



Medical Data Processing and Analysis for Remote Health and Activities Monitoring

Salvatore Vitabile¹✉, Michal Marks², Dragan Stojanovic³,
Sabri Pllana⁴, Jose M. Molina⁵, Mateusz Krzyszton², Andrzej Sikora²,
Andrzej Jarynowski⁶, Farhoud Hosseinpour⁷, Agnieszka Jakobik⁸,
Aleksandra Stojnev Ilic³, Ana Respicio⁹, Dorin Moldovan¹⁰,
Cristina Pop¹⁰, and Ioan Salomie¹⁰

¹ University of Palermo, Palermo, Italy
salvatore.vitabile@unipa.it

² Research and Academic Computer Network, Warsaw, Poland
{michal.marks,mateusz.krzyszton,andrzej.sikora}@nask.pl

³ University of Nis, Niš, Serbia
{dragan.stojanovic,aleksandra.stojnev}@elfak.ni.ac.rs

⁴ Linnaeus University, Växjö, Sweden
sabri.pllana@lnu.se

⁵ Universidad Carlos III de Madrid, Madrid, Spain
molina@ia.uc3m.es

⁶ Interdisciplinary Research Institute in Wrocław, Wrocław, Poland
andrzej.jarynowski@doctoral.uj.edu.pl

⁷ University of Turku, Turku, Finland
farhoud.hosseinpour@utu.fi

⁸ Cracow University of Technology, Kraków, Poland
ajakobik@pk.edu.pl

⁹ Universidade de Lisboa, Lisbon, Portugal
alrespicio@ciencias.ulisboa.pt

¹⁰ Technical University of Cluj-Napoca, Cluj-Napoca, Romania
Ioan.Salomie@cs.utcluj.ro

Abstract. Recent developments in sensor technology, wearable computing, Internet of Things (IoT), and wireless communication have given rise to research in ubiquitous healthcare and remote monitoring of human's health and activities. Health monitoring systems involve processing and analysis of data retrieved from smartphones, smart watches, smart bracelets, as well as various sensors and wearable devices. Such systems enable continuous monitoring of patients psychological and health conditions by sensing and transmitting measurements such as heart rate, electrocardiogram, body temperature, respiratory rate, chest sounds, or blood pressure. Pervasive healthcare, as a relevant application domain in this context, aims at revolutionizing the delivery of medical services through a medical assistive environment and facilitates the independent living of patients. In this chapter, we discuss (1) data collection, fusion, ownership and privacy issues; (2) models, technologies and solutions for

medical data processing and analysis; (3) big medical data analytics for remote health monitoring; (4) research challenges and opportunities in medical data analytics; (5) examples of case studies and practical solutions.

Keywords: e-Health · Internet of Things (IoT) · Remote health monitoring · Pervasive healthcare (PH)

1 Introduction

Recent developments in sensor technologies, wearable computing, Internet of Things (IoT), and wireless communications have given rise to research on mobile and ubiquitous health-care and remote monitoring of peoples health and activities [61]. Health monitoring systems include processing and analysis of data retrieved from smart-phones, smart watches, smart bracelets (that is, wristbands), as well as various connected sensors and wearable devices [69]. Such systems enable continuous monitoring of patients psychological and health conditions by sensing and transmitting measurements such as heart rate, electrocardiogram (ECG), body temperature, respiratory rate, chest sounds, and blood pressure. Collected, fused, and analyzed sensor data are important for diagnosis and treatment of patients with chronic diseases (such as, hypertension and diabetes) or for monitoring and assistance of elderly people. The area of health informatics is explored by researchers with different academic backgrounds: computer scientists, physicians, mathematicians, statisticians, and sociologists. All of them have something to contribute: from medical knowledge, though computer science (that is, simulations and data analysis), to sociological and marketing know-how (such as, apps dissemination and social interventions). There are mathematical/computational models (for instance, differential equations or system dynamics) that can be involved in understanding of health-relevant processes. For example, infectious disease models can use sensoric data that are collected from human contacts; these models can be useful in risk assessment of patients.

This chapter addresses major challenges and opportunities related to the medical data collection, modelling and processing in the context of monitoring of human's health and behavior. To illustrate the use of discussed concepts and technologies we present three case studies. The first case study describes a Big Data solution for remote health monitoring that considers challenges pertaining to monitoring the health and activities of patients. We present the design and implementation of a remote health monitoring system architecture that may be employed in a public, private or hybrid cloud. The second case study describes a system designed to support distributed monitoring of human health during a trip. It is assumed that a travel agency that organizes a trip to high mountains equips the guide and tourists with sensors that monitor health parameters. The third case study describes the use of machine learning in the context of monitoring daily living activities of people with dementia. Detected anomalies

in the daily living activities of people with dementia may help medical experts to distinguish between health decline symptoms caused by dementia and health decline symptoms incurred by the side effects of the prescribed medications.

The rest of this chapter is structured as follows. Section 2 discusses data collection, fusion, ownership and privacy issues. Models, technologies and solutions for medical data processing and analysis are addressed in Sect. 3. Section 4 discusses big medical data analytics for remote health monitoring. Research challenges and opportunities in medical data analytics are addressed in Sect. 5. Section 6 describes examples of case studies and practical scenarios. The chapter is concluded in Sect. 7.

2 Medical Data Collection, Fusion, Ownership and Privacy

2.1 Medical Data Collection

Medicine has always been trying to measure properties of patients (such as, clinical picture) and how to understand their health and disease. Computer science would then incorporate the results of these measurements, experiments and observation into models. Recently, the utilisation of IT tools for medicine (such as, machine learning) has undergone an accelerating growth, however all models of such system are incomplete without real data, especially register-based. For example with patient Electronic Health Record/Documentation (EHR/D) data can be provided with reasonably accuracy on clinical picture, procedures, or co-morbidity. International standard - HL7 (Health Level Seven) allows interoperability between IT providers and solutions. However, still in many countries most of medical data is analog and must be digitized (e.g. image and sound processing). Moreover, powerful computer facilities and ubiquitous wearable devices with multiple sensors have made possible the collection, processing and storage of data on individual or group of patients. Individual data can be used for risk assessment for a particular health disorder. For example the risk of being affected by an infectious disease can be calculated based on sensor data and questionnaires [36]. Data collected from group of patients potentially being in contact can be used for infectious disease risk assessment in given community. There is concern about Data Protection Act in EU from legal perspective about providing such kind of data analysis, which will be described in detail in this chapter. The consent of each patient should be requested, which could be very difficult to obtain.

2.2 Blockchain for Privacy and Security

The main application of Blockchain in medicine is traceability of drugs in many actors setup. Such a technique has been already applied in Sub-Saharan Africa and South-East Asia for verifying originality of drugs (fake drugs are big problem there). Blockchain technology can also help in investigating food-borne outbreaks, where tractability of items in supply chain with storage/transport characteristics is crucial.

2.3 Data Fusion

Data fusion (DF) is a multi-domain growing field aiming to provide data for situation understanding. Globally, threat detection and facility protection are some of the vital areas in DF research [49]. Fusion systems aim to integrate sensor data and information in databases, knowledge bases, contextual information, user mission, etc., in order to describe dynamically-changing situations [32]. In a sense, the goal of information fusion is to attain a real-time simulation of a subset of the world based on partial observations of it [50].

The ability to fuse digital data into useful information is hampered by the fact that the inputs, whether device-derived or text-based, are generated in different formats, some of them unstructured. Whether the information is unstructured by nature or the fact that information exchange (metadata) standards are lacking (or not adhered to) all of this hinders automatic processing by computers. Furthermore, the data to be fused may be inaccurate, incomplete, ambiguous, or contradictory; it may be false or corrupted by hostile measures. Moreover, much information may be hard to formalize, i.e., imaging information. Consequently, information exchange is often overly complex. In many instances, there is a lack of communication between the various information sources, simply because there is no mechanism to support this exchange.

The high data flow, either device-based or text-based, is unable to process and leads to time delays, extra costs, and even inappropriate decisions due to missing or incomplete information [13, 63].

The key aspect in modern DF applications is the appropriate integration of all types of information or knowledge: observational data, knowledge models (a priori or inductively learned), and contextual information [28, 73]. Each of these categories has a distinctive nature and potential support to the result of the fusion process.

- **Observational Data:** Observational data are the fundamental data about the individual, as collected from some observational capability (sensors of any type). These data are about the observable characteristic of a person that are of interest [40].
- **Contextual Information:** Context and the elements of what could be called Contextual Information could be defined as “the set of circumstances surrounding the acquired data that are potentially of relevance to its completion.” Because of its data-relevance, fusion implies the development of a best-possible estimate taking into account this lateral knowledge [29]. We can see the context as background, i.e., not the specific entity, event, or behaviour of prime interest but that information which is influential to the formation of a best estimate of these items [27].
- **Learned Knowledge:** In those cases where a priori knowledge for DF process development cannot be formed, one possibility is to try and excise the knowledge through online machine learning processes operating on the observational and other data [59]. These are procedural and algorithmic methods for discovering relationships among concepts of interest that being captured by the system (sensor data and contextual information). There is a tradeoff

involved in trying to develop fully-automated algorithmic DF processes for complex problems where the insertion of human intelligence at some point in the process may be a much more judicious choice [23].

At this time, there are a multitude of problem-specific solutions to fusion, most of which are centred in specific applications, producing problem-specific results [26]. The ability to automatically combine the results of smaller problems into a larger context is still missing.

2.4 Medical Data Security Requirements Enforced by the Law and Related Ownership and Privacy Issues

All organizations that are collecting, processing and storing medical are obliged by the international and local law regulations and standards to improve their data protection strategies. The example of data protection laws required by USA is the The Health Insurance Portability and Accountability Act (HIPAA)¹, used for compliance and secure adoption of electronic health records (EHR).

HIPAA sets the rules for sensitive patient data protection with its Privacy and Security Rules. According to HIPA, a company that processes protected health information must store it in a system that is secure as far as physical, network, and process security is concerned. Security Rule of HIPAA defines requirements for health information that is held or transferred in electronic form. It requires that HIPAA-covered entities implement the following protections:

- Ensure the confidentiality, integrity, and availability of all e-PHI they create, receive, maintain or transmit;
- Identify and protect against reasonably anticipated threats to the security or integrity of the information;
- Protect against reasonably anticipated, impermissible uses or disclosures; and
- Ensure compliance by their workforce.

Additionally, it regulates technical safeguards when hosting sensitive patient data, including facility access in place, as well as policies for using workstations, electronic media, and all processes of transferring, removing, disposing, and re-using electronic media or electronic information. It enforces authorized unique user IDs, emergency access procedures, automatic log off, encryption of data, and audit reports. Also, activity tracking logs of all activity on hardware and software are necessary.

Health-care organizations must ensure the security and availability of patients data for both health-care professionals and patients.

