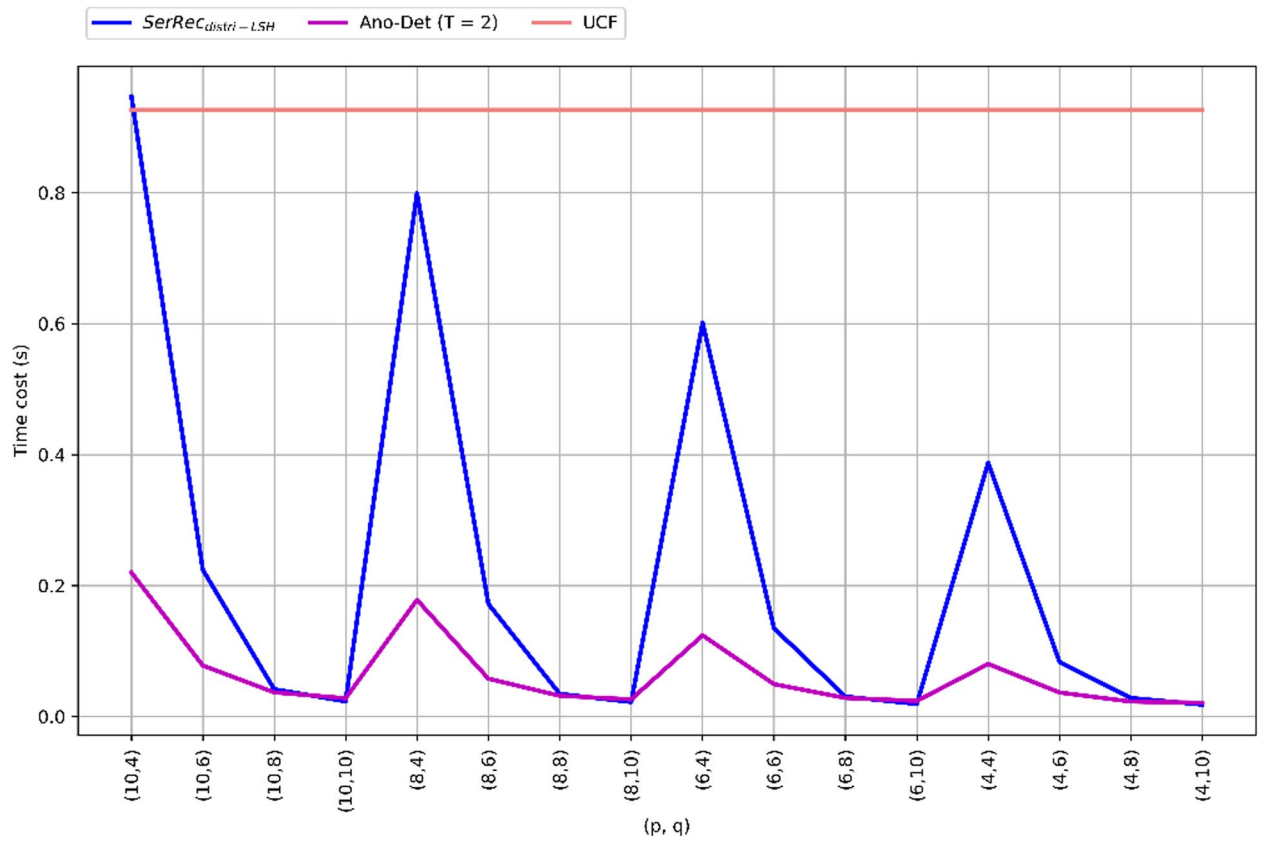

Fig. 3 Anomaly detection efficiency of three methods

