Urban Healthcare Big Data System Based on Crowdsourced and Cloud-Based Air Quality Indicators

Min Chen, Jun Yang, Long Hu, M. Shamim Hossain, and Ghulam Muhammad

The authors introduce a method of integrating multi-source air quality data for the data preparation of artificial-intelligence-based smart urban services. Then a testbed of UH-BigDataSys is set up with the deployment of air-quality-aware healthcare applications. Finally, they provide health guidance for urban residents in aspects of respiratory diseases, outdoor travel, sleep quality, and so on.

Abstract

The ever-accelerating process of globalization enables more than half the population to live in cities. Thus, the air quality in cities exerts critical influence on the health status of more and more urban residents. In this article, based on urban air quality data collected through meteorological sites, mobile crowdsourcing, and IoT sensing, along with users' body signals, we propose an urban healthcare big data system named UH-BigDataSys. In this article, we first introduce a method of integrating multi-source air quality data for the data preparation of artificial-intelligence-based smart urban services. Then a testbed of UH-BigDataSys is set up with the deployment of air-quality-aware healthcare applications. Finally, we provide health guidance for urban residents in aspects of respiratory diseases, outdoor travel, sleep quality, and so on. The ultimate goal of UH-BigDataSys is for urban residents to lead healthier lives.

Introduction

In the 2016 edition of World Health Statistics, the World Health Organization (WHO) reported that people dying of diseases related to air pollution reached 6,500,000 in 2012. This number exceeds 11 percent of deaths in 2012. Meanwhile, it is explicitly pointed out by WHO in the report "Ambient (Outdoor) Air Quality and Health" that air pollution is a major environmental risk that exerts influence on health. Diseases such as stroke, heart disease, lung cancer, and chronic and acute respiratory illnesses [1] may be prevented by decreasing air pollution levels. The lower the air pollution level is, the healthier the cardiovascular and respiratory systems of people are, no matter whether in the long term or the short term. The impact of air pollution on the health of humans in urban environments is presented in [2, 3]. Thus, air quality evaluation can have a significant impact on health status of the enormous numbers of urban residents.

Air quality indicators (AQIs) include an internationally used parameter set to evaluate air quality, and the statistics reflects five pollution standards, including ground-level ozone, particulate matter, carbonic oxide, sulfur dioxide, and nitrogen dioxide. At present, AQIs are usually measured by weather monitoring sites. Due to their high construction cost, limited coverage, and insufficient quantity, the air quality data collected by traditional meteorological sites are not enough to portray the real situations. Fortunately, with the ever growing number of smart devices and mobile terminals, urban residents with portable mobile devices can contribute to sense ambient air quality in real time [4, 5]. Then the data are stored and shared in clouds. Thus, the public can participate in the collection of urban air quality data through mobile devices, and the perception of urban air quality may be established over large-scale networks [6]. Crowdsourcing methods for urban environment sensing are discussed in [7, 8]. Furthermore, with the technology advances regarding the Internet of Things (IoT) and vehicular networking [9], smart buildings and vehicles are also equipped with monitoring facilities. The related designs of hardware and software as well as architecture for urban environment sensing are presented [10, 11]. Zheng *et al.* [12] propose to predict air quality by analyzing the correlation among air quality data collected by weather monitoring and vehicles. Through the above methods, more comprehensive analytics is applied to more abundant urban air quality data, which are acquired at lower cost. The seamless integration between public daily life and the perception of air quality can be realized based on sensing assisted by mobile users. The purpose is not only to improve the coverage area and the efficiency of air quality data collection, but also to enable the efficient processing and sharing of real-time data flows in combination with historical information.

However, to the best of our knowledge, there is no work considering the integration of meteorological site data, mobile device crowdsourced data, and IoT sensing data to improve the accuracy of AQI analysis. Furthermore, users' physiological indicators are not considered in the solution. In the face of abominable urban air quality conditions in many developing countries, how to conduct joint urban air quality sensing and health status analytics is critical to provide personalized health monitoring solutions. Although challenging, it brings extensive social value to diagnose individual physiological state based on urban air quali-

Digital Object Identifier:
10.1109/MCOM.2018.1700571

Min Chen, Jun Yang, and Long Hu are with Huazhong University of Science and Technology; M. Shamim Hossain and Ghulam Muhammad are with King Saud University.

ty data for improving the quality of people's life. Thus, an innovative healthcare solution should be discussed based on user-oriented physiological data and urban AQI.