The complete suite of HIPAA Administrative Simplification Regulations can be found at 45 CFR Part 160, Part 162, and Part 164, and includes:

- Transactions and Code Set Standards;
- Identifier Standards;

¹ <https://www.hhs.gov/hipaa>.

- Privacy Rule;
- Security Rule;
- Enforcement Rule;
- Breach Notification Rule.

General Data Protection Regulation (GDPR)² is a regulation in EU law on data protection and privacy within the European Union (EU) and the European Economic Area (EEA). It also specifies the conditions for exporting of personal data outside the EU and EEA areas. It specifies:

- Rights of the data subject
- Controller role
- Transfers of personal data to third countries or international organizations
- Independent supervisory authorities
- Cooperation and consistency
- Remedies, liability and penalties
- Delegated acts and implementing acts

It specifies actions to be taken that enforces the processing on systems that are processing medical data. Several examples may be:

- Data holders must notify the authorities in case of breach;
- Subjects have a right to access their data;
- The right to be forgotten that gives individuals the power to request the removal of their personal data;
- Privacy must be designed in;
- Privacy is included by default, not by unchecking a box;
- Right to rectification, the data shall have the right to obtain from the controller without undue delay. The data subject have the right to have incomplete personal data completed, including by means of providing a supplementary statement.
- Every organization holding personal third party data must have a data protection officer.

The GDPR imposes stiff fines on data controllers and processors for non-compliance, up to 20 million, or 4% of the worldwide annual revenue of the prior financial year, in case of breaking the basic principles for processing medical data.

To obtain the HIPAA required security level, the company may follow two paths:

1. choose a reputable HIPAA training company that offers certification credentials, and analyze the company procedures and systems by itself;
2. obtain evaluation by an independent, third party auditor, like for instance, Coalfire Systems³ or ComplySmart⁴.

² <https://eur-lex.europa.eu/eli/reg/2016/679/oj>.

³ <http://www.coalfire.com>.

⁴ <http://www.complysmart.com>.

For ensuring GRPD standards, the company may:

1. obtain the customized GDPR certificate for protection officers⁵;
2. hire authorized experts⁶;

For both HIPAA and GDPR compliance there is no one main central independent party that may certify the company.

GDPR specific challenges rises inside remote systems. First of all, under the GDPR personal data may not be stored longer then needed. This is why, retention procedures have to be implemented and when the data expired they have to be removed from systems. Data can be stored on multiple locations, under multiple vendors, processed by many services. The deletion of data completely have to consider also backups and copies of data stored on the remote equipment. Additionally, breaching response may also be the issue. Breach notification have to be shared among data holders. A breach event have to be well defined.

Processing of personal data outside the European Economic Area (EEA) is the next problem due to multiple location. Controllers will need to define a multi-country data strategy taking into account localization laws.

The transparency of security procedures of medical systems or third party certificates are necessary to ensure security controllers about the quality of security services.

Medical service providers must be subject of an audit to perform a control framework with privacy and privacy by design control measures.

Medical services should be monitored to address any changes in technology and recommended updates to the system. It includes newly introduced equipment, and sensors.

During processing large set of data, visibility regarding metadata and data anonymization is a big challenge. The level of protection of metadata, the respective ownership rights, rights to process the collections of metadata, intended uses of metadata should be examined carefully.

2.5 Publicly Available Medical Datasets and Reuse of Already Collected Data

Comprehensive and real-world personal health and activities datasets have a very important role in data processing and analysis; therefore several attempts have been made to create big and representative examples of real-world datasets [67].

Examples of such datasets are UbiqLog [68] and CrowdSignals [82], which contains both the data from the smartphones, and from wearable devices, smart watch/bracelet.

Biomedical signals datasets can be found at PhysioNet⁷ which offers free Web access to large collections of recorded physiologic signals (PhysioBank) and related open-source software for collecting them (PhysioToolkit).

⁵ <https://www.eugdpr.institute/gdpr-certification/>.

⁶ <https://www.forcepoint.com>.

⁷ www.physionet.org.

The ExtraSensory dataset contains data from 60 users and more than 300K labeled examples (minutes) originated from smartphone and smartwatch sensors. The dataset is publicly available for context recognition research [77].

Within already implemented Hospital Information System (HIS) framework a lot of data is collected (e.g. for administrative purposes), but not have been analysed from patients health perspective. Organizational structure as dedicating staff to patients, transport paths rerouting, shift distributions, geo-localizing isolated/cohorted patients, etc. can be used for example in infectious disease field. There are models and tools describing the possible spread of pathogens and could indicate an effective strategy in the fight against infectious diseases⁸, however available solutions are not using HPC as it would be possible.

3 Models, Technologies and Solutions for Medical Data Processing and Analysis

Usability of e-health can rely on collecting techniques and data analysis. Various data sets (survey, diagnosis and tests, time series, spatial, panel, longitudinal data, etc.) shall be analysed by different methods (regressions, decision trees, structural equation modelling, social network analysis, agent-based modelling, machine learning, etc.).

3.1 Innovative Use of Commercial Electronic Devices (Smartphones and Watches, Bracelets, Wearables,...) for Remote Monitoring

Low-power and low cost electronic devices can be easily purchased and used. Some properties as body temperature and blood pulse can be measured with high accuracy, but developer are trying to target more and more features. FDA (Federal Drug Agency) as well as EMA (European Medicines Agency) certify digital Health Software and Devices. Most of devices on the market do not satisfy minimum conditions for certification and accreditation. Some of them as air quality monitoring devices (e.g. PM2.5 laser dust sensor for less than 5\$) have been massively used in heavily polluted cities. However no medical organization recommends using home devices due to very low data quality and data provided from certified official stations should be enough. More data is not always better as it is shown in this example.

Some issues come with digital health apps and software (mhealth). Nowadays an app can be found for almost everything. Let's consider advances in image recognition using deep learning by a smartphone for measuring blood pressure by photoplethysmography. The biggest advantage of this method is simplicity, so nothing more than smartphone is needed. However, even it works well after calibration for most users, there are patients for whom the error of this procedure can reach 30%. So apps offering this feature (like icare and HDCH) were taken away from Google Play and App Store, but their equivalents are still very popular, for example in China.

⁸ www.sirsz.pl.

Another type of sensors are bluetooth beacons to track social interactions that can be used for infectious disease control and psychological health (one of certified suppliers for these beacons is kontakt.io). Smartphone or wearable sensors can communicate between each other and collect contact data.

3.2 IoT Platforms for e-Health Applications

Thingier.io. The IoT platform is an open source alternative. Thingier.io is relatively new to the IoT ecosystem but is being used extensively in different research projects [25, 43, 48], or even for education [38]. It provides a ready to use cloud service for connecting devices to the Internet to perform any remote sensing or actuation over the Internet. It offers a free tier for connecting a limited number of devices, but it is also possible to install the software outside the cloud for a private management of the data and devices connected to the platform, without any limitation.

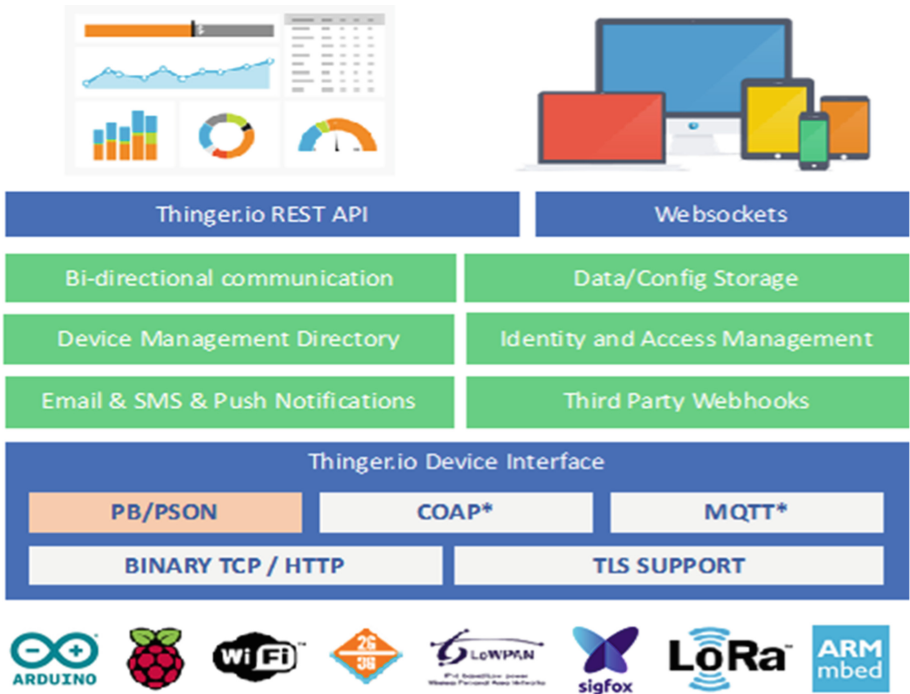


Fig. 1. Thingier.io platform

This platform is hardware agnostic, so it is possible to connect any device with Internet connectivity, from Arduino devices, Raspberry Pi, Sigfox devices, Lora solutions over gateways, or ARM devices, to mention a few. The platform

provides some out of the box features like device registry; bi-directional communication in real-time, both for sensing or actuation; data and configuration storage, so it is possible to store time series data; identity and access management (IAM), to allow third party entities to access the platform and device resources over REST/Websocket APIs; third party Webhooks, so the devices can easily call other Web services, send emails, SMS, push data to other clouds, etc. It also provides a web interface to manage all the resources and generate dashboards for remote monitoring. The general overview of this platform is available at Fig. 1. The main benefit of using this platform, aside that it is open source, is the possibility to obtain a bi-directional communication with the devices, in real-time, by using standard REST-API interfaces. This way, it is possible to develop any application, i.e., desktop, mobile, Web service, that interacts with devices by using a well-known and proven interface based on REST [62]. Meanwhile, the devices can use more efficient (in terms of bandwidth, or memory footprint) binary protocols to communicate with the cloud. In the other way, this platform provides client libraries for connecting several state of the art IoT devices like Arduino-based, ESP8266, ESP32, LinkitOne, Texas Instruments CC3200, Libelium Wasmotes, Raspberry Pi, etc. The client libraries provide a comprehensive way of connecting devices and sending information to the cloud, without having to deal with complex IoT protocols.

Libelium MySignals. This is a development platform for medical devices and eHealth applications. The platform can be used to develop eHealth web applications and to test own sensors for medical applications. MySignals is an example of commercial product which is offered and supported by a Spanish company called Libelium⁹. It allows measurement of more than 20 biometric parameters such as pulse, breath rate, oxygen in blood, electrocardiogram signals, blood pressure, muscle electromyography signals, glucose levels, galvanic skin response, lung capacity, snore waves, patient position, airflow and body scale parameters. Data gathered by sensors can be stored in the MySignals or third party Cloud to be visualized in external Web and mobile app's. Libelium offers an API for developers to access the information as well as Open source HW version which is based on Arduino. One of the drawbacks is the restriction for sending the information coming from MySignals to a third party cloud server using directly WiFi radio which is limited only for HW version – option is not available for SW version running on libelium Atmega 2560 node.