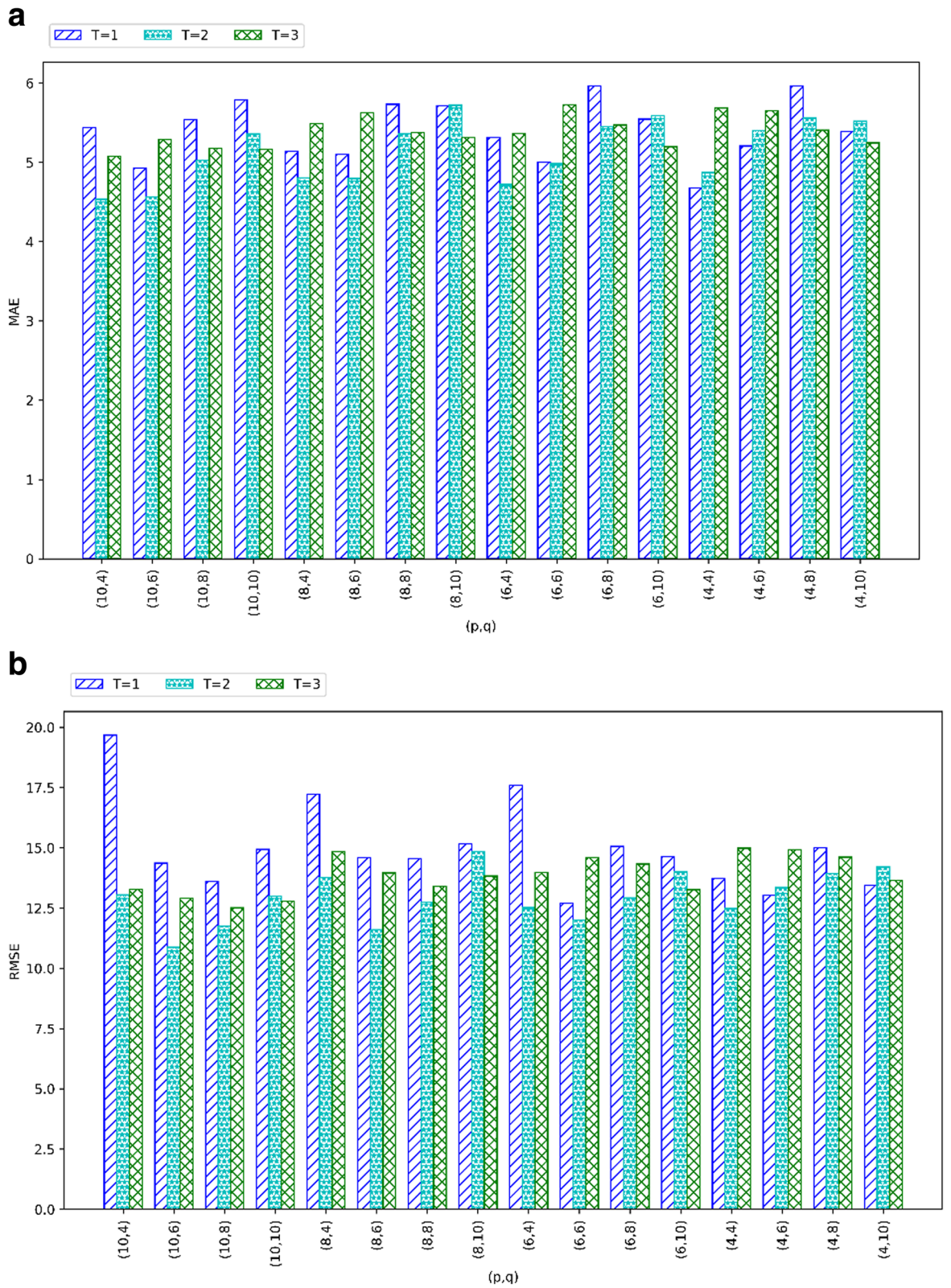

Fig. 4 Anomaly detection accuracy of Ano-Det w.r.t. (p, q, T)

