

Health Care Virtual Environment System Reinforced by Hadoop and Cloud

Vidyashree S K

Department of CS&E
BGS Institute of technology
Mandya, Karnataka, India
Vidyagowda.sap@gmail.com

Abstract—The advances in information technology have witnessed great progress on healthcare technologies in various domains nowadays. However, these new technologies have also made healthcare data not only much bigger but also much more difficult to handle and process. Moreover, because the data are created from a variety of devices within a short time span, the characteristics of these data are that they are stored in different formats and created quickly, which can, to a large extent, be regarded as a big data problem. To provide a more convenient service and environment of healthcare, this paper proposes a cyber-physical system for patient-centric healthcare applications and services, called Health-CPS, built on cloud and big data analytics technologies. This system consists of a data collection layer with a unified standard, a data management layer for distributed storage and parallel computing, and a data-oriented service layer. The results of this study show that the technologies of cloud and big data can be used to enhance the performance of the healthcare system so that humans can then enjoy various smart healthcare applications and services.

Index Terms—Body area networks (BANs), big data, cloud computing, cyber-physical systems (CPS), healthcare.

I. INTRODUCTION

With the assistance of cloud computing and big data, healthcare data from cyber-physical systems (CPS) [9], [10] (e.g., huge files, complex structures, and different features) can be For the healthcare industry, cloud and big data not only are important techniques but also are gradually becoming the trend in healthcare innovation. Nowadays, medicine is relying much more on specific data collection and analysis, whereas

IN the past two decades, information technology has been widely utilized in medicine [1]. Electronic health records (EHRs), biomedical database, and public health have been enhanced not only on the availability and traceability but also on the liquidity of data [2]. As healthcare-related data are con-

- 1) **Large Scale:** With the improvement of medical informationization, particularly the development of hospital information systems, the volume of medical data has been increasing [3]. In addition, the promotion of wearable health devices accelerates the explosion of healthcare data [4].
- 2) **Rapid Generation:** Most medical equipment, particularly wearable devices, continuously collects data. The rapidly generated data need to be processed promptly for responding to emergency immediately [5].
- 3) **Various Structure:** Clinical examination, treatment, monitoring, and other healthcare devices generate complex and heterogeneous data (e.g., text, image, audio, or video) that are either structured, semistructured, or nonstructured at all [6], [7].
- 4) **Deep Value:** The value hidden in an isolated source is limited. However, through data fusion of EHR and electronic medical records (EMRs), we can maximize the

medical knowledge is explosively growing. Therefore, medical knowledge published and shared via cloud is popular in practice [21], [22]. Patients typically will know more than a doctor. As such, the information and knowledge base can be enriched and shared by the doctors over the cloud [23]. The patients can also actively participate in medical activities assisted by big data. Through smart phones, cloud computing, 3-D printing, gene sequencing, and wireless sensors, the medical right returns to the patients, and the role of a doctor is as a consultant to provide decision support to the patients [24].

The revolutions of cloud and big data may have a strong impact on the healthcare industry, which has even been reformed as a new complex ecosystem. Fig. 1 illustrates the extended healthcare ecosystem, including traditional roles (e.g., patients

efficiently managed [11]. In [12], Lin proposed a NoSQL-based approach for rapid processing, storage, indexing, and

sistently explosive, there are challenges for data management, storage, and processing, as follows.

deep value from healthcare data, such as personalized health guidance and public health warning [8].

rational database. Nie *et al.* [13] also designed a multilingual health search engine to return one multifaceted answer that is well structured and precisely extracted from multiple heterogeneous healthcare sources. In another study [14], Takeuchi and Kodama presented a personal dynamic health system based on cloud computing and big data to store daily healthcare-related information collected by mobile devices. In addition, they also proposed a health data mining algorithm to find out the correlation between a health condition and lifestyle.