FIWARE. FIWARE¹⁰ is an open source IoT platform “combining components that enable the connection to IoT with Context Information Management and Big Data services in the cloud” [8]. FIWARE uses rather simple standard APIs for data management and exchange that can be used to develop smart applications. It provides enhanced OpenStack-based cloud hosting capabilities

⁹ www.libelium.com/product/mysignals.

¹⁰ <https://www.fiware.org/>.

and a number of components for functions offered as a Service. The adoption of FIWARE technology by the eHealth market is promoted by the industrial accelerator FICHe project [9] launched by European Commission and several industrial projects that have been developed in the FICHe context. CardioWeel [6] is an advanced driver assistance system that provides information about drivers health status using the human electrocardiogram (ECG) acquired from the drivers hands. This system is hosted on FIWARE cloud using the available virtual machines. The development of healthcare applications using FIWARE components is described in [75] and in [19].

4 Big Medical Data Analytics for Remote Health Monitoring

With the advance of remote health monitoring and pervasive healthcare concepts, an increased research work has been published to cover the topics ranging from theory, concepts, and systems, to applications of Big medical data systems for ubiquitous healthcare services.

4.1 Big Medical Data System Architecture

The analysis and health problem detection can be performed on a mobile device (phone, watch, bracelet, etc.) leveraging edge computing principles, or at a nearby computing infrastructure, e.g. IoT gateway, or a home/hospital server [22].

Physiological, health and social media data, providing insight into people activities and health conditions, could be enriched and correlated with external and environmental data collected within Smart City infrastructure (weather conditions, environment pollution/noise, temperature, city events, traffic conditions, etc.). As such, fusion, analysis and mining of IoT medical data and external/environmental data could better detect potential health problems and their cause-effect relationships with the environment.

The fusion, processing and analytics of sensor data, collected from personal mobile and health devices, as well as Smart city infrastructure are performed at the edge/fog computing components providing efficiency and minimal latency in detection of critical medical conditions that requires prompt actions. Also, this can provide a personalized health system for general well-being where individuals can be provided with healthcare tailored to their needs.

Moreover, such system should support efficient collection and analysis of massive quantities of heterogeneous and continuous health and activities data (Big Data) from a group, or a crowd of users. The storage, aggregation, processing and analysis of Big health data could be performed within public, private or hybrid cloud infrastructure. The results of data analysis and mining are provided to physicians, healthcare professionals, medical organisations, pharmaceutical companies, etc. through tailored visual analytics, dashboard applications. There is a number of research and development challenges that must be addressed

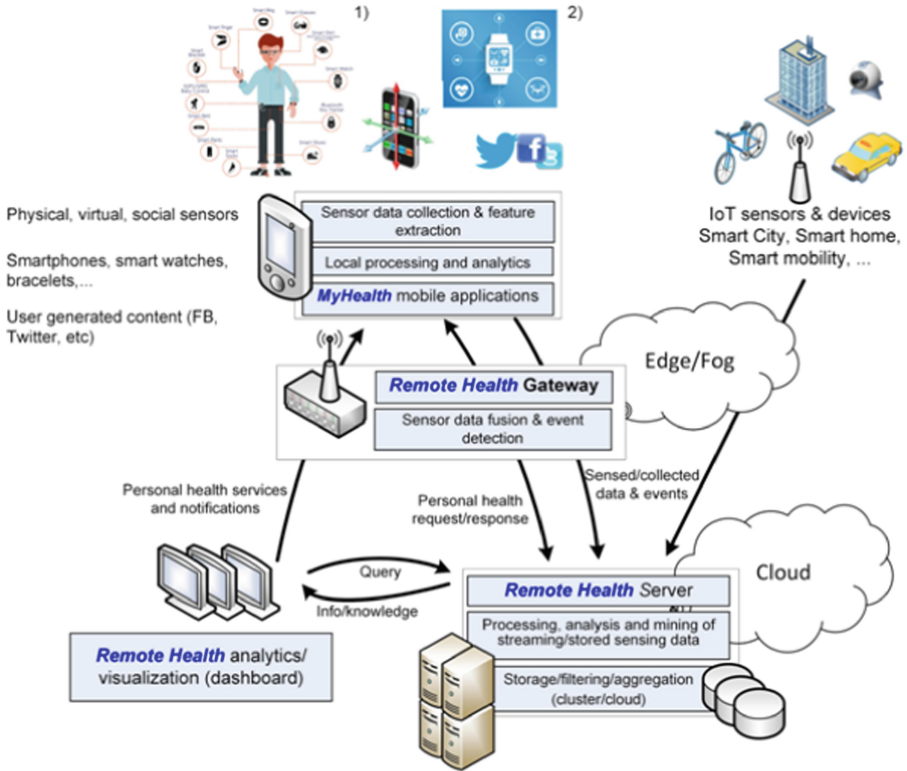


Fig. 2. A system for remote monitoring of people health and activities - general architecture

prior to wider application of such personalized healthcare systems that continuously monitor peoples health and activities and respond appropriately on critical events and conditions. These include, but are not limited to security, privacy and data ownership, sensor data fusion, scalable algorithms and systems for analytics and data mining, and edge-cloud models for processing and analytics.

The general architecture of a large-scale distributed system based on medical IoT devices and Big Data processing and analytics for remote monitoring of peoples health and activities is given in Fig. 2.

Remote monitoring systems can operate at multiple scales, providing personal and global sensing. As illustrated in [44], there are three distinct scales for sensing: personal, group and community sensing.

Personal or individual monitoring systems are designed for single individual, and are often focused on personal data collection and analysis. Typical scenarios include tracking the users exercise routines, measuring activity levels or identification of symptoms connected with psychological disorders. Although collected data might be for sole consumption of the user, sharing with medical

professional is common. Popular personal monitoring technologies are Google GlassTM, FitBitTM and The Nike+ FuelBandTM.

If individuals share a common interest, concern or goal while participating in monitoring applications they form a group. Group monitoring systems can be popular in social networks or connected groups, where data can be shared freely, or with privacy protection. Common examples are health monitoring applications including contests related to specific goals: running distance, weight loss, calorie intake, etc.

If a number of people participating in health and activities monitoring is very large, it is called community or crowd monitoring. Crowd modeling and monitoring implies collective data analytics for the good of the community. However, it involves cooperation of persons who will not trust each other, highlighting the need for strong privacy protection and possibly low commitment levels from users. Examples of community monitoring involve tracking the spread of diseases across area, finding patterns for specific medical conditions etc. The impact of scaling to monitoring applications is to be explored, but many research issues related to information sharing, data ownership, data fusion, security and privacy, algorithms used for data mining, providing useful feedback, etc. remain open.

Khan et al. in [41] presented the latest research and development efforts and achievements in the field of smart healthcare regarding high performance computing (HPC) and large-scale healthcare architectures, data quality and large-scale machine learning models for smart healthcare, Internet-of-Things, fog computing, and m-Health, as well as wearable computing, Quality of Service (QoS), and context-awareness for smart healthcare.

The challenges in designing algorithms, methods and systems for healthcare analytics and applications have been examined in [45], along with a survey on smart healthcare technologies and solutions. The next-generation healthcare applications, services and systems related to Big healthcare data analytics are reviewed and challenges in developing such systems are discussed.

The current state and projected future directions for integration of remote health monitoring technologies into clinical and medicine practice have been presented in [31]. Several of the challenges in sensing, analytics, and visualization that need to be addressed before systems can be designed for seamless integration into clinical practice are highlighted and described.

4.2 Big IoT Data for Remote Health Monitoring

With the development of Internet of Things (IoT) concepts, the idea of IoT healthcare systems has been an interesting topic for large number of researchers. Baker et al. in [16] have presented state-of-the-art research related to wearable IoT healthcare system. A standard model for application in future IoT healthcare system was proposed, along with an overview of existing challenges including security, privacy, wearability and low-power operation. Also, the authors reviewed several research works that address IoT healthcare challenges related to rehabilitation, diabetes management, disease monitoring, heart-attack detection and general activity and health monitoring purposes. Furthermore, a four-

part model was proposed as an aid in the development of future IoT healthcare systems, and it includes:

- wearable sensor and central nodes,
- short range communications,
- long range communications,
- secure cloud storage architecture and machine learning.

Islam et al. in [34] have given an overview of existing IoT-based healthcare network studies, and state-of-the-art network architectures, platforms, applications, and industrial trends in this area. Furthermore, they highlighted security and privacy issues and proposed a security model aiming to minimize security risk related to health care. Specifically, they presented an extensive overview of IoT healthcare systems, one of them being the Internet of m-Health Things (m-IoT): an operating symbiosis of mobile computing, medical sensors, and communications technologies for healthcare services.

An important part of IoT healthcare is obtaining insights from large data generated by IoT devices. The focus of [53] was on IoT architecture, opportunities and challenges, but from the data analysis point of view. In this paper a brief overview of research efforts directed toward IoT data analytics, as well as the relationship between Big Data analytics and IoT were given. Furthermore, analytic types, methods and technologies for big IoT data mining are discussed.

In [77], authors have used smartphone and smartwatch sensors to recognize detailed situations of people in their natural behavior. Initial tests were conducted over labeled data from over 300k minutes from 60 subjects. A dedicated application for data retrieval was implemented and presented. As stated in the paper, the main contribution is the emphasis on in-the-wild conditions, namely naturally used devices, unconstrained device placement, natural environment and natural behavioral content. The main challenge when recognizing context in non-controlled conditions is high diversity and variance of the data. However, it was shown that everyday devices, in their natural usage, can capture information about a wide range of behavioral attributes.

In [57] a distributed framework based on the IoT paradigm has been proposed for monitoring human biomedical signals in activities involving physical exertion. Also, a validation use case study was presented which includes monitoring footballers heart rates during a football match, and it was shown that data from BAN devices can be used to predict not only situations of sudden death but also possible injuries.

Ma et al. in [52] have presented a Big health application system based on health IoT devices and Big Data. The authors presented the architecture of health care application leveraging IoT and Big Data and introduced technology challenges and possible m-health applications based on this architecture.

One important subset of the data used for healthcare and improved quality of life are the data retrieved from healthcare services running on smartphones such as Endomondo, Nike+, RunKeeper, and Runtastic. These applications can provide data about different lifestyles, such as sleep, diet and exercise habits all of

which can be correlated with various medical conditions. In such manner, physicians can calculate risk for conditions or in cases of diagnosed ones, can adapt treatments or provide emergency responses when needed. Cortes et al. in [21] have provided an extensive analysis of traces gathered from Endomondo sport tracker service. The main Big Data challenges associated with analysis of the data retrieved from this type of applications were discussed. They include data acquisition (data redundancy, workload, flow rate), data cleaning (invalid and uncertain data from sensors), data integration, aggregation and representation (define common data representation across various applications, spatio-temporal properties of the data), query processing, data modeling and analysis (queries are important aspect of knowledge discovery), and interpretation (online and offline analysis).