In this article, we propose an urban healthcare big data system, UH-BigDataSys, where data integration and physiological indicator are considered. Meteorological sites, mobile crowdsourcing, and IoT sensing are adopted for air quality data collection in order to provide urban residents with more comprehensive and more accurate air quality monitoring services. Besides air quality monitoring, real-time physiological index monitoring for users is also realized by the use of wearable devices. Corresponding guidance for health and daily activities is provided to urban residents dynamically as air quality changes. As for a user who participates in outdoor activities, a personalized and healthy activity plan is recommended by UH-BigDataSys, with the analytics of his or her current physiological status and air quality in his or her surrounding environment. The introduction of an air quality index provides health analysis for the user with more abundant dimensionalities of information and improves efficiency in health guidance. The main contributions of this article are as follows:

•The multidimensional air quality indicator (M-AQI) big data integration method is proposed based on AQI sensing at three networking levels. First, crowdsourced AQI data is collected as the first level. Then data fusion of AQI data is considered at the edge-cloud-based level. Finally, AQI data are integrated on a remote cloud or meteorological supercomputing platform.

•An innovative healthcare monitoring system via urban big data (i.e., UH-BigDataSys) is proposed based on M-AQI big data and physiological data of a user. Compared to a traditional health monitoring system on the basis of physiological indices, the range of applications for UH-BigDataSys is broader. UH-BigDataSys is not only limited to traditional health monitoring; it can also give suggestions at a higher level in combination with the current health status of a user and their surrounding environment; for example, travel guidance to patients with respiratory diseases may be personalized by this system.

•The demonstration application platform for UH-BigDataSys is established. Typical applications include early forcasting for diseases related to urban air quality, travel guidance to patients with respiratory diseases, and monitoring for the user's emotions and sleep quality related to indoor air quality.

The remaining contents are arranged as follows. We introduce the method for the integration of M-AQI big data, we introduce the design of UH-BigDataSys, we discuss the testbed of UH-BigDataSys, and finally we give our conclusion.

M-AQI Big Data Integration Based on Crowdsourcing and Edge Clouds

In this section, we mainly consider two categories of smart city data related to the health status of urban residents (i.e., AQI data and users' body signals). To collect AQI data, we use various AQI sensing devices that are widely spread out in urban environments [13]. For the collection of body signals, wearable 2.0 devices (e.g., smart clothing) can be utilized. Figure 1 shows the procedure of the integration of M-AQI big data. The heterogeneous AQI data are integrated for data fusion and sharing in edge clouds and the data center cloud. With the accumulation of M-AQI data, the related big data analytics is conducted in the clouds. This section addresses issues in sensing and the fusion process of AQI data.

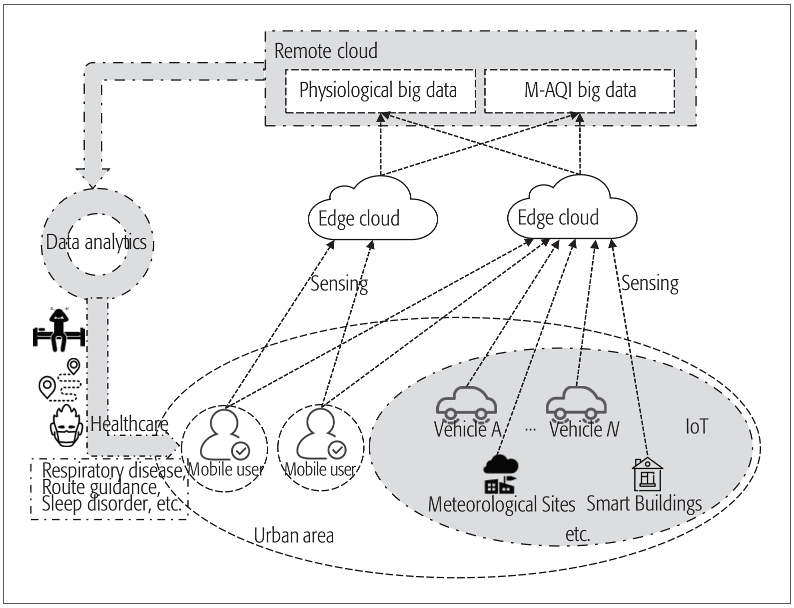

Figure 1. Flowchart of M-AQI big data integration.