Although the innovations are in the healthcare field, there are some issues that need to be solved, particularly the heterogeneous data fusion and the open platform for data access and analysis of healthcare data, to overcome the limitation of a

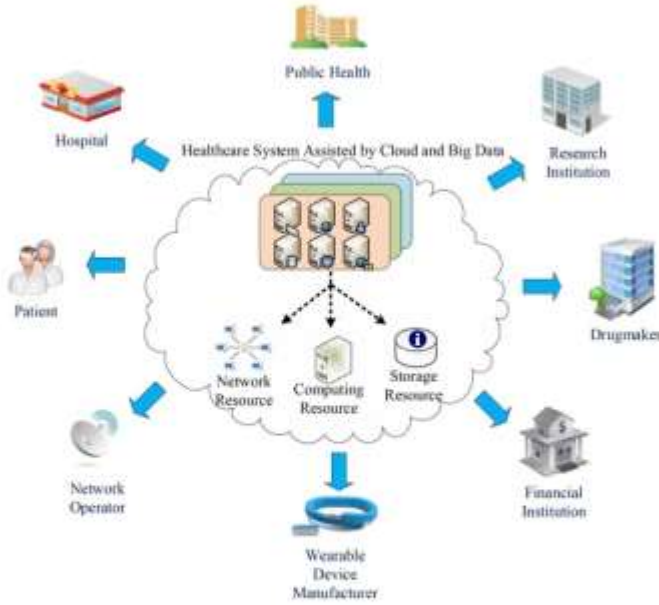


Fig. 1. Illustration for the extended healthcare ecosystem.

building a comprehensive healthcare system is in the handling of the heterogeneous healthcare data captured from multiple sources. That is why a healthcare CPS using technologies of cloud and big data (Health-CPS) is presented in this paper, the contributions of which can be summarized as follows: 1) A unified data collection layer for the integration of public medical resources and personal health devices is presented; 2) a cloud-enabled and data-driven platform for the storage and analysis of multisource heterogeneous healthcare data is established; and 3) a healthcare application service cloud is designed, which provides a unified application programming interface (API) for the developers and a unified interface for the users.

II. HEALTH-CPS ARCHITECTURE

A. Design Issues

For the healthcare industry, cloud and big data not only are important techniques but also are gradually becoming the trend in healthcare innovation. Nowadays, medicine is relying much more on specific data collection and analysis, whereas medical knowledge is explosively growing. Therefore, medical knowledge published and shared via cloud is popular in practice [21], [22]. Patients typically will know more than a doctor. As such, the information and knowledge base can be enriched and shared by the doctors over the cloud [23]. The patients can also actively participate in medical activities assisted by big data. Through smart phones, cloud computing, 3-D printing, gene sequencing, and wireless sensors, the medical right returns to the patients, and the role of a doctor is as a consultant to provide decision support to the patients [24].

The revolutions of cloud and big data may have a strong impact on the healthcare industry, which has even been reformed as a new complex ecosystem. Fig. 1 illustrates the extended healthcare ecosystem, including traditional roles (e.g., patients

and medical institutions) and other newcomers. That is why we need to design a more suitable healthcare system to cope with the following challenges in this novel revolution.

- 1) **Multisource Heterogeneous Data Management With Unified Standards.** A variety of data types and heterogeneity of homogeneous data make healthcare data hard to harness. On one hand, the system must support various healthcare equipment devices to ensure scalability. On the other hand, the data formats should be converted with the unified standards, to improve the efficiency of data storage, query, retrieval, processing, and analysis.
- 2) **Diversified Data Analysis Modules With Unified Programming Interface.** Diversified healthcare data include structured, semistructured, and other unstructured data. According to these different data structures, it is necessary to deploy suitable methods for efficient online or offline analysis, such as stream processing, batch processing, iterative processing, and interactive query. To reduce system complexity and improve development and access efficiency, a unified programming interface will be a fundamental component.
- 3) **Application Service Platform With Unified North-bound Interface.** As shown in Fig. 1, the system is expected to provide various applications and services for different roles. To provide available and reliable healthcare services, the application service platform of this system is essential for resource optimization, technical support, and data sharing.