The concept of smart health (s-Health), which integrates context-aware mobile health principles with sensors and information originated in smart cities, has been introduced in [74]. The authors provided a detailed overview of the smart health principles and discussed the main research challenges and opportunities in augmentation of smart healthcare with smart city concepts.

4.3 Processing and Analysis of Big Health and Mobility Data Streams

The healthcare industry generates large amounts of data, driven by record keeping, compliance and regulatory requirements and patient care [65]. The current trend is toward rapid digitalization of these massive amounts of data, and their fusion with data retrieved from personal and mobility sensors. When it comes to health care in terms of patient records, treatment plans, prescription information etc., everything needs to be done quickly, accurately and in some cases transparently enough to satisfy stringent industry regulations. However, with extended datasets, effective data processing and analysis can uncover hidden insights that improve patient care. If the data is collected out of the medical offices, they can model a patient's behavior more accurately. Furthermore, the results can be used to provide better real-time healthcare at lower cost. Real-time component requires that systems for analysis and processing must deal not only with huge amounts of data, but also with Big Data streams. This requirement has led to a proliferation of studies related to processing and analysis of big, heterogeneous streams for healthcare purposes. These streams have all main characteristics of Big Data:

- Volume: in order to create a complete overview of person's or group health, it is necessary to take into account data collected from related sources, obtained not only from medical instruments and BAN sensors, but also from social media, mobile devices or machine-to-machine data. This data can be from terabytes to exabytes, does not need to be stored, but must be effectively and analyzed and processed in timely manner.
- Velocity: Data streams with unparalleled rate and speed are generated and must be processed and analysed accordingly. In order to be effective, and to

provide responses in emergency situations, healthcare systems must process and analyze torrents of data in real time, but also be capable to perform long-term batch operations.

- Variety: Data comes in all varieties, from structured, numeric data obtained from sensors and medical devices to unstructured text documents, email, video, audio and mobility data. Systems must be capable of processing all varieties of data from text to graph data.

In order to effectively process and analyze Big health data streams, techniques for Big Data must be taken into account and adapted for data streams. Big Data in health informatics is a pool of technologies, tools, and techniques that manage, manipulate, and organize enormous, varied, and intricate datasets in order to improve the quality of patients' status. Fang et al. in [24] have given a comprehensive overview of challenges and techniques for big computational health informatics, both historical and state-of-the art. Several techniques and algorithms in machine learning were characterized and compared. Identified challenges were summarized into four categories, depending on Big Data characteristics they tackle (i.e., volume, velocity, variety, and veracity). Furthermore, a general pipeline for healthcare data processing was proposed. This pipeline includes data capturing, storing, sharing, analyzing, searching, and decision support.

Wang et al. in [81] have presented strategic implications of Big Data in healthcare, including but not limited to historical development, architectural design and component functionalities of Big Data analytics. Different capabilities of healthcare data analytics were recognized from different implementation cases, and various benefits were identified. Recognized capabilities include:

- Analytical capability for patterns of care - analytical techniques typically used in a Big Data analytics system to process data with an immense volume, variety, and velocity via unique data storage, management, analysis, and visualization technologies.
- Unstructured data analytical capability - unstructured and semi-structured data in healthcare refer to information that can neither be stored in a traditional relational database nor fit into predefined data models.
- Decision support capability - the ability to produce reports about daily healthcare services to aid managers' decisions and actions.
- Predictive capability - ability to build and assess a model aimed at generating accurate predictions of new observations, where new can be interpreted temporally and or cross-sectionally.
- Traceability - ability to track output data from all the system's IT components throughout the organization's service units.

As a result, the best practice for Big Data analytics architecture has been proposed. This architecture loosely consists of five main architectural layers: data, data aggregation, analytics, information exploration, and data governance. Ma et al. in [51] have underlined the importance of data fusion based approach to provide more valuable personal and group services, such as personalized health guidance and public health warnings. Various technologies and areas that are

related to this topic were covered, including mobile, wearable and cloud computing, Big Data, IoT and Cyber Physical systems. Authors proposed cloud- and big-data-assisted layered architecture, and described each layer along with tools and technologies that can be used. Also, different applications were identified related to medical recommendations and wearable healthcare systems.

4.4 High Performance Data Analytics

Efficient analytics of large amounts of data in health care (such as, patient genomic data or X-ray images) demands computational, communication and memory resources of High Performance Computing (HPC) architectures.

HPC involves the use of many interconnected processing elements to reduce the time to solution of given a problem. Many powerful HPC systems are heterogeneous, in the sense that they combine general-purpose CPUs with *accelerators* such as, Graphics Processing Units (GPUs), or Field Programmable Gates Arrays (FPGAs). For instance ORNL's Summit—which is currently credited as the leading HPC system in the TOP500 list of most powerful computer systems [76]—is equipped with 4608 computing nodes and each node comprises two IBM Power9 22-core CPUs and six NVIDIA V100 GPUs, for a theoretical peak power of approximately 200 petaflops.

Many approaches have proposed for using HPC systems [18, 39, 56, 60]. While multi-core CPUs are suitable for general-purpose tasks, many-core processors (such as the Intel Xeon Phi [20] or GPU [58]) comprise a larger number of lower frequency cores and perform well on scalable applications [54] (such as, DNA sequence analysis [55] or deep learning [79]).

5 Medical Data Analytics: Research Challenges and Opportunities

5.1 Adaptive Edge/Cloud Processing and Analytics

In the context of medical IoT, the particular benefits could arise from the fog and edge computing paradigms. They represent the model where sensitive data generated by body worn medical devices and smart phone sensors are processed, analyzed and mined close to where it is generated, on these devices themselves, instead of sending vast amounts of sensor data to the cloud, that could exhaust network, processing or storage resources and violate user privacy. Only aggregated, context enriched data/events are sent to the cloud for further processing and analytics. The right balance between Edge-Cloud should provide fast response to detected events, as well as conservation of network bandwidth, minimization of latency, security and preservation of privacy.

Typically, Internet of Things (IoT) is composed of small, resource constrained devices that are connected to the Internet. In essence, they are dedicated to perform specific tasks without the need of providing a general means for performing complicated, resource-consuming computations. This means that most

of the data is transmitted to the cloud without preprocessing or analysis. This implies that the amount of data that is transmitted to the cloud is increasing even more rapidly than the number of IoT devices itself. Indeed, cloud computing is being recognized as a success factor for IoT, providing ubiquity, reliability, high-performance and scalability. However, IoT solutions based on cloud computing fail in applications that require very low and predictable latency or are implemented in a geographically distributed manner with a lot of wireless links. A promising technology to tackle the low-latency and geographical distribution required by IoT devices is fog computing.

The fog computing layer is an intermediate layer between the edge of the network and the cloud layer. The fog computing layer extends the computation paradigm geographically providing local computing and storage for local services. Fog computing does not outsource cloud computing. It aims to provide a computing and storage platform physically closer to the end nodes provisioning new breed of applications and services with an efficient interplay with the cloud layer. The expected benefit is a better quality of service for applications that require low latency. Lower latency is obtained by performing data analysis already at the fog computing layer. Data analysis at the fog computing layer is lightweight, and therefore more advanced analyses and processing should be done at the cloud layer. Naturally, some applications do not require real-time computation, or they need high processing power, and therefore they are performed at the cloud layer. For example, in the case of a smart community, where homes in a neighborhood are connected to provide community services, low latency is expected for making urgent decisions, and thus computation is performed within the neighborhood, instead of a cloud layer which can be located on another continent [33].

5.2 Virtualisation Technologies in e-Health

To support multi-tenancy of different applications and to achieve elasticity in large-scale shared resources, fog computing takes advantages of virtualization technologies. Virtualization and application partitioning techniques are the two key technology solutions that are employed in a fog computing platform. Virtualization includes the process of abstracting and slicing the heterogeneous computing nodes into virtual machines (VMs), hiding the details about the heterogeneity of hardware devices from the applications that are running on the fog layer. Among different virtualization technologies, Linux containers, i.e., containerization, has advantages in short implementation time, efficient resource utilization and low management cost. Application partitioning divides each task into smaller sub-tasks to be executed concurrently on distributed grids, which can improve the task offloading efficiency of applications when the node is overloaded by computationally intensive applications.

Containers provide OS level virtualization without a need for deployment of a virtual OS. Hence, they are lightweight and significantly smaller in size than VMs. Containers provide a self-contained and isolated computing environment for applications and facilitate lightweight portability and interoperability for IoT

applications. Moreover, data and resource isolation in containers offers improved security for the applications running in fog nodes [17].

Recently, researchers have shown increased interest in studying and deploying container-based healthcare applications. Kačeniauskas et al. [37] have developed a software to simulate blood flows through aortic valves, as a cloud service. They have experimentally evaluated the performance of XEN VMs, KVM VMs and Docker containers and have found that Docker containers outperform KVM and XEN VMs. In [35] a framework to monitor patients symptoms has been proposed. This solution employs a Raspberry Pi that read the medical data through the sensors attached and sent it to a server. The server is running Docker containers.

A home-based healthcare framework has been proposed by Li et al. [47]. Different sensors send streams of medical data to docker containers for processing. Koliogeorgi et al. [42] has presented the cloud infrastructure of the AEGLE project, which integrates cloud technologies and heterogeneous reconfigurable computing in large scale healthcare system for Big Bio-Data analytics. AEGLE runs each service and library in different Docker containers. It also uses Kubernetes for fault tolerant, management and scaling of the containerized applications.

A conceptual model and minimal viable product implementation, that enables the analyses of genomic data, harvested by a portable genome sequencer using mobile devices, has been proposed by Araújo et al. [14]. They have used Docker to process the data and to ensure that their system will work across different devices and operating systems. Moreover they would like to prepare the Docker image available for mobile phones. In this way, it will be possible to run their system containers on mobile phones. Rassias et al. [66] has introduced a web-based platform for real-time teleconsultation services on medical imaging. Their platform consists of a client application and an ecosystem of microservices. Each system's service is running in Docker container.

Bahrami et al. [15] has proposed a method to better protect sensitive data. Their method provides real-time resource recommendation for a container scheduler according to HIPAA's (Health Insurance Portability and Accountability Act) regulations. Finally, a recent paper by Sahoo et al. [70] presents a fault tolerant scheduling framework for distribution of healthcare jobs based on their types. They have evaluated the framework with both hypervisor-based virtualization and container-based virtualization. Containers achieved faster start-up time and less response time for both CPU-intensive and data-intensive task.