and q are both varied from 4 to 10. Comparison results are demonstrated in Fig. 3. As shown in Fig. 3, Ano-Det and $\text{SerRec}_{\text{distri-LSH}}$ consume less time than the baseline UCF method, because the former two methods both use index technique which is often time efficient in big data context while UCF does not. Furthermore, Ano-Det consumes less time than $\text{SerRec}_{\text{distri-LSH}}$ because time factor is considered in Ano-Det and therefore, less but more similar students could be obtained in Ano-Det. Correspondingly, the time consumed in anomaly detection phase is reduced considerably. Another obvious observation is available from Fig. 3: the time costs of Ano-Det and $\text{SerRec}_{\text{distri-LSH}}$ both decline with the growth of q and the drop of p. The reason can be analyzed as follows: a larger q and a smaller p both mean more rigid similarity judgment conditions according to Eqs. (9)-(12); in this situation, only fewer similar students are returned for clustering and anomaly detection. Therefore, the time cost is decreased accordingly. In summary, the time cost of Ano-Det method is relatively small and hence can be applicable to big data analysis scenarios where a quick response is often necessary.

Profile 3: detection accuracy of Ano-Det w.r.t. (p, q, T) According to the algorithm analysis of Ano-Det method, three important parameters are existent including q in Eq. (9), p in Eq. (12) and threshold T in Step 3. Through analyzing Algorithm 1, we can find that these three factors are all related to the similarity calculation as well as the subsequent student clustering and anomaly detection. Therefore, we design a set of experiments in this profile to observe the relationship of Ano-Det's performances (e.g., MAE and RMSE) with respect to the three parameters. In the parameter settings, student volume $N = 142$, health criteria volume $M = 4500$, time point volume $K = 64$, threshold T is varied from 1 to 3, p and q are both varied from 4 to 10. Comparison results are demonstrated in Fig. 4. As shown in Fig. 4a, the MAE of Ano-Det method is often the largest when $T = 1$. This can be explained as follows: $T = 1$ means that two students are similar as long as their indexes are equal in terms of any of the q dimensions (in Eq. (9)) in any of p hash tables (in Eq. (12)). The above similarity evaluation condition is relatively looser compared to the conditions corresponding to $T = 2$ and $T = 3$. As a consequence, more similar students are returned for subsequent student clustering and anomaly detection even the returned similar students are actually not very similar with each other. Therefore, the anomaly detection accuracy is decreased more or less. Similar results can be obtained from Fig. 4b whose reason is the same as that in Fig. 4a.

Conclusions

It has become a key research issue to guarantee the healthy living of people especially the students who are of relative weak physique. In this situation, precise recognition of the anomaly in student health conditions is beneficial to the quick discovery of potential patients. Fortunately, the quick development of mobile cloud computing technologies [46] and wearable sensors has provided a promising way to monitor the real-time health conditions of students and find out the anomalies timely. However, two challenges are present in the above anomaly detection issue. First, the health data monitored by massive wearable sensors and transmitted to the cloud platform are often massive and updated frequently, which probably leads to low efficiency of anomaly detection. Second, the health data of students are often sensitive enough, which probably impedes the integration of health data in cloud platform. In view of these challenges, a time-efficient and privacy-aware anomaly detection solution for students is proposed with wearable sensors in mobile cloud computing. Finally, we prove the feasibility of our research proposal via a set of simulated experiments.

In this paper, we only discuss the health data with identical formats. However, data format variety is one of the key characteristics of big data applications [47–52]. Therefore, we will further improve our anomaly detection algorithm by accommodating the diversity of health data types or formats in future. In addition, energy saving is an important issue to tackle the challenge raised by big data [36, 53–56]; therefore, we will consider to introduce some effective energy saving techniques into our proposal in future.

Abbreviations

CNN: Convolutional Neural Networks; GRU: Gate Recurrent Unit; MSD: Markov square distance; SHM: Structured Health Data.

Acknowledgements

We would like to thank the providers of WS-DREAM dataset.

Authors' contributions

Yu Xie: idea, writing and experiment design; Kuilin Zhang: motivating example and model; Huaizhen Kou: English writing and experiments; Mohammad Jafar Mokarram: related work and editing. All authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

WS-DREAM: http://wsdream.github.io/.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Chengdu University of Information Technology, Chengdu, China. [2]Shandong Provincial University Laboratory for Protected Horticulture, Weifang University of Science and Technology, Weifang, China. [3]School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. [4]Department of Electrical and Electronics Engineering, Shiraz University of Technology, Shiraz, Iran.