B. Cloud- and Big-Data-Assisted Architecture

Fig. 2 illustrates the architecture of Health-CPS, which consists of three layers, namely, data collection layer, data management layer, and application service layer.

- 1) **Data collection layer:** This layer consists of data nodes and adapters, provides a unified system access interface for multisource heterogeneous data from hospitals, Internet, or user-generated content. Through an adapter, raw data in a variety of structures and formats can be preprocessed to ensure the availability and security of the data transmission to the data management layer.
- 2) **Data management layer:** This layer consists of a distributed file storage (DFS) module and a distributed parallel computing (DPC) module. Assisted by big-data-related technologies, DFS will enhance the performance of the healthcare system by providing efficient data storage and I/O for heterogeneous healthcare data. According to the timeliness of data and priority of analysis task, DPC provides the corresponding processing and analysis methods.
- 3) **Application service layer:** This layer provides users the basic visual data analysis results. It also provides an open unified API for the developers aiming at user-centric application to provide rich, professional, and personalized healthcare services.

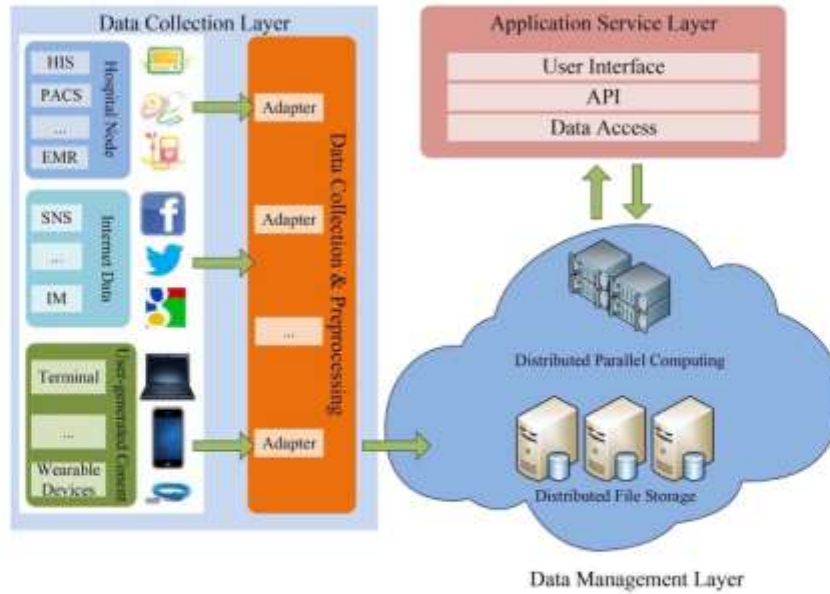


Fig. 2. Health-CPS architecture.



Fig. 3. Data collection layer.

III. DATA COLLECTION LAYER

As shown in Fig. 3, in the data collection layer, various healthcare data are collected by the data nodes and are transmitted to the cloud through the configurable adapters that provide the functionality to preprocess and encrypt the data.

A. Data Node

The user-oriented Health-CPS collects not only the traditional medical data but also the healthcare-related daily

behavior data. According to the data sources, data nodes can be divided into the following four groups.

- 1) **Research data.** Drug research and development institutions and other scientific research institutions have accumulated a large amount of research data, such as clinical trial data and high-throughput screening data. These digital data, including individual or clinical gene or protein data, can help identify the drug side effect and the new effect.
- 2) **Medical expense data.** Medical behaviors generate massive expense data, such as medical bill and medical insurance reimbursement, which are not the traditional healthcare data, but it can be used to analyze and estimate the medical cost. These data are generally stored in different databases of medical institutions, which are geographically dispersed and adopt unified data formats.
- 3) **Clinical data.** This is the typical medical data, usually collected by medical service providers for clinical diagnosis such as EMR, and medical image. These data can be unified, managed, and opened to researchers with a necessary precondition for ensuring the privacy of the patient, to maximize the value of clinical medical data mining.
- 4) **Individual activity and emotion data.** This kind of data is generated apart from the healthcare sector, but it is also relevant to personal health. For example, individual retail consumption records can reflect living habits, which can be used to evaluate individual health risk and make a personalized health plan. Furthermore, based on the physiological data collected by wearable devices, the health status of a user can be easily monitored and traced. The individual emotion data are available to be collected through the information published on the social networks, which can be used in the mental health measuring and affective computing. Particularly for the recovering

patients, a doctor may be able to adjust the treatment plan according to a patient's emotion. An emotion-aware healthcare service promotes the innovation of modern medical with humanistic treatment.