6 Case Studies and Practical Solutions

6.1 Big Data Solution for Remote Health Monitoring

Taking into account all the challenges related to health/activities monitoring data and control flows that should be employed in a public, private or hybrid cloud, we have designed Remote Health Monitoring system architecture and have implemented the system tailored to analysis of mobility data for the purposes

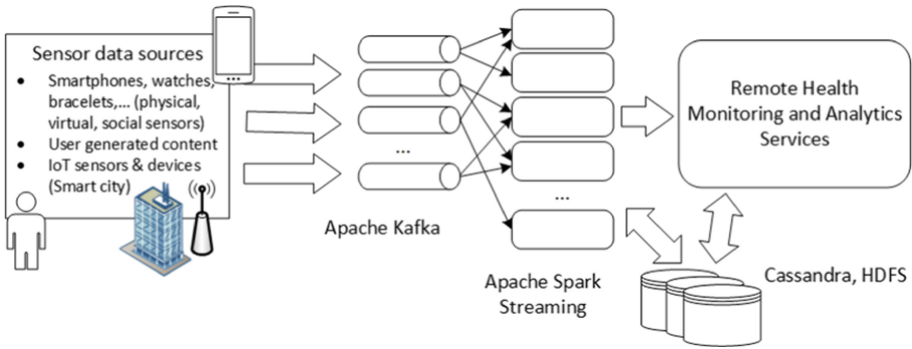


Fig. 3. The general architecture of remote health monitoring cloud infrastructure

of healthcare. The general architecture of Remote Health Monitoring system is given in Fig. 3.

To provide collection of massive and fast sensor data streams, a distributed message broker is employed that maintain a set of topics, for different users or medical conditions, where different sensor data and events will be streamed. These topics would be monitored by number of processing and analytics jobs within Big Data stream engine. Each job acquires only the data from the topics of interest, so different jobs are used to identify specific medical condition and patient behaviour. This architecture could be applied to both personalized and group healthcare systems. The main advantages are:

- Data decoupling: every measurement is streamed as a separate record to a predefined topic, so occurrence of missing data in records is easily detectable: they are absent from the topic; if data from certain sensors is missing, it can be easily left out.
- The separation at jobs level: jobs can monitor only a subset of topics; provides a perfect ground for implementation of different privacy protocols for each job.
- The user level separation: this model allows both personal and collective analysis as jobs can monitor data on individual, or a group level.
- Comprehensive visualization: the results from each jobs can be visualized separately, which can simplify the visualization itself; however, alerts and notifications from different jobs can be combined and visualized to provide an insight to general state.

Big health data stream processing and analysis currently relies on systems for massive stream analytics. A comparative analysis of the existing state-of-the art stream processing solutions is presented in [72], including both open source solutions such as Storm, Spark Streaming and S4, and commercial streaming solutions such as Amazon Kinesis and IBM Infosphere Stream. Comparison includes processing model and latency, data pipeline, fault tolerance and data guarantees. These and similar solutions can be used in healthcare use cases.

For the implementation of prototype system Apache Kafka¹¹ and Apache Spark Streaming¹² have been used. Kafka is chosen as it can support different topics, and messaging capabilities necessary for handling Big Data streams. Apache Spark Streaming is selected because of its main paradigm which implies that stream data is treated as series of bounded batch data, a technique called micro batching, and for its machine learning library MLlib. We do not use a mobile health application for data collection, but we created a simulation of multiple users based on pre-recorded personal health/activities data used for streaming such data to the system. Each users simulation component reads a file that contains different records in csv format, and publishes them to adequate Kafka topic. Actual data have been obtained from UbiqLog [68] and CrowdSignals [82] datasets. These datasets contain information retrieved from smartphones and smart bracelets, such as a heart rate, step count, accelerometer and gyroscope data, location, application usage and WiFi. A distinct record is generated for every measurement or event, and it contains user id, topic, and value. Id is used to separate different users, topic to identify type of measurement or event, and value to hold actual data. For each type of record there is a predefined Kafka topic. These topic streams are monitored by a number of Apache Spark Streaming jobs. Each job acquires the data from monitored topics, and performs specific health and activities related analysis. We implemented a demo job for prediction of skin temperature based on heart rate and step count values. Firstly, in offline mode we created a regression model for these parameters. Then, in Spark Streaming job, we loaded that model, and generated alert in the form of console notification if predicted temperature is higher than the predefined threshold. This illustrates that the system can be used for complex event processing. Our vision is to create jobs that can detect complex events, extrapolate valuable information, and run spatio-temporal mining algorithms on available data. Furthermore, visualization and alert generating component of the system will be implemented.

6.2 Smartphone Ad Hoc Network (SPAN) for e-Health Application

Practical Scenario. In this section, we propose a system designed to support distributed monitoring of trip participants' health. Let's assume that there is a travel agency which organizes the climbing in a high mountains, e.g. Kili-manjaro. As the agency makes reasonable policy, they equip the guide and the participants with sensors that measure health parameters (heart rate, pressure, body temperature, etc.) in order to monitor the people's health.

The guide-leader is equipped with a satellite mobile phone and can exchange information with the agency and rescue teams. The communication between the guide and the participants is organized in ad hoc manner using Bluetooth. Each participant is equipped with a smartphone and the health sensor. The sensor is connected to the smartphone by Bluetooth Low Energy.

¹¹ kafka.apache.org.

¹² spark.apache.org/streaming/.

Technology. One of the trends observed recently in the global telecommunications is a constant growth of the number of smart devices. Mobile devices are getting richer in functionality thanks to more and more computing resources, better batteries and more advanced embedded equipment (e.g. sensors, cameras, localization modules). An important parts of each mobile device are the wireless communication interfaces (e.g. WiFi, Bluetooth), which are used to exchange data between the mobile device and peripheral devices (e.g. headphones, hands-free system, etc.) or other mobile devices. Thanks to the wireless communication it is also possible to integrate smartphones with external sensors.

Thus Sikora et al. [71] decided to create the SmartGroup@Net (SGN) mobile application which allows on direct communication between people without the necessity to be in a range of cellular network infrastructure. The communication is realized over Bluetooth Low Energy (BLE) to create mobile network allowing local exchange of messages, which contain selected information about the state of the network nodes. This is an example of Smart Phone Ad hoc Network (SPAN) solution¹³. In the considered test study the Polar H7 Bluetooth Smart 4.0 heart rate sensor and Wiko Sunny phones are used – see Fig. 4.



Fig. 4. A smartphone with SGN application running together with Polar H7 heart rate sensor.

Mobile Application. SmartGroup@Net (SGN) is the mobile application, which allows on supporting outdoor search actions performed by group of people. The application is useful for different public services, that could be supported during rescue action by civilian volunteers equipped with mobile phones. Other potential applications of SGN include: increasing safety of people moving in

¹³ SPAN network can be formed by portable devices (e.g. smartphones, tablets, etc.) which communicate over WiFi or Bluetooth and operate in ad-hoc mode. The main advantage of the SPAN network is independence not only from the network operator, but also from the tariff plan – the communication in an ad-hoc mode is free of charges.

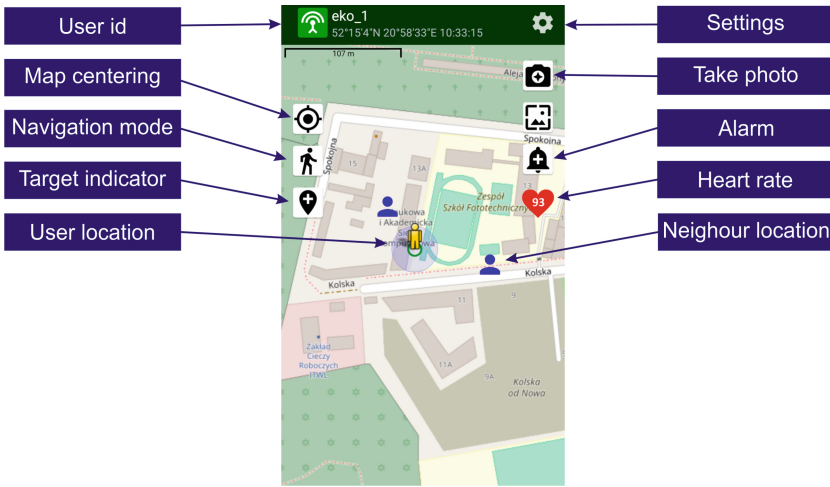


Fig. 5. An example of the SGN application user screen with buttons and fields description

groups in areas with limited access to the mobile network – like high mountains climbing, coordination and management of rescue actions in a crisis situation area (e.g. earthquake, plane crash etc.) and other coordinated actions involving sensors carried by people or self-propelled devices. In all mentioned scenarios it is important to maintain a proper group formation and topology.

An exemplary screen of the SGN application is shown in Fig. 5. Information about the current location and the destination of the mobile phone user as well as the locations of other participants of the action are presented on the map downloaded online from the Open Street Map service¹⁴ and stored in the cache or installed in the offline mode as image file.

The presented application uses the BLE protocol for local exchange of information between the participants of the action. The information sent by each participant include his identification data (action number, user's name and id), location (geographical coordinates), status (normal or emergency), heart rate. These data are saved in the broadcast packet, encrypted and sent to other users depending on the radio transmission range. Each of the participants of the action can simultaneously broadcast messages with their own data and listen to messages from the other group members. Thus the leader (guide) can obtain the information about health status of the participants. Depending on the situation appropriate preventive measures can be taken.

¹⁴ <http://www.openstreetmap.org>.

6.3 Machine Learning for Monitoring Daily Living Activities of People with Dementia

Dementia is a very complex condition affecting the mental health of people and having a negative impact upon their daily life, independence and abilities. The current statistics show that dementia has a rapid spread worldwide as each year around 10 million new cases are reported [7]. Moreover, it is estimated that since 2016, dementia has become the 5th cause of deaths worldwide [12]. Even if there is no cure for dementia, the lives of people with dementia could be improved by detecting and treating challenging behavioural and psychological symptoms [7]. In this context, our vision is to leverage on novel Big Data ICT technologies and Machine Learning (ML) algorithms to analyze the daily living activities of the people with dementia aiming to determine anomalies in the pattern of their daily living activities. Such patterns might be analyzed by medical experts or professional caregivers to take actions to improve patients overall health. For example, the anomalies found in the daily living activities of people with dementia could help medical experts in differentiating between health decline symptoms introduced by dementia and health decline symptoms generated by the side effects of the prescribed medications and their interactions (i.e. polypharmacy side-effects). This would allow medical experts to better evaluate the health status of a patient and better adjust its medication plan.

This section proposes a system designed based on the Lambda architecture for the real-time analysis of daily living activities of people with dementia using machine learning algorithms. The system collects data about the daily living activities patterns of the monitored persons using a set of Raspberry Pi's that are installed in the house of the monitored person. The collected data is sent to a Big Data platform where machine learning algorithms are applied to identify anomalies in the behavior of people with dementia. The proposed system has been implemented in the context of the MEDGUIDE project [10] which is an innovative Ambient Assisted Living (AAL) project that aims to provide solutions for supporting the well-being of people with dementia and their caregivers.