Received: 8 June 2022 Accepted: 26 July 2022
Published online: 05 September 2022

References

1. Ferraz Junior N, Silva AA, Guelfi AE, Kofuji ST (2022) Performance evaluation of publish-subscribe systems in IoT using energy-efficient and context-aware secure messages. J Cloud Comput 11(1):1–17
2. Wang S, Cong Y, Zhu H, Chen X, Qu L, Fan H et al (2020) Multi-scale context-guided deep network for automated lesion segmentation with endoscopy images of gastrointestinal tract. IEEE J Biomed Health Inform 25(2):514–525
3. Uslu BÇ, Okay E, Dursun E (2020) Analysis of factors affecting IoT-based smart hospital design. J Cloud Comput 9(1):1–23
4. Cui WH, Ye J (2019) Logarithmic similarity measure of dynamic neutrosophic cubic sets and its application in medical diagnosis. Comput Ind 111:198–206
5. Fu J, Ye J, Cui W (2018) An evaluation method of risk grades for prostate cancer using similarity measure of cubic hesitant fuzzy sets. J Biomed Inform 87:131–137
6. Ye J, Fu J (2016) Multi-period medical diagnosis method using a single valued neutrosophic similarity measure based on tangent function. Comput Methods Programs Biomed 123:142–149
7. Feng S, Shen S, Huang L, Champion AC, Yu S, Wu C et al (2019) Three-dimensional robot localization using cameras in wireless multimedia sensor networks. J Netw Comput Appl 146:102425
8. Huang J, Lv B, Wu Y et al (2022) Dynamic Admission Control and Resource Allocation for Mobile Edge Computing Enabled Small Cell Network. IEEE Trans Veh Technol 71(2):1964–1973
9. Dai Y, Wu J, Fan Y, Wang J, Niu J, Gu F, et al. (2022) MSEva: A Musculoskeletal Rehabilitation Evaluation System Based on EMG Signals. ACM Trans Sensor Netw (TOSN). https://doi.org/10.1145/3522739
10. Huang J, Tong Z, Feng Z. Geographical POI recommendation for Internet of Things: A federated learning approach using matrix factorization. Int J Commun Syst 1-1. https://doi.org/10.1002/dac.5161
11. Liu Y, Song Z, Xu X, Rafique W, Zhang X, Shen J, et al. (2021) Bidirectional GRU networks-based next POI category prediction for healthcare. Int J Intel Syst. https://doi.org/10.1002/int.22710
12. Liu H, Xu X, Li E, Zhang S, Li X. (2021) Anomaly detection with representative neighbors. IEEE Trans Neural Netw Learn Syst. https://doi.org/10.1109/TNNLS.2021.3109898
13. Kong L, Wang L, Gong W, Yan C, Duan Y, Qi L (2021) LSH-aware multitype health data prediction with privacy preservation in edge environment. World Wide Web 1-16. https://doi.org/10.1007/s11280-021-00941-z
14. Liu J, Shen S, Yue G, Han R, Li H (2015) A stochastic evolutionary coalition game model of secure and dependable virtual service in sensor-cloud. Appl Soft Comput 30:123–135
15. Li Y, Xu H, Cao Q, Li Z, Shen S (2015) Evolutionary game-based trust strategy adjustment among nodes in wireless sensor networks. Int J Distrib Sensor Netw 11(2):818903
16. Zhou X, Liang W, Li W, Yan K, Shimizu S, Kevin I, et al. (2021) Hierarchical Adversarial Attacks Against Graph Neural Network Based IoT Network Intrusion Detection System. IEEE Internet Things J. https://doi.org/10.1109/JIOT.2021.3130434
17. Liu J, Wang X, Shen S, Fang Z, Yu S, Yue G, et al. (2021) Intelligent jamming defense using DNN Stackelberg game in sensor edge cloud. IEEE Internet Things J. https://doi.org/10.1109/JIOT.2021.3103196
18. Zhou H, Shen S, Liu J (2020) Malware propagation model in wireless sensor networks under attack-defense confrontation. Computer Communications 162:51–58
19. Zhou X, Yang X, Ma J, Kevin I, Wang K (2021) Energy efficient smart routing based on link correlation mining for wireless edge computing in IoT. IEEE Internet Things J. https://doi.org/10.1109/JIOT.2021.3077937
20. Zhang H, Shen S, Cao Q, Wu X, Liu S (2020) Modeling and analyzing malware diffusion in wireless sensor networks based on cellular automaton. Int J Distrib Sensor Netw 16(11):1550147720972944