B. Adapter

Adapter is a middleware to provide a data node with access to the system, which is not only a physical data link but also a raw data preprocessor and encrypter. Apart from cleaning data, removing redundancy, and doing compression, the preprocessing module supports data format transformation. According to the type of collected data, the adapter adopts a system-defined data standard for the format conversion. The encryption module encrypts the preprocessed data to ensure security through a hierarchical privacy preservation mechanism. Any unauthorized devices cannot decrypt the data package even if they have access to the system. To improve the scalability of such a system, the functional unit of an adapter is configurable. When the following conditions are met, the corresponding modules of the adapter can be updated online.

- 1) **Data node variation:** When the data node is replaced or upgraded, if the data format of the updated device is not consistent with the former one, the functional units will not work properly. Then, the adapter must send a request to the server for reconfiguring the preprocessing module to make it compatible with the updated one, whereas the server records the kind of updated data node and reauthorize the encryption module online.
- 2) **Data standards update:** When a new type of device without a system-defined data standard has access to the system, the data standard library should be extended, which is expected to be pushed to the corresponding module for update.

IV. DATA MANAGEMENT LAYER

As shown in Fig. 4, the data management layer consists of a DFS module and a DPC module to support efficient management and analysis of heterogeneous data.

A. DFS Module

The primary challenges of big healthcare data are how to establish an efficient distributed storage mechanism for massive data and how to support efficient data processing and analysis. HealthCPS deploys DFS techniques to provide big healthcare data with efficient storage, high-throughput data upload and download, high data fault tolerance, multitenant user management, access control list, and rapid data retrieval and exchange.

To uniformly manage multisource heterogeneous data, the following three components are involved to establish the data standards.

- 1) **Data Description:** Considering data availability, reliability, and traceability, the stored data are expected to provide a description including data source, data size, metadata number, data age, and data administrator.

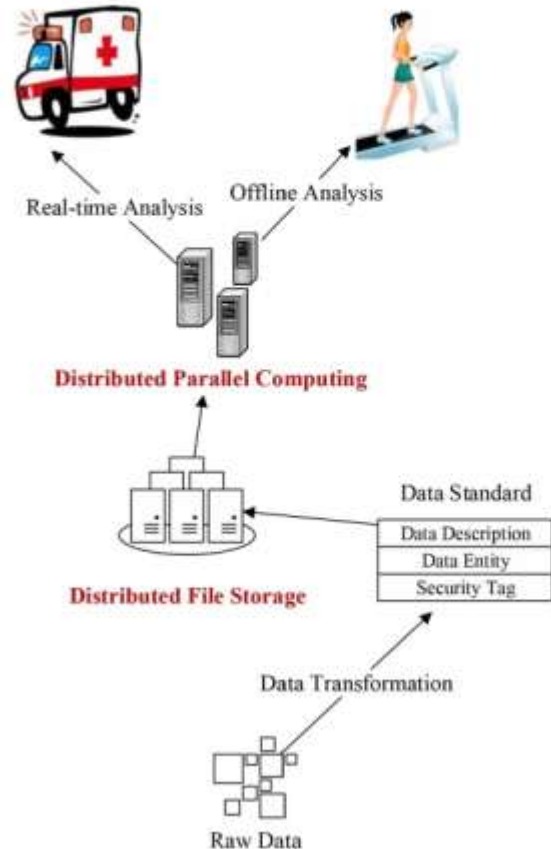


Fig. 4. Data management layer.