Challenges of ML for Heterogeneous Data Streams. ML for data streams is a subclass of the machine learning domain in which the analyzed data is generated in real-time. In [46] several challenges that might appear when ML is applied on heterogeneous distributed data streams are identified: (a) the input data is generated by many sources, (b) the data formats are both structured and unstructured, (c) the streaming data has a high speed, (d) the data is not completely available for processing (in some cases), (e) the data might be noisy, characterized by missing values and of poor quality. Moreover several ML algorithms must be adapted to larger datasets because they were designed for smaller sets of data and the samples usually have many features because the data is monitored by a big number of sensors. According to [64] several critical issues that must be considered when analyzing real time data streams are: data incompleteness, data streams heterogeneity, high speed of data streams and data large scale.

In [78] two major approaches for data streams processing are presented. In the first approach the in-memory persists only a small part of the elements which characterize the entire data stream while in the second approach fixed length windows are used. In the second approach a selection of fixed parts of data are loaded temporarily into a memory buffer and the ML algorithms take as input the data from that memory buffer.

The authors of [30] identify another major challenge that might exist when the ML algorithms are applied on data streams: in many cases the speed of the underlying ML algorithm is much slower than the speed of the coming data. A possible solution for this challenge is to skip some of the instances from the data stream. However, in that case the outcome might be negative if the instances that are skipped contain insightful information. In [80] two steps that are typically applied for predictive analytics and for prescriptive analytics are presented. The first step is the analytic model building – in which the analytic model is generated by trying out different approaches iteratively. The second step is dedicated to the analytic model validation – when the model is tested on different datasets and improved either by changing the configurable parameters or the complete algorithm.

The Experimental Platform. Our system design is based on Lambda architecture and that can be used for the real-time analysis of the daily living activities of the people that have dementia using different ML algorithms.

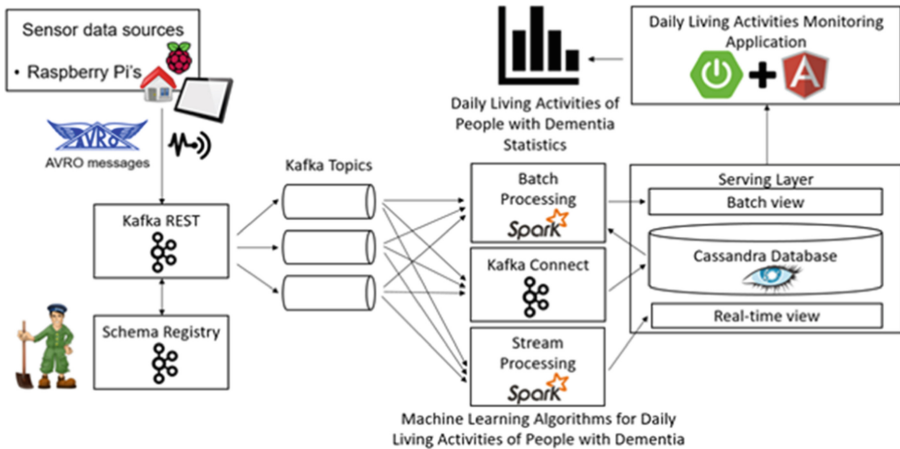


Fig. 6. System for analyzing daily living activities of people with dementia using ML algorithms

Figure 6 presents the high-level architecture of a system that is used for the analysis of the daily living activities of people with dementia. The system applies ML algorithms to detect anomalies in the daily living activities patterns of the monitored patients that might require special attention from a specialized healthcare professional.

The system collects data about the daily living activities patterns of the monitored persons using a set of Raspberry Pi's that are installed in the house of the monitored person. The Raspberry Pi's collect information about different types of activities such as sleeping, feeding, hygiene, indoor activities and outdoor activities. The number of times that the elder goes to the toilet and the wandering behavior during night can be used as indicatives of different stages of dementia. The caregiver being notified that the monitored patient presents anomalous behavior with respect to the regular pattern of activities may take preventive actions as to minimize the effects of dementia.

The data collected by the Raspberry Pi's is sent in the form of AVRO messages to a Big Data platform created using the following technologies: Apache Zookeeper [5], Apache Kafka [3], Apache Spark [4] and Apache Cassandra [2]. Zookeeper is a server that manages the coordination of tasks for the nodes of the distributed system that is used for the processing of the streaming data. Kafka is a streaming platform that handles the real-time data. The data comes in the form of AVRO messages and is inserted in the system through the Kafka REST server. The Kafka REST server uses an AVRO schema that is persisted in the Schema Registry server to get the information from the messages. The AVRO schema imposes a set of restrictions on the messages such as the number of fields and the data types of the fields. The data is then processed in an implementation of a Lambda architecture.

Kafka Connect is a server that takes the messages from the Kafka topics and inserts them in Cassandra. In the case of batch processing the entire data stored in the database is used as training data for the classification algorithms that are used for the prediction of the daily living activities of the people with dementia, while in the case of stream processing only the last streaming information is considered. The function in the case of Lambda architecture is represented by the ML algorithm that is used for the prediction of the daily living activities. In the case of stream processing, it is unfeasible to retrain the ML algorithm each time new streaming data comes into the system because this process would take too much time compared to the time in which the prediction should be made. Considering this time restriction, depending on the size of the streaming data, the ML algorithms should be retrained at a frequency of several minutes or several hours.

The predicted data can then be sent to the caregivers or to the healthcare professionals using a healthcare application created with Spring and Angular. Spring is used for back-end and for the communication with Apache Cassandra and also for retrieving the data generated by the monitoring sensors and the results obtained after the application of the ML algorithms. Spring is also used for sending notifications to the caregivers if corrective measures must be taken. Angular is a JavaScript framework that is used for the creation of the front-end side. By using the healthcare system the caregivers should be able to see data analytics, to detect anomalies in the daily living activities of the monitored persons and should also be able to take immediate actions when they are notified.

Detecting the Side-Effects of Polypharmacy upon the Daily Life Activities of People with Dementia Using Machine Learning Algorithms.

This section presents how we have used the previously described experimental platform in the context of detecting the deviations in the daily life activities, and their polypharmacy-related causes, of people with dementia using ML algorithms. We have chosen to focus on polypharmacy as it represents one of the main challenges in the treatment of dementia affecting over 21% of the elders with dementia. In particular, the Random Forest Classification algorithm has been used for detecting the deviations of the daily life activities from the patients' baseline routine, while the k-Means clustering algorithm has been used for correlating these deviations with the side-effects of drug-drug interactions. The two ML algorithms have been implemented using two Spark jobs.

Concepts Definitions. In our approach, a day containing monitored activities is defined as:

$$day = (duration(a_1), duration(a_2), \dots, duration(a_n)) \quad (1)$$

where $duration(a_i)$ represents the total duration in hours of the activity a_i while n is the number of total activities considered. In our approach, we have considered the following five types of daily life activities as significant enough for allowing the detection of polypharmacy side effects: $a_1 =$ sleeping, $a_2 =$ feeding, $a_3 =$ toilet hygiene, $a_4 =$ functional mobility and $a_5 =$ community mobility. Table 1 presents an example with information provided by sensors regarding the activities performed by a patient with dementia during January 2nd, 2018.

Table 1. Information provided from sensors regarding the activities performed by a patient with dementia during a day

Start time	End time	Activity
2018.01.02 00:00:00	2018.01.02 7:00:00	Sleeping
2018.01.02 7:00:00	2018.01.02 7:30:00	Toilet hygiene
2018.01.02 7:30:00	2018.01.02 8:00:00	Feeding
2018.01.02 8:00:00	2018.01.02 12:00:00	Functional mobility
2018.01.02 12:00:00	2018.01.02 13:30:00	Sleeping
2018.01.02 13:30:00	2018.01.02 14:00:00	Feeding
2018.01.02 14:00:00	2018.01.02 16:00:00	Community mobility
2018.01.02 16:00:00	2018.01.02 18:00:00	Functional mobility
2018.01.02 18:00:00	2018.01.02 18:30:00	Feeding
2018.01.02 18:30:00	2018.01.02 19:00:00	Toilet hygiene
2018.01.02 19:00:00	2018.01.02 21:00:00	Functional mobility
2018.01.02 21:00:00	2018.01.02 23:59:59	Sleeping

The representation of the day January 2nd, 2018 according to (1) is $\text{day} = (10.5, 1.5, 1, 8, 2)$.

The *baseline* for a patient with dementia is a day when the patient has a typical behavior and it is represented similar to the representation of a day (see (1)). In our approach, the baseline for each monitored patient is defined by the doctor based on discussions with the patient, its family and caregivers.

We consider that a monitored day is a *deviated day* if it contains at least one activity which has the total duration higher or lower than a pre-defined threshold compared to the same activity type in the baseline. A deviated day may be associated with a semantic annotation consisting of a drug-drug interaction and an associated side-effect. The information regarding dementia drugs, drug-drug interactions and associated side-effects are stored as instances in a Polypharmacy Management Ontology we have designed. This ontology is a simplified version of the Drug-Drug Interactions Ontology (DINTO) [83] enhanced with drugs side effects taken from [1].

Classifying a Day as Normal or as Having Significant Deviations from the Baseline Using the Random Forest Classification Algorithm. To classify a day as normal or as having significant deviations from the baseline we have used the Random Forest algorithm suitable for Big Data classification provided by Apache Spark [11]. We have trained the Random Forest classification algorithm using a data set consisting of days labeled as having or not having a significant deviation from the normal baseline.

In the testing phase, a set of new monitored days is classified, based on the data set used in the training phase, as having/not having a deviation from the baseline.

Identifying the Causes that Produced the Deviation of a Day from the Baseline. To identify the causes which led to a deviation in the daily life activities of a patient with dementia we have defined a k-Means clustering method consisting of two steps:

1. *Cluster the deviated days.* The Spark MLlib implementation of the K-Means clustering algorithm [11] has been applied on a set of deviated days, each day being annotated with a drug-drug interaction and a side-effect causing the deviation. Each cluster will contain similar annotated days and will be labeled with the annotation (i.e. drug-drug interaction and its adverse effect) of the cluster's centroid.
2. *Assign a new deviated day to a cluster resulted in the previous step.* A new deviated day will be assigned to the cluster for which the Euclidean distance between that day and the cluster's centroid is minimum. Consequently, the deviated day will be annotated with the drug-drug interaction and its adverse effect corresponding to the assigned cluster's centroid.

Figure 7 presents the resources and the data flows for detecting the side-effects of polypharmacy upon the daily life activities of people with dementia using the Random Forrest Classifier and the k-Means Clustering Method.