AQI Data Sensing through Crowdsourcing

In Table 1, three categories of data are classified. There are six fundamental indicators in the category of "air quality data": PM2.5, PM10, CO, NO2, O3, and SO2. Through the use of mobile devices carried by urban residents, the technique of crowdsourcing is useful for sensing AQI data. With various advantages such as low cost, large sensing coverage, location awareness, and personalized data collection, it can be widely used in various aspects of daily life, such as traffic, environment monitoring, and healthcare. However, this method also exhibits some disadvantages such as poor data quality and intermittent data provisioning.

AQI sensing via crowdsourcing requires distributing numerous data sensing tasks to available caching and computing resources of mobile devices carried by urban residents. The specific steps of crowdsourcing are listed as follows. First, UH-BigDataSys should establish an association matrix using mobility data of urban residents such as GPS, acceleration, and moving speed. Based on the association matrix, mobility pattern and movement features of the residents are extracted. Then we can establish the residents' behavioral model. Finally, optimal selection can be achieved regarding those urban residents whose routes of mobility match the sensing points of AQI. Furthermore, UH-BigDataSys should choose those urban residents with higher credits (i.e., users who contribute high-quality sensing data, such as data quality evaluated by accuracy, timeliness, correlativity, integrity, etc.) and assign the data sensing

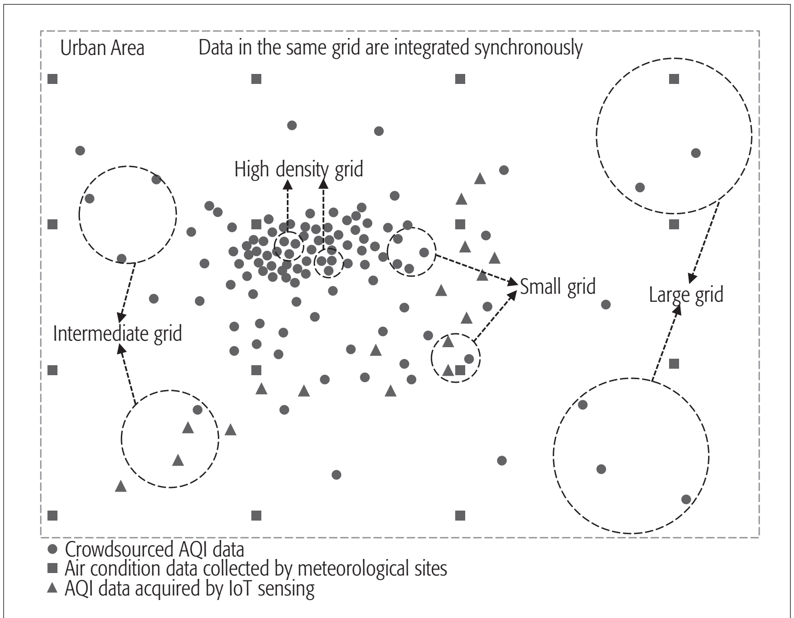

Figure 2. Schematic diagram for dynamic data integration based on time-space characteristic and data density distribution.

task to them.