- 2) **Data Entity:** As mentioned above, to support the integration of data in the same category, unified standards are essential for data transformation. Data entity is a standardized definition of data objects, which includes attributes, type, value range, and the relationships with other objects.
- 3) **Security Tag:** It defines data security and authorization levels.

B. DPC Module

The DPC module analyzes and processes data from DFS and ultimately discovers knowledge. DPC provides offline computation for massive unstructured data, supports real-time data analysis and query, and integrates various data mining and machine learning algorithms. In different scenarios, DPC supports real-time analysis as well as offline analysis.

- 1) **Real-time analysis.** In the scenarios of intensive care, sudden disease detection, or vital signs monitoring, the changing data reflects individual health status in real time. Therefore, these data need to be processed rapidly, and the analysis results are expected to return quickly for responding to emergency situations. With the memory analysis framework, hotspot data, such as heart rate, blood pressure, and other relevant data, will be kept in memory for improving the analysis efficiency.

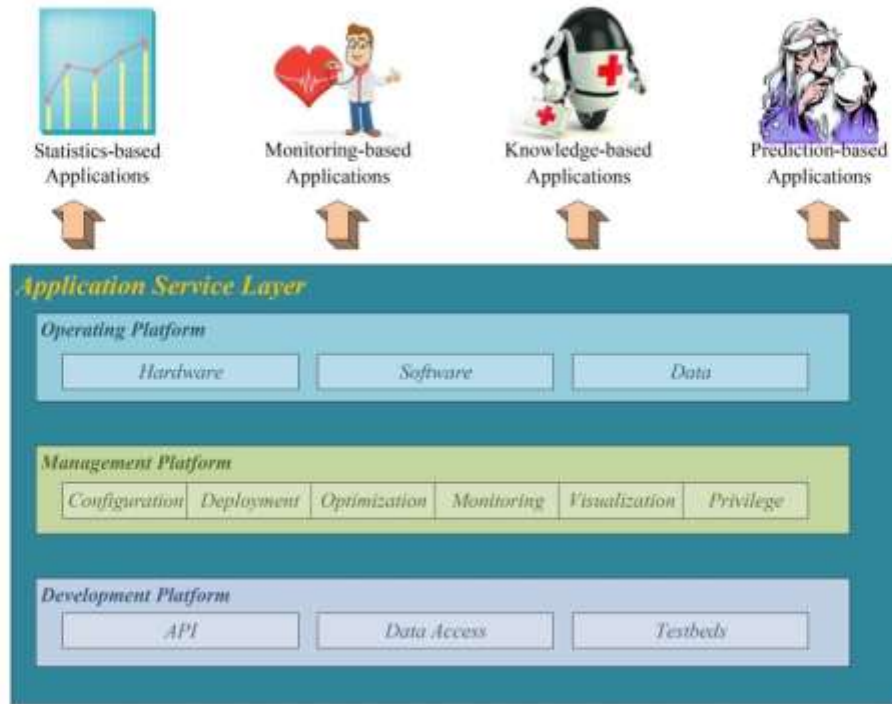


Fig. 5. Framework of the application service layer.

- 2) **Offline analysis.** In the scenarios without high expectations of response time (e.g., health status evaluation, medical recommendation, or health planning), common offline analysis methods are available in DPC, including machine learning, statistical analysis, and recommendation algorithms.

V. APPLICATION SERVICE LAYER

A. Framework of Application Service Layer

As shown Fig. 5, the application service layer provides an operating platform, a management platform, and a development platform for both users and developers.

- 1) **Operating platform** provides resources for running healthcare applications, i.e., hardware, software, and data.
- 2) **Management platform** manages various applications in Health-CPS, including configuration management, deployment management, optimization management, monitoring management, visualization management, and privilege management.
- 3) **Development platform** provides a uniform API, data access, and testbeds.

B. Data-Oriented Healthcare Applications and Services

According to the technical complexity and commercial value, the applications can be divided into the following four groups.