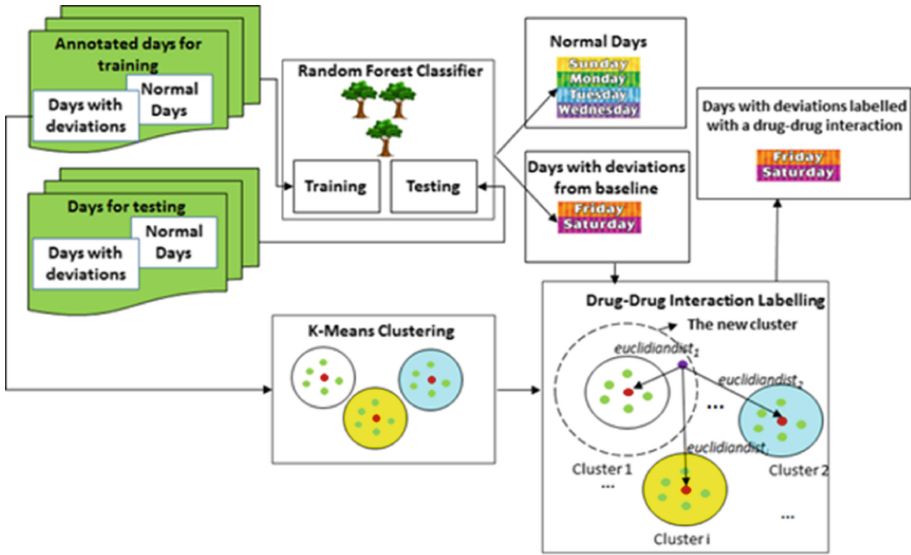


Fig. 7. The resources and the data flows for detecting the side-effects of polypharmacy upon the daily life activities of people with dementia

7 Summary

Health informatics is already an established scientific field and advances of e-health and m-health are already part of clinical practices of the XXI century medicine. Digital medicine enables optimization of decision-making processes and precise/personal medicine through the possibility of analyzing a huge amount of data at low cost. The computer assisted decision-making tools may be more efficient and safer than “analog” approaches that involve a physician.

In this chapter we have described (1) data collection, fusion, ownership and privacy issues; (2) models, technologies and solutions for medical data processing and analysis; (3) big medical data analytics for remote health monitoring; (4) research challenges and opportunities in medical data analytics; (5) three case studies. The first case study described a Big Data solution for remote health monitoring that considers challenges pertaining to monitoring the health and activities of patients. The second case study described a system designed to support distributed monitoring of human health during a trip. The third case study described the use of machine learning in the context of monitoring daily living activities of people with dementia.

References

1. <https://www.drugs.com/>
2. Apache Cassandra. <http://cassandra.apache.org/>
3. Apache Kafka: A distributed streaming platform. <https://kafka.apache.org/>

4. Apache Spark: Lightning-fast cluster computing. <http://spark.apache.org/>
5. Apache Zookeeper. <https://zookeeper.apache.org/>
6. CardioWheel. <https://www.cardio-id.com/cardiowheel>
7. Dementia. <https://www.who.int/en/news-room/fact-sheets/detail/dementia>
8. FIWARE. <https://www.fiware.org/>
9. FIWARE Accelerator Programme: FICHe project. <https://www.f6s.com/fiche>
10. ICT integrated system for coordinated polypharmacy management in elders with dementia. <https://dsrl.coned.utcluj.ro/medguide/>
11. Machine learning library (MLlib) guide. <https://spark.apache.org/docs/latest/ml-guide.html>
12. The top 10 causes of death. <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>
13. Akbar, A., et al.: Real-time probabilistic data fusion for large-scale IoT applications. *IEEE Access* **6**, 10015–10027 (2018). <https://doi.org/10.1109/ACCESS.2018.2804623>
14. Araújo, R., Kim, A., Crisan, A.: EpiMobile: pathogen point of care diagnosis and global surveillance using mobile devices (2017)
15. Bahrami, M., Malvankar, A., Budhraj, K.K., Kundu, C., Singhal, M., Kundu, A.: Compliance-aware provisioning of containers on cloud. In: 2017 IEEE 10th International Conference on Cloud Computing (CLOUD), pp. 696–700. IEEE (2017)
16. Baker, S.B., Xiang, W., Atkinson, I.: Internet of Things for smart healthcare: technologies, challenges, and opportunities. *IEEE Access* **5**, 26521–26544 (2017). <https://doi.org/10.1109/ACCESS.2017.2775180>
17. Bellavista, P., Zanni, A.: Feasibility of fog computing deployment based on Docker containerization over RaspberryPi. In: Proceedings of the 18th International Conference on Distributed Computing and Networking, ICDCN 2017, pp. 16:1–16:10. ACM, New York (2017)
18. Benkner, S., et al.: PEPPHER: efficient and productive usage of hybrid computing systems. *IEEE Micro* **31**(5), 28–41 (2011). <https://doi.org/10.1109/MM.2011.67>
19. Celesti, A., et al.: How to develop IoT cloud e-health systems based on FIWARE: a lesson learnt. *J. Sens. Actuator Netw.* **8**(1), 7 (2019)
20. Chrysos, G.: Intel® Xeon Phi Coprocessor-the Architecture. Intel Whitepaper (2014)
21. Cortés, R., Bonnaire, X., Marin, O., Sens, P.: Stream processing of healthcare sensor data: studying user traces to identify challenges from a big data perspective. *Procedia Comput. Sci.* **52**, 1004–1009 (2015)
22. Dubey, H., et al.: Fog computing in medical Internet-of-Things: architecture, implementation, and applications. In: Khan, S.U., Zomaya, A.Y., Abbas, A. (eds.) *Handbook of Large-Scale Distributed Computing in Smart Healthcare*. SCC, pp. 281–321. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58280-1_11
23. Esteban, A.P., Patricio, M.A., García, J., Molina, J.M.: A feature selection approach to the group behavior recognition issue using static context information. *IJDSN* **9** (2013). <https://doi.org/10.1155/2013/383906>
24. Fang, R., Pouyanfar, S., Yang, Y., Chen, S.C., Iyengar, S.S.: Computational health informatics in the big data age: a survey. *ACM Comput. Surv.* **49**(1), 12:1–12:36 (2016). <https://doi.org/10.1145/2932707>. <http://doi.acm.org/10.1145/2932707>
25. Fisher, D.K., Fletcher, R.S., Anapalli, S.S., Pringle III, H.: Development of an open-source cloud-connected sensor-monitoring platform. *Adv. Internet Things* **8**(01), 1 (2017)

26. Garcia, A.C.B., Vivacqua, A.S., Pi, N.S., Martí, L., Molina, J.M.: Crowd-based ambient assisted living to monitor the elderly's health outdoors. *IEEE Softw.* **34**(6), 53–57 (2017). <https://doi.org/10.1109/MS.2017.4121217>
27. Gil, G.B., de Jesús, A.B., Molina, J.M.: InContexto: multisensor architecture to obtain people context from smartphones. *IJDSN* **8** (2012). <https://doi.org/10.1155/2012/758789>
28. Gómez-Romero, J., Serrano, M.A., García, J., Molina, J.M., Rogova, G.L.: Context-based multi-level information fusion for harbor surveillance. *Inf. Fusion* **21**, 173–186 (2015). <https://doi.org/10.1016/j.inffus.2014.01.011>
29. Gómez-Romero, J., Serrano, M.A., Patricio, M.A., García, J., Molina, J.M.: Context-based scene recognition from visual data in smarthomes: an information fusion approach. *Pers. Ubiquit. Comput.* **16**(7), 835–857 (2012). <https://doi.org/10.1007/s00779-011-0450-9>
30. Hapfelmeier, A., Mertes, C., Schmidt, J., Kramer, S.: Towards real-time machine learning. In: *Proceedings of the ECML PKDD 2012 Workshop on Instant Interactive Data Mining* (2012)
31. Hassanalieragh, M., et al.: Health monitoring and management using Internet-of-Things (IoT) sensing with cloud-based processing: opportunities and challenges. In: *2015 IEEE International Conference on Services Computing*, pp. 285–292, June 2015. <https://doi.org/10.1109/SCC.2015.47>
32. Herrero, J.G., Patricio, M.A., Molina, J.M., Cardoso, L.A.: Contextual and human factors in information fusion. In: *Human Systems Integration to Enhance Maritime Domain Awareness for Port/Harbour Security*, pp. 79–92 (2010). <https://doi.org/10.3233/978-1-60750-621-8-79>
33. Hosseinpour, F., Meng, Y., Westerlund, T., Plosila, J., Liu, R., Tenhunen, H.: A review on fog computing systems. *Int. J. Adv. Comput. Technol.* **8**, 48–61 (2016)
34. Islam, S.M.R., Kwak, D., Kabir, M.H., Hossain, M., Kwak, K.: The Internet of Things for health care: a comprehensive survey. *IEEE Access* **3**, 678–708 (2015). <https://doi.org/10.1109/ACCESS.2015.2437951>
35. Jaiswal, K., Sobhanayak, S., Mohanta, B.K., Jena, D.: IoT-cloud based framework for patient's data collection in smart healthcare system using Raspberry-Pi. In: *2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, pp. 1–4, November 2017
36. Jarynowski, A., Marchewka, D., Buda, A.: Internet-assisted risk assessment of infectious diseases in women sexual and reproductive health. *E-methodology*, pp. 135–144 (2017)
37. Kačeniauskas, A., Pacevič, R., Starikovičius, V., Maknickas, A., Staškūnienė, M., Davidavičius, G.: Development of cloud services for patient-specific simulations of blood flows through aortic valves. *Adv. Eng. Softw.* **103**, 57–64 (2017)
38. Kamar, I., Chatterjee, P., Hamie, A.: Internet of Things in learning systems—a perspective of platforms. *Int. J. Adv. Res. Comput. Sci.* **7**(2), 52–59 (2016)
39. Kessler, C., et al.: Programmability and performance portability aspects of heterogeneous multi-/manycore systems. In: *Design, Automation and Test in Europe Conference and Exhibition (DATE)*, pp. 1403–1408. IEEE (2012)
40. Khaleghi, B., Khamis, A.M., Karray, F., Razavi, S.N.: Multisensor data fusion: a review of the state-of-the-art. *Inf. Fusion* **14**(1), 28–44 (2013). <https://doi.org/10.1016/j.inffus.2011.08.001>
41. Khan, S.U., Zomaya, A.Y., Abbas, A. (eds.): *Handbook of Large-Scale Distributed Computing in Smart Healthcare*. SCC. Springer, Cham (2017). <https://doi.org/10.1007/978-3-319-58280-1>