Edge-Cloud-Based AQI Integration

As shown in Fig. 2, there are three categories of AQI data collected:

- AQI data sensed by urban residents via crowdsourcing
- AQI monitored by IoT sensing with monitoring facilities (vehicles, intelligent buildings, etc.)
- AQI data acquired at meteorological sites deployed by a meteorological agency

These data are synchronized to edge clouds through various communication modes. However, the diversity of data source causes heterogeneity of AQI data in time-space characteristic, data density distribution, and data accuracy. Heterogeneity of data density refers to the different densities of AQI data in different time slots or different regions. The heterogeneity of data accuracy refers to the difference of accuracies for data from different sources. Intuitively, the accuracy of AQI data from meteorological monitoring sites is highest. In comparison, AQI data acquired from mobile devices are low, especially under their high-speed movements. Thus, how to integrate multi-source and multi-quality AQI data is a challenging problem.

In this section, we propose an edge-cloud-based AQI integration method that can be conducted on an edge cloud and a remote cloud synchronously. First, we extract the location (i.e., longitude and latitude) feature and time feature of AQI data, as shown in Table 1. Then we partition those data based on time slot feature and determine the size of a grid on the basis of data density in a single time slot. The grid size is set to be smaller if the data density in that region is higher, thus guaranteeing a grid with finer granularity in those regions with high data density. Typically, those regions correspond to residential areas with highly dense population. Thus, it needs to provide precise data with finer granularity. Finally, for the heterogeneity of data accuracy, we use weighted average for the sake of simplicity,

$$v = \frac{w_1}{m}\sum_{i=1}^{m} v_i + \frac{w_2}{n}\sum_{i=1}^{n} v_i + \frac{w_3}{l}\sum_{i=1}^{l} v_i$$

where m represents the number of data segments in a certain time slot through crowdsourcing, n

Data category	Item	Typical value	Nullable
Air quality data	Location (Longitude)	114.3672	No
	Location (Latitude)	30.5719	No
	Time	2017-05-15 12:30:01	No
	PM2.5	29 g/m3	Yes
	PM10	–g/m3	Yes
	CO	1.3 mg/m3	Yes
	NO2	25 g/m3	Yes
	O3	180 g/m3	Yes
	SO2	14 g/m3	Yes
Physiological data	Location (Longitude)	114.3894	No
	Location (Latitude)	30.4822	No
	Time	2017-05-15 09:21:45	No
	ECG	1221 (ADC Sampling)	Yes
	EMG	2542 (ADC Sampling)	Yes
	Heart rate	74 (beats per minute)	Yes
	Body temperature	29°C	Yes
	Blood oxygen	98%	Yes

Data category	(Longitude)	(Latitude)	Time	AQI	PM2.5	PM10	CO	NO2	O3	SO2
M-AQI	114.3672	30.5719	2017-05-15(12:00-13:00)	75	29	60	1.3	25	180	14
	114.2511	30.5514	2017-05-15(12:00-13:00)	70	9	90	0.5	30	144	7
	114.2836	30.6197	2017-05-15(12:00-13:00)	52	9	53	0.5	21	141	8
	114.3006	30.5494	2017-05-15(12:00-13:00)	35	10	22	0.6	24	112	10

Table 1. Profile of sensing data.