21. Liu J, Wang X, Yue G, Shen S (2018) Data sharing in VANETs based on evolutionary fuzzy game. Futur Gener Comput Syst 81:141–155
22. Chen Y, Gu W, Li K. Dynamic task offloading for Internet of Things in mobile edge computing via deep reinforcement learning. Int J Commun Syst https://doi.org/10.1002/dac.5154
23. Bao Y, Tang Z, Li H, Zhang Y (2019) Computer vision and deep learning-based data anomaly detection method for structural health monitoring. Structural Health Monitoring 18(2):401–421
24. Tang Z, Chen Z, Bao Y, Li H (2019) Convolutional neural network-based data anomaly detection method using multiple information for structural health monitoring. Struct Control Health Monit 26(1):e2296
25. Lee K, Kim JK, Kim J, Hur K, Kim H (2018) CNN and GRU combination scheme for bearing anomaly detection in rotating machinery health monitoring. In: 2018 1st IEEE International conference on knowledge innovation and invention (ICKII). IEEE, pp 102-105
26. Nguyen LH, Goulet JA (2018) Anomaly detection with the switching kalman filter for structural health monitoring. Struct Control Health Monit 25(4):e2136
27. Sarmadi H, Karamodin A (2020) A novel anomaly detection method based on adaptive Mahalanobis-squared distance and one-class kNN rule for structural health monitoring under environmental effects. Mech Syst Signal Process 140:106495
28. Qi L, Yang Y, Zhou X, Rafique W, Ma J (2021) Fast Anomaly Identification Based on Multi-Aspect Data Streams for Intelligent Intrusion Detection Toward Secure Industry 4.0. IEEE Trans Ind Inform https://doi.org/10.1109/TII20213139363
29. Wang W, Wang Z, Zhou Z, Deng H, Zhao W, Wang C et al (2021) Anomaly detection of industrial control systems based on transfer learning. Tsinghua Sci Technol 26(6):821–832
30. Guezzaz A, Asimi Y, Azrour M, Asimi A (2021) Mathematical validation of proposed machine learning classifier for heterogeneous traffic and anomaly detection. Big Data Mining and Analytics 4(1):18–24
31. Ying C, Hua X, Zhuo M, et al. (2022) Cost-Efficient Edge Caching for NOMA-enabled IoT Services. China Commun
32. Zhang K, Tian Z, Cai Z, Seo D (2021) Link-privacy preserving graph embedding data publication with adversarial learning. Tsinghua Sci Technol 27(2):244–256
33. Sandhu AK (2021) Big data with cloud computing: Discussions and challenges. Big Data Mining Analytics 5(1):32–40
34. Qi L, Hu C, Zhang X, Khosravi MR, Sharma S, Pang S et al (2021) Privacy-aware data fusion and prediction with spatial-temporal context for smart city industrial environment. IEEE Trans Ind Inform 17(6):4159–4167
35. Yuan Q, Wang D, Zhao Y, Sang Y, Wang F, Liu Y et al (2021) Privacy-aware examination results ranking for the balance between teachers and mothers. Tsinghua Sci Technol 27(3):581–588
36. Li F, Yu X, Ge R, Wang Y, Cui Y, Zhou H (2021) BCSE: Blockchain-based trusted service evaluation model over big data. Big Data Min Analytics 5(1):1–14
37. Yuan L, He Q, Chen F, Zhang J, Qi L, Xu X et al (2021) CSEdge: Enabling Collaborative Edge Storage for Multi-Access Edge Computing Based on Blockchain. IEEE Trans Parallel Distrib Syst 33(8):1873–1887
38. Xu J, Li D, Gu W et al (2022) UAV-assisted Task Offloading for IoT in Smart Buildings and Environment via Deep Reinforcement Learning. Building and Environment. https://doi.org/10.1016/j.buildenv.2022.109218
39. Zheng X, Zhang L, Li K, Zeng X (2021) Efficient publication of distributed and overlapping graph data under differential privacy. Tsinghua Sci Technol 27(2):235–243
40. Wang H, Cao Z, Zhou Y, Guo ZK, Ren Z (2021) Sampling with prior knowledge for high-dimensional gravitational wave data analysis. Big Data Min Analytics 5(1):53–63
41. Chen Y, Zhao F, Lu Y, Chen X. Dynamic task offloading for mobile edge computing with hybrid energy supply. Tsinghua Science and Technology https://doi.org/10.26599/TST.2021.9010050