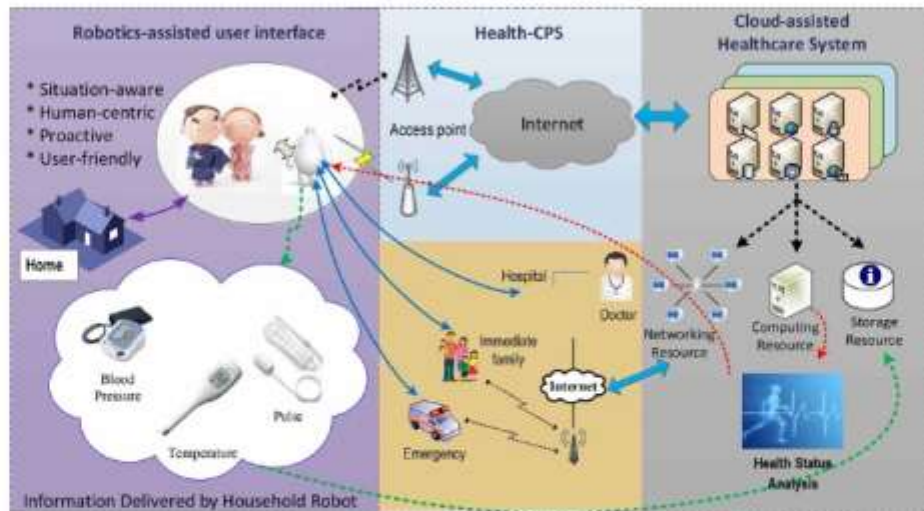
- 1) **Statistics-based** applications only provide basic statistics and report services. For example, an individual health

status report is the representative application. In addition, drug misuse and outdated reports are available through the statistics of clinical trial data.

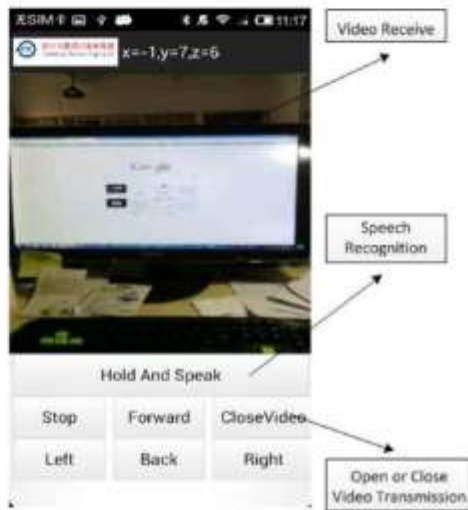
- 2) **Monitoring-based** applications are typically utilized to monitor individual vital signs. Through real-time analysis, a user's physiological changes can be immediately detected to avoid sudden diseases. Through offline analysis of historical data, the recovery procedure can be traced, which supports treatment optimization.
- 3) **Knowledge-based** applications are the most representative big data application. Supported by data mining and machine learning techniques, it is available to discover data correlation and dependence. Typical applications include chronic disease diagnosis, genetic disease analysis, treatment evaluation, side effect identification, and public health warning.
- 4) **Prediction-based** applications have the highest technical complexity and greatest commercial value. For example, individual eating habits can be deduced through retail records, and some potential health risks can be predicted, particularly diet-related diseases, such as obesity and high blood pressure. In addition, considering individual physiological features, individual treatment simulation is available to assess risk and make the optimal medical plan.

VI. ROCHAS: A TESTBED FOR HEALTH-CPS BASED ON ROBOT TECHNOLOGY

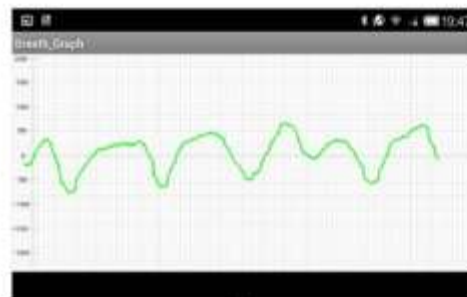
Here, as an example, a brief introduction to a robotic testbed developed by the Embedded and Pervasive Computing Laboratory at Huazhong University of Technology and



(a)



(b)



(c)

Fig. 6. ROCHAS testbed architecture and software interfaces. (a) ROCHAS testbed architecture. (b) Interface to control robot. (c) Sensory data representation.