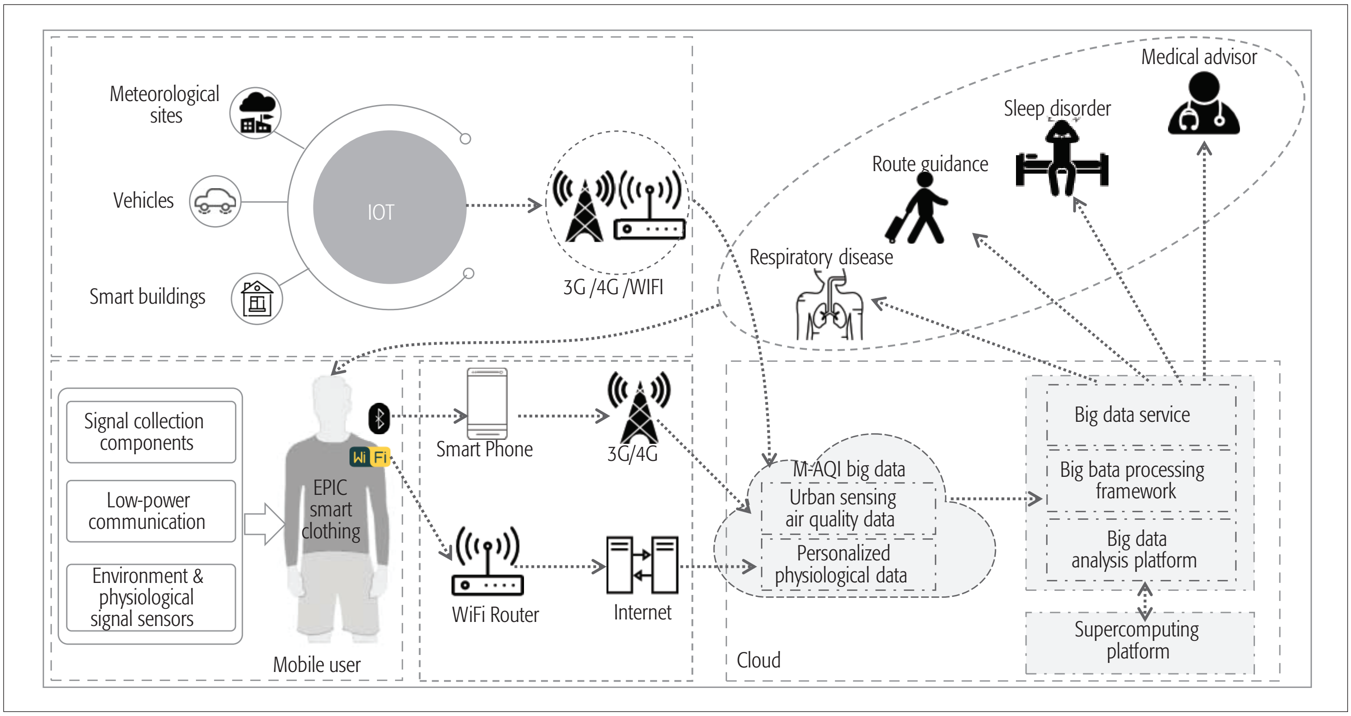

Figure 3. System architecture of UH-BigDataSys.

stands for the number of data segments collected from meteorological sites, l stands for the number of data segments via IoT sensing, v_i stands for the sampling value of an attribute (e.g., PM2.5), where w_1, w_2, and w_3 denote the weights of various measurements, and $w_1 + w_2 + w_3 = 1$. After the calculation is completed, combination for each attribute value in the same category in a certain time slot is conducted to generate a new M-AQI record. The profile of M-AQI data is shown in the M-AQI column in Table 1, from which the value of AQI will be obtained by converting the six air quality attributes.

AQI Integration on Remote Cloud or Supercomputing Platform

M-AQI data produced at edge clouds will be synchronized and shared to a remote cloud. Thus, data integration can also be conducted at a remote cloud with the same method as at edge clouds. There are stringent requirements in computing and storage for M-AQI big data applications. With limited resources, edge clouds cannot fulfill the requirements of authentic big data applications. With its particular performance in terms of convenience, scalability, and on-demand services, a remote cloud is able to provide ultimate guarantee for M-AQI big data analysis.

Furthermore, a supercomputing platform has massive high-performance computing resources, providing new possibilities for expanding the capability of the remote cloud. Because of the deficiency of the user-friendly interface, the current supercomputing platform cannot provide convenient and interactive services. By the construction of a virtual resource pool, uniform management of supercomputing resources may be formed, on the basis of which connectivity is convenient with affordable cost in terms of supercomputing platform or remote cloud. Furthermore, uniform management of supercomputing resources could define the virtualized interface, protocol, and software module for the supercomputing platform to abstract resources of the supercomputing platform into open services. Finally, the common user-oriented supercomputing resource services can be obtained, on the basis of which M-AQI big data services may provide urban residents with more accurate and personalized services.

Design of UH-BigDataSys

The architecture of UH-BigDataSys is shown in Fig. 3. Based on the acquired M-AQI big data, UH-BigDataSys also collects personalized physiological data of users via wearable 2.0 devices (e.g., smart clothing). Furthermore, the system jointly analyzes physiological big data of the user and M-AQI big data based on users' location information. Finally, it provides users health advice for respiratory diseases, outdoor travel, sleep quality control, and so on.