42. Koliogeorgi, K., et al.: AEGLE's cloud infrastructure for resource monitoring and containerized accelerated analytics. In: 2017 IEEE Computer Society Annual Symposium on VLSI (ISVLSI), pp. 362–367. IEEE (2017)
43. La Ode Hasnuddin, S.S., Abidin, M.S.: Internet of Things for early detection of landslides. In: Prosiding Seminar Nasional Riset Kuantitatif Terapan 2017, vol. 1 (2018)
44. Lane, N.D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., Campbell, A.T.: A survey of mobile phone sensing. *IEEE Commun. Mag.* **48**(9), 140–150 (2010). <https://doi.org/10.1109/MCOM.2010.5560598>
45. Lee, C., et al.: Big healthcare data analytics: challenges and applications. In: Khan, S.U., Zomaya, A.Y., Abbas, A. (eds.) *Handbook of Large-Scale Distributed Computing in Smart Healthcare*. SCC, pp. 11–41. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58280-1_2
46. L'Heureux, A., Grolinger, K., Elyamany, H.F., Capretz, M.A.M.: Machine learning with big data: challenges and approaches. *IEEE Access* **5**, 7776–7797 (2017). <https://doi.org/10.1109/ACCESS.2017.2696365>. <https://ieeexplore.ieee.org/document/7906512>
47. Li, P., Xu, C., Luo, Y., Cao, Y., Mathew, J., Ma, Y.: CareNet: building a secure software-defined infrastructure for home-based healthcare. In: *Proceedings of the ACM International Workshop on Security in Software Defined Networks & #38; Network Function Virtualization, SDN-NFVSec 2017*, pp. 69–72. ACM, New York (2017)
48. Likotiko, E., Petrov, D., Mwangoka, J., Hilleringmann, U.: Real time solid waste monitoring using cloud and sensors technologies. *Online J. Sci. Technol.* **8**(1), 106–116 (2018)
49. Llinas, J.: Information fusion process design issues for hard and soft information: developing an initial prototype. In: Yager, R.R., Reformat, M.Z., Alajlan, N. (eds.) *Intelligent Methods for Cyber Warfare*. SCI, vol. 563, pp. 129–149. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-08624-8_6
50. López, J.M.M., Herrero, J.G., Rodríguez, F.J.J., Corredera, J.R.C.: Cooperative management of a net of intelligent surveillance agent sensor. *Int. J. Intell. Syst.* **18**(3), 279–307 (2003). <https://doi.org/10.1002/int.10089>
51. Ma, X., Wang, Z., Zhou, S., Wen, H., Zhang, Y.: Intelligent healthcare systems assisted by data analytics and mobile computing. *Wirel. Commun. Mob. Comput.* **2018**, 16 (2018). Article ID 3928080
52. Ma, Y., Wang, Y., Yang, J., Miao, Y., Li, W.: Big health application system based on health Internet of Things and big data. *IEEE Access* **5**, 7885–7897 (2017). <https://doi.org/10.1109/ACCESS.2016.2638449>
53. Marjani, M., et al.: Big IoT data analytics: architecture, opportunities, and open research challenges. *IEEE Access* **5**, 5247–5261 (2017). <https://doi.org/10.1109/ACCESS.2017.2689040>
54. Memeti, S., Li, L., Pllana, S., Kolodziej, J., Kessler, C.: Benchmarking OpenCL, OpenACC, OpenMP, and CUDA: programming productivity, performance, and energy consumption. In: *Proceedings of the 2017 Workshop on Adaptive Resource Management and Scheduling for Cloud Computing, ARMS-CC 2017*, pp. 1–6. ACM, New York (2017). <https://doi.org/10.1145/3110355.3110356>
55. Memeti, S., Pllana, S.: Accelerating DNA sequence analysis using Intel® Xeon Phi™. In: 2015 IEEE Trustcom/BigDataSE/ISPA, vol. 3, pp. 222–227, August 2015

56. Memeti, S., Pllana, S.: HSTREAM: a directive-based language extension for heterogeneous stream computing. In: 2018 IEEE International Conference on Computational Science and Engineering (CSE), pp. 138–145, October 2018. <https://doi.org/10.1109/CSE.2018.00026>
57. Mora, H.M., Gil, D., Terol, R.M., López, J.A., Szymanski, J.: An IoT-based computational framework for healthcare monitoring in mobile environments. *Sensors* **17**, 2302 (2017)
58. NVIDIA: What is GPU-Accelerated Computing? April 2017. <http://www.nvidia.com/object/what-is-gpu-computing.html>. Accessed 03 Nov 2018
59. Padilla, W.R., García, J., Molina, J.M.: Information fusion and machine learning in spatial prediction for local agricultural markets. In: Bajo, J., et al. (eds.) PAAMS 2018. CCIS, vol. 887, pp. 235–246. Springer, Cham (2018). <https://doi.org/10.1007/978-3-319-94779-2.21>
60. Padua, D. (ed.): *Encyclopedia of Parallel Computing*. Springer, Boston (2011). <https://doi.org/10.1007/978-0-387-09766-4>
61. Perez, D., Memeti, S., Pllana, S.: A simulation study of a smart living IoT solution for remote elderly care. In: 2018 Third International Conference on Fog and Mobile Edge Computing (FMEC), pp. 227–232, April 2018. <https://doi.org/10.1109/FMEC.2018.8364069>
62. Sanchez-Pi, N., Martí, L., Bustamante, Á.L., Molina, J.M.: How machine learning could detect anomalies on thinger.io platform? In: Bajo, J., et al. (eds.) PAAMS 2018. CCIS, vol. 887, pp. 259–269. Springer, Cham (2018). <https://doi.org/10.1007/978-3-319-94779-2.23>
63. Sanchez-Pi, N., Martí, L., Molina, J.M., Bicharra García, A.C.: Information fusion for improving decision-making in big data applications. In: Pop, F., Kołodziej, J., Di Martino, B. (eds.) *Resource Management for Big Data Platforms*. CCN, pp. 171–188. Springer, Cham (2016). <https://doi.org/10.1007/978-3-319-44881-7.9>
64. Qiu, J., Wu, Q., Ding, G., Xu, Y., Feng, S.: A survey of machine learning for big data processing. *EURASIP J. Adv. Sig. Proces.* (1) (2016). <https://doi.org/10.1186/s13634-016-0355-x>. <https://link.springer.com/article/10.1186/s13634-016-0355-x>
65. Raghupathi, W., Raghupathi, V.: Big data analytics in healthcare: promise and potential. *Health Inf. Sci. Syst.* **2**(1), 3 (2014)
66. Rassias, G., Andrikos, C.O., Tsanakas, P., Maglogiannis, I.: Versatile cloud collaboration services for device-transparent medical imaging teleconsultations. In: 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS), pp. 306–311. IEEE (2017)
67. Rawassizadeh, R., Kotz, D.: Datasets for mobile, wearable and IoT research. *GetMobile: Mob. Comput. Commun.* **20**(4), 5–7 (2017). <https://doi.org/10.1145/3081016.3081018>. <http://doi.acm.org/10.1145/3081016.3081018>
68. Rawassizadeh, R., Momeni, E., Dobbins, C., Mirza-Babaei, P., Rahnamoun, R.: Lesson learned from collecting quantified self information via mobile and wearable devices. *J. Sens. Actuator Netw.* **4**(4), 315–335 (2015). <https://doi.org/10.3390/jsan4040315>
69. Rodgers, M.M., Pai, V.M., Conroy, R.S.: Recent advances in wearable sensors for health monitoring. *IEEE Sens. J.* **15**(6), 3119–3126 (2015). <https://doi.org/10.1109/JSEN.2014.2357257>
70. Sahoo, P.K., Dehury, C.K.: Efficient data and CPU-intensive job scheduling algorithms for healthcare cloud. *Comput. Electr. Eng.* **68**, 119–139 (2018)

71. Sikora, A., Krzyszto, M., Marks, M.: Application of bluetooth low energy protocol for communication in mobile networks. In: 2018 International Conference on Military Communications and Information Systems (ICMCIS), pp. 1–6, May 2018. <https://doi.org/10.1109/ICMCIS.2018.8398689>
72. Singh, M.P., Hoque, M.A., Tarkoma, S.: A survey of systems for massive stream analytics. arXiv preprint [arXiv:1605.09021](https://arxiv.org/abs/1605.09021) (2016)
73. Snidaro, L., García, J., Llinas, J.: Context-based information fusion: a survey and discussion. *Inf. Fusion* **25**, 16–31 (2015). <https://doi.org/10.1016/j.inffus.2015.01.002>
74. Solanas, A., et al.: Smart health: a context-aware health paradigm within smart cities. *IEEE Commun. Mag.* **52**(8), 74–81 (2014). <https://doi.org/10.1109/MCOM.2014.6871673>
75. Sotiriadis, S., Vakanas, L., Petrakis, E., Zampognaro, P., Bessis, N.: Automatic migration and deployment of cloud services for healthcare application development in FIWARE. In: 2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA), pp. 416–419. IEEE (2016)
76. TOP500: The TOP500 List of Most Powerful Computer Systems (2018). <http://www.top500.org/>. Accessed 24 Dec 2018
77. Vaizman, Y., Ellis, K., Lanckriet, G.: Recognizing detailed human context in the wild from smartphones and smartwatches. *IEEE Pervasive Comput.* **16**(4), 62–74 (2017). <https://doi.org/10.1109/MPRV.2017.3971131>
78. Valsamis, A., Tserpes, K., Zissis, D., Anagnostopoulos, D., Varvarigou, T.: Employing traditional machine learning algorithms for big data stream analysis: the case of object trajectory prediction. *J. Syst. Softw.* **127**, 249–257 (2017). <https://doi.org/10.1016/j.jss.2016.06.016>. <https://dl.acm.org/citation.cfm?id=3066671>
79. Viebke, A., Pillana, S.: The potential of the Intel (R) Xeon Phi for supervised deep learning. In: 2015 IEEE 17th International Conference on High Performance Computing and Communications, pp. 758–765, August 2015. <https://doi.org/10.1109/HPCC-CSS-ICCESS.2015.45>
80. Waehner, K.: How to apply machine learning to event processing (2016). <https://www.rtinsights.com/big-data-machine-learning-software-event-processing-analytics/>
81. Wang, Y., Kung, L., Byrd, T.A.: Big data analytics: understanding its capabilities and potential benefits for healthcare organizations. *Technol. Forecast. Soc. Change* **126**, 3–13 (2018). <https://doi.org/10.1016/j.techfore.2015.12.019>. <http://www.sciencedirect.com/science/article/pii/S0040162516000500>
82. Welbourne, E., Tapia, E.M.: CrowdSignals: a call to crowdfund the community’s largest mobile dataset. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, UbiComp 2014 Adjunct, pp. 873–877. ACM, New York (2014). <https://doi.org/10.1145/2638728.2641309>
83. Zazo, M.H., Hastings, J., Segura-Bedmar, I., Croset, S., Martinez, P., Steinbeck, C.: An ontology for drug-drug interactions. In: Proceedings of the 6th International Workshop on Semantic Web Applications and Tools for Life Sciences (2014). http://ceur-ws.org/Vol-1114/Session3_Herrero-Zazo.pdf

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