Science is given to provide home users with situation-aware, human-centric, proactive, and user-friendly healthcare services via CPS.

A. Testbed Architecture

As shown in Fig. 6(a), the testbed consists of the following components.

- 1) **Robotics-assisted user interface** includes: 1) robot hardware, e.g., mechanical devices, environment sensors, biosensors, and camera, and 2) embedded software, e.g., motion control for robot, speech recognition, and sensory data preprocessing.
- 2) **Health-CPS** provides a robot with healthcare services supported by a cloud-assisted system, and communication to the doctor, family, and emergency.
- 3) **Cloud-assisted healthcare system** provides rich networking, computing, and storage resource for sensory data analysis and health modeling.

B. Technical Details

The controller of the robot is an ARM board integrated with Cortex-A8, and the operation system is Android 4.0. The embedded and driver software are written in C, whereas the interface is written in Java. The robot is equipped with various sensors to collect temperature, humidity, heart rate, and electrocardiogram.

As shown in Fig. 6(b), both home and remote users can control the robot via a smart phone, such as mechanical movement, speech recognition, and video transmission. As shown in Fig. 6(c), after processing of sensory data, various physical conditions are stored in the cloud and represented in the user interface, which can be used for the assisted decision of treatment.

VII. CONCLUSION

Nowadays, space (from hospital to home and carry) and time (from discrete sampling to continuous tracking and monitoring) are no longer a stumbling stone for modern healthcare by

using more powerful analysis technologies. Medical diagnosis is evolving to patient-centric prevention, prediction, and treatment. The big data technologies have been developed gradually and will be used everywhere. Consequently, healthcare will also enter the big data era. More precisely, the big data analysis technologies can be used as guide in lifestyle, as a tool to support in the decision-making, and as a source of innovation in the evolving healthcare ecosystem. This paper has presented a smart health system assisted by cloud and big data, which includes 1) a unified data collection layer for the integration of public medical resources and personal health devices, 2) a cloud-enabled and data-driven platform for multisource heterogeneous healthcare data storage and analysis, and 3) a unified API for developers and a unified interface for users. Supported by Health-CPS, various personalized applications and services are developed to address the challenges in the traditional healthcare, including centralized resources, information island, and patient passive participation. In the future, we will focus on developing various applications based on the Health-CPS to provide a better environment to humans.