For physiological big data collection, smart clothing is a great choice for urban residents under various scenes [14]. It integrates textile clothing and body sensors, and exhibits good performance in terms of sensor deployment, a user's comfortability, and low-power communications. The physiological data acquired by smart clothing are shown in Table 1, including five fundamental indicators: electrocardiograph (ECG), electromyography (EMG), heart rate, saturation of blood oxygen, and body temperature. The ECG and EMG are acquired via textile dry electrodes in smart clothing. Heart rate can be figured out from original ECG signals. As for the saturation of blood oxygen, an optical sensor may be integrated into smart clothing to achieve non-invasive detection. Body temperature can be acquired by an NTC thermistor sensor. Finally, data processing modules in smart clothing receive and process original signals generated by each sensor, then convert them into digital signals and store them

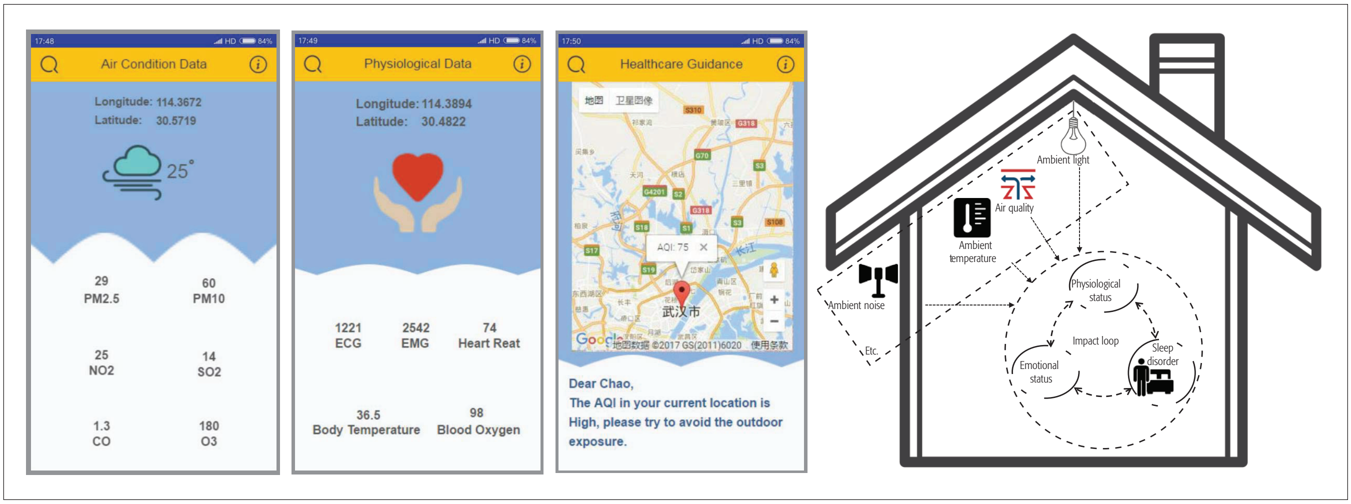

Figure 4. a) Air condition data; b) physiological data; c) healthcare guidance; d) impact of sleep disorder.

into its memory. A local processing unit determines whether to send physiological data to intelligent terminals or not according to the status of the communication module. Typically, the intelligent terminal is a smartphone to further transmit them to the cloud, thus finally realizing persistent storage for physiological data of the user in the cloud.

With sustainable sensing and analysis on physiological data of urban residents, UH-BigDataSys can provide guidance based on the physiological and mental status of the user. Furthermore, through the detection of air quality, UH-BigDataSys enriches the perception of the surrounding environment of urban residents. Finally, UH-BigDataSys can provide health guidance to urban residents with the combination of physiological data and air quality data around them. For example, in an outdoor environment, a resident with physiological diseases, especially respiratory disease, will get timely warning from UH-BigDataSys about air quality around him or her, reminding him or her to pay attention to air quality conditions to thus avoid deterioration of his or her respiratory disease. Meanwhile, through sensing of air quality data widely and dynamically in the city, UH-BigDataSys will also advise the user on outdoor activities, plan the outdoor route for the user when necessary, and guide the user to avoid an epidemic area or an area with high air pollution. As for the indoor environments, air quality would also exert influence on disease and emotion of humans. For example, dreary air would bring discomfort to the human body and exert influence on the sleep of residents at night. Based on physiological data of the user and indoor air quality data, UH-BigDataSys extracts the correlations among sleep, daily activities, air quality, and the health of the user. Furthermore, on the basis of analyzing the correlations among all factors, UH-BigDataSys constructs a measurement model to figure out the influence of air quality on sleep status of user.