42. Zhou X, Hu Y, Wu J, Liang W, Ma J, Jin Q (2022) Distribution Bias Aware Collaborative Generative Adversarial Network for Imbalanced Deep Learning in Industrial IoT. IEEE Trans Ind Inform. https://doi.org/10.1109/TII.2022.3170149
43. Wu Z, Shen S, Zhou H, Li H, Lu C, Zou D (2021) An effective approach for the protection of user commodity viewing privacy in e-commerce website. Knowl-Based Syst 220:106952
44. Liang W, Hu Y, Zhou X, Pan Y, Kevin I, Wang K (2021) Variational few-shot learning for microservice-oriented intrusion detection in distributed industrial IoT. IEEE Trans Ind Inform 18(8):5087–5095
45. Qi L, Zhang X, Dou W, Ni Q (2017) A Distributed Locality-Sensitive Hashing based Approach for Cloud Service Recommendation from Multi-Source Data. IEEE J Sel Areas Commun 35(11):2616–2624
46. Chen Y, Liu Z, Zhang Y, Wu Y, Chen X, Zhao L (2021) Deep reinforcement learning-based dynamic resource management for mobile edge computing in industrial internet of things. IEEE Trans Ind Inform 17(7):4925–4934
47. Su YS, Ruan Y, Sun S, Chang YT (2020) A Pattern Recognition Framework for Detecting Changes in Chinese Internet Management System. J Soc Comput 1(1):28–39
48. Li T, Li C, Luo J, Song L (2020) Wireless recommendations for Internet of vehicles: Recent advances, challenges, and opportunities. Intell Converged Netw 1(1):1–17
49. Evans J (2020) Social computing unhinged. J Soc Comput 1(1):1–13
50. Qi L, Lin W, Zhang X, Dou W, Xu X, Chen J (2022) A Correlation Graph based Approach for Personalized and Compatible Web APIs Recommendation in Mobile APP Development. IEEE Trans Knowl Data Eng. https://doi.org/10.1109/TKDE.2022.3168611.
51. Gao X, Luo JD, Yang K, Fu X, Liu L, Gu W (2020) Predicting Tie Strength of Chinese Guanxi by Using Big Data of Social Networks. J Soc Comput 1(1):40–52
52. Catlett C, Beckman P, Ferrier N, Nusbaum H, Papka ME, Berman MG et al (2020) Measuring cities with software-defined sensors. J Soc Comput 1(1):14–27
53. Rahman GS, Dang T, Ahmed M (2020) Deep reinforcement learning based computation offloading and resource allocation for low-latency fog radio access networks. Intelligent and Converged Networks. 1(3):243–257
54. Meng S, Huang W, Yin X, Khosravi MR, Li Q, Wan S et al (2020) Security-aware dynamic scheduling for real-time optimization in cloud-based industrial applications. IEEE Trans Ind Inform 17(6):4219–4228
55. Nath S, Wu J (2020) Deep reinforcement learning for dynamic computation offloading and resource allocation in cache-assisted mobile edge computing systems. Intell Converged Netw 1(2):181–198
56. Chen Y, Zhao F, Chen X, Wu Y (2022) Efficient Multi-Vehicle Task Offloading for Mobile Edge Computing in 6G Networks. IEEE Trans Veh Technol 71(5):4584–4595

Publisher's Note