REFERENCES

- [1] J. Wan *et al.*, "Cloud-enabled wireless body area networks for pervasive healthcare," *IEEE Netw.*, vol. 27, no. 5, pp. 56-61, Sep./Oct. 2013.
- [2] G. Kim, S. Trimi, and J. Chung, "Big-data applications in the government sector," *ACM Commun.*, vol. 57, no. 3, pp. 78-85, 2014.
- [3] K. Lee, T. T. Wan, and H. Kwon, "The relationship between healthcare information system and cost in hospital," *Pers. Ubiquitous Comput.*, vol. 17, no. 7, pp. 1395-1400, Oct. 2013.
- [4] M. Chen, Y. Ma, S. Ullah, W. Cai, and E. Song, "ROCHAS: Robotics and cloud-assisted healthcare system for empty nester," in *Proc. BodyNets*, Boston, MA, USA, Sep. 30-Oct. 2 2013, pp. 1-4.
- [5] M. Chen, "NDNC-BAN: Supporting rich media healthcare services via named data networking in cloud-assisted wireless body area networks," *Inf. Sci.*, vol. 284, pp. 142-156, Nov. 2014.
- [6] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile Netw. Appl.*, vol. 19, no. 2, pp. 171-209, Apr. 2014.
- [7] M. Chen, Y. Hao, Y. Li, D. Wu, and D. Huang, "Demo: LIVES: Learning through interactive video and emotion-aware system," in *Proc. ACM Mobihoc*, Hangzhou, China, Jun. 22-25, 2015.
- [8] H. Zhang, S. Mehrotra, D. Liebovitz, C. Gunter, and B. Malin, "Mining deviations from patient care pathways via electronic medical record system audits," *ACM TMIS*, vol. 4, no. 4, p. 17, 2013.
- [9] J. Wan, H. Yan, H. Suo, and F. Li, "Advances in cyber-physical systems research," *KSII TIS*, vol. 5, no. 11, pp. 1891-1908, 2011.
- [10] J. Wan, D. Zhang, S. Zhao, L. T. Yang, and J. Lloret, "Context-aware vehicular cyber-physical systems with cloud support: Architecture, challenges, and solutions," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 106-113, Aug. 2014.
- [11] M. Chen, Y. Hao, Y. Li, C. Lai, and D. Wu, "On the computation offloading at ad hoc cloudlet: Architecture and service models," *IEEE Commun. Mag.*, vol. 53, no. 3, pp. 1-7, Jun. 2015.
- [12] C. Lin *et al.*, "Temporal event tracing on big healthcare data analytics," in *Proc. IEEE Int. BigData Congress*, 2014, pp. 281-287.
- [13] L. Nie, T. Li, M. Akbari, J. Shen, and T. Chua, "Wenzher: Comprehensive vertical search for healthcare domain," in *Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2014, pp. 1245-1246.
- [14] H. Takeuchi and N. Kodama, "Validity of association rules extracted by healthcare-data-mining," in *Proc. IEEE 36th Annu. Int. Conf. EMBC*, 2014, pp. 4960-4963.
- [15] V. Chandola, S. Sukumar, and J. Schryver, "Knowledge discovery from massive healthcare claims data," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2013, pp. 1312-1320.
- [16] M. Chen, S. Gonzalez, A. Vasilakos, H. Cao, and V. Leung, "Body area networks: A survey," *ACM/Springer Mobile Netw. Appl.*, vol. 16, no. 2, pp. 171-193, Apr. 2011.
- [17] M. Chen, S. Gonzalez, Q. Zhang, M. Li, and V. Leung, "A 2G-RFID based E-healthcare system," *IEEE Wireless Commun. Mag.*, vol. 17, no. 1, pp. 37-43, Feb. 2010.
- [18] M. Chen, Y. Zhang, Y. Li, S. Mao, and V. Leung, "EMC: Emotion-aware mobile cloud computing in 5G," *IEEE Netw.*, vol. 29, no. 2, pp. 32-38, Mar. 2015.
- [19] M. G. R. Alam, E. J. Cho, E. Huh, and C. S. Hong, "Cloud based mental state monitoring system for suicide risk reconnaissance using wearable biosensors," in *Proc. 8th Int. Conf. Ubiquitous Inf. Manage. Commun.*, 2014, p. 56.
- [20] C. He, X. Fan, and Y. Li, "Toward ubiquitous healthcare services with a novel efficient cloud platform," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 1, pp. 30-234, Jan. 2013.
- [21] O. Samuel, M. Omisore, B. Ojokoh, and E. Atajeromavwo, "Enhanced cloud based model for healthcare delivery organizations in developing countries," *Int. J. Comput. Appl.*, vol. 74, pp. 0975-8887, 2013.
- [22] C. Yang *et al.*, "Accessing medical image file with co-allocation HDFS in cloud," *Future Gener. Comput. Syst.*, vol. 43, pp. 61-73, Feb. 2015.
- [23] M. Chen, Y. Zhang, Y. Li, M. Hassan, and A. Alamri, "AIWAC: Affective interaction through wearable computing and cloud technology," *IEEE Wireless Commun. Mag.*, vol. 22, no. 1, pp. 20-27, Feb. 2015.
- [24] M. Chen, Y. Zhang, L. Hu, T. Taleb, and Z. Sheng, "Cloud-based wireless network: Virtualized, reconfigurable, smart wireless network to enable 5G technologies," *ACM/Springer Mobile Netw. Appl.*, DOI: 10.1007/s11036-015-0590-7, 2015, to be published.