A Testbed for UH-BigDataSys

We have deployed a testbed to evaluate the performance of UH-BigDataSys. The infrastructure we design for our testbed is a minimal cloud platform that consists of four different servers: one controller, one networker, and two computing nodes. The executive environment for our testbed is constructed on the basis of Openstack technology. We utilize Spring Framework to implement the UH-BigDataSys system and define all services as RESTful application programming interfaces (APIs). Finally, all kinds of clients access RESTful APIs to fetch all implemented services.

Figures 4 and 5 exhibit the user interface of our UH-BigDataSys testbed. Figures 4a and 5a show urban air quality data acquired by portable sensors carried by urban residents through mobile crowdsourcing. Figures 4b and 5b show physiological data of residents acquired by smart clothing. Figure 4c shows health advice given by UH-BigDataSys based on the acquired air quality around residents and the physiological status of residents.

As shown in Fig. 4d, the sleep quality of a resident is closely related to his or her daily activities, surrounding air quality and his or her physiological and psychological state. UH-BigDataSys perceives physiological information of residents such as heart rate, blood oxygen, body temperature, and exercise status via smart clothing, and obtains air quality data via crowdsourcing and IoT sensing, then utilizes big data analysis and machine learning to establish an efficient prediction model, which guides behaviors of residents to improve their sleep quality. We found that psychological state, physiological state, and sleep quality can affect each other.

Conclusion

In this article, we first discuss AQI data collection based on meteorological sites' data, mobile crowdsourcing sensing, and IoT sensing. Then the integration of M-AQI is proposed based on the edge clouds. M-AQI big data exhibits higher data quality and finer granularity. Next, physiological data of urban residents based on smart clothing is discussed, and UH-BigDataSys is proposed. The system analyzes M-AQI and physiological big data to provide guidance for urban residents in aspects such as respiratory disease, travel advice, and sleep quality in order to improve the quality of life of urban residents. Finally, a tested of UH-BigDataSys is established toward the smart applications for enhanced healthcare based on AQI in

Figure 5. Web interface of testbed: a) air quality data; b) physiological data.

urban environments.

Acknowledgment

This work is supported by the Deanship of Scientific Research at King Saud University, Riyadh, Saudi Arabia through the Research Group Project No. RGP-1436-023. Long Hu is the corresponding author.

References

[1] M. Chen *et al.*, "Disease Prediction by Machine Learning over Big Healthcare Data," *IEEE Access*, vol. 5, no. 1, 2017, pp. 8869–79.

[2] E. Fotopoulou *et al.*, "Linked Data Analytics in Interdisciplinary Studies: The Health Impact of Air Pollution in Urban Areas," *IEEE Access*, vol. 4, 2016, pp. 149–64.

[3] K. B. Shaban, A. Kadri, and E. Rezk, "Urban Air Pollution Monitoring System with Forecasting Models," *IEEE Sensors J.*, vol. 16, no. 8, 2016, pp. 2598–2606.

[4] C. Bermejo and P. Hui, "Steal Your Life Using 5 Cents: Hacking Android Smartphones with NFC Tags," arXiv preprint arXiv:1705.02081, 2017.

[5] D. Chatzopoulos and P. Hui, "ReadMe: A Real-Time Recommendation System for Mobile Augmented Reality Ecosystems," *Proc. 2016 ACM Conf. Multimedia*, 2016, pp. 312–16.

[6] D. Wu *et al.*, "ADDSEN: Adaptive Data Processing and Dissemination for Drone Swarms in Urban Sensing," *IEEE Trans. Computers*, vol. 66, no. 2, 2017, pp. 183–98.

[7] R. Tse and G. Pau, "Enabling Street-Level Pollution and Exposure Measures: A Human-Centric Approach," *Proc. 6th ACM Int'l. Wksp. Pervasive Wireless Healthcare*, 2016, pp. 1–4.

[8] B. Han, F. Qian, and L. Ji, "When Should We Surf the Mobile Web Using Both WiFi and Cellular?," *Proc. 5th Wksp. All Things Cellular: Operations, Applications and Challenges*, 2016, pp. 7–12.

[9] D. Tian *et al.*, "Robust Energy Efficient MIMO Transmission for Cognitive Vehicular Networks," *IEEE Trans. Vehic. Tech.*, vol. 65, no. 6, 2016, pp. 3845–59.

[10] K. Zheng *et al.*, "Design and Implementation of LPWA-Based Air Quality Monitoring System," *IEEE Access*, vol. 4,

2016, pp. 3238–45.
[11] B. Predić *et al.,* "Exposuresense: Integrating Daily Activities with Air Quality Using Mobile Participatory Sensing," *2013 IEEE Int'l. Conf. Pervasive Computing and Commun. Wksps.*, 2013, pp. 303–05.
[12] Y. Zheng, F. Liu, and P. H. Hsieh, "U-air: When Urban Air Quality Inference Meets Big Data," *Proc. 19th ACM SIGKDD Int'l. Conf. Knowledge Discovery and Data Mining*, 2013, pp. 1436–44.
[13] K. Hu *et al.*, "Design and Evaluation of a Metropolitan Air Pollution Sensing System," *IEEE Sensors J.*, vol. 16, no. 5, 2016, pp. 1448–59.
[14] M. Chen *et al.,* "Wearable 2.0: Enable Human-Cloud Integration in Next Generation Healthcare System," *IEEE Commun.*, vol. 55, no. 1, Jan. 2017, pp. 54–61.
[15] M. Chen, P. Zhou, and G. Fortino, "Emotion Communication System," *IEEE Access*, vol. 5, 2017, pp. 326–37.

Biographies

Min Chen [SM'09] (minchen2012@hust.edu.cn) has been a full professor in the School of Computer Science and Technology at Huazhang University of Science and Technology (HUST) since February 2012. He is Chair of the IEEE Computer Society STC on Big Data. His Google Scholars Citations reached 11,700+ with an h-index of 53. He received the IEEE Communications Society Fred W. Ellersick Prize in 2017. His research focuses on cognitive computing, IoT sensing, 5G networks, emotion detection, healthcare big data, and so on.

Jun Yang (junyang_cs@hust.edu.cn) received his Bachelor's and Master's degrees in software engineering from HUST in 2008 and 2011, respectively. Currently, he is a Ph.D candidate with the Embedded and Pervasive Computing (EPIC) Lab in the School of Computer Science and Technology, HUST. His research interests include cognitive computing, software intelligence, the Internet of Things, cloud computing, big data analytics, and so on.

Long Hu (hulong@hust.edu.cn) is a lecturer in the School of Computer Science and Technology at HUST. He also received his Ph.Dd, Master's, and B.S. degrees from HUST. Currently, his research includes 5G mobile communication systems, big data mining, marine-ship communication, the Internet of Things, multimedia transmission over wireless networks, and so on.

M. Shamim Hossain [SM'09] (mshossain@ksu.edu.sa)) is a full professor in the Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia. He has co-authored more than 165 publications. He is the recipient of the ACM TOMM Nicolas D. Georganas Best Paper Award. He currently serves on the Editorial Board of *IEEE Multimedia*. His research focuses on cloud networking, social media, IoT, cloud media for healthcare, and smart health.

Ghulam Muhammad (ghulam@ksu.edu.sa) is a full professor in the College of Computer and Information Sciences at King Saud University. His research interests include image, speech, and signal processing, He has authored and co-authored many publications including papers in refereed journals, conference papers, books, and book chapters. He has supervised a number of Ph.D. and Master's theses. He owns a U.S. patent.