IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 11, 2018, accepted July 9, 2018, date of publication July 24, 2018, date of current version August 20, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2859340

A Review on the Role of Nano-Communication in Future Healthcare Systems: A Big Data Analytics Perspective

ALI RIZWAN[1], AHMED ZOHA[1], RUI ZHANG[2], WASIM AHMAD[1], (Member, IEEE), KAMRAN ARSHAD[3], (Senior Member, IEEE), NAJAH ABU ALI[4], AKRAM ALOMAINY[2], (Senior Member, IEEE), MUHAMMAD ALI IMRAN[1], (Senior Member, IEEE), AND QAMMER H. ABBASI[1], (Senior Member, IEEE)

[1]School of Engineering, University of Glasgow, Glasgow G12 8QQ, U.K.
[2]School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K.
[3]Faculty of Engineering, Ajman University, Ajman, United Arab Emirates
[4]Faculty of Information Technology, United Arab Emirates University, Al Ain, United Arab Emirates

Corresponding author: Najah Abu Ali (najah@uaeu.ac.ae)

This work was supported by the ADEC Award for Research Excellence, Abu Dhabi, United Arab Emirates University, under Project AARE17-019.

ABSTRACT This paper presents a first-time review of the open literature focused on the significance of big data generated within nano-sensors and nano-communication networks intended for the future healthcare and biomedical applications. It is aimed toward the development of modern smart healthcare systems enabled with P4, i.e., predictive, preventive, personalized, and participatory capabilities to perform diagnostics, monitoring, and treatment. The analytical capabilities that can be produced from the substantial amount of data gathered in such networks will aid in exploiting the practical intelligence and learning capabilities that could be further integrated with conventional medical and health data leading to more efficient decision making. We have also proposed a big data analytics framework for gathering intelligence, form the healthcare big data, required by futuristic smart healthcare to address relevant problems and exploit possible opportunities in the future applications. Finally, the open challenges, the future directions for researchers in the evolving healthcare domain, are presented.

INDEX TERMS Nano-sensors, nano communication, big data analytics, body-centric communication, smart healthcare.

I. INTRODUCTION

Current healthcare scheme is a reactive approach to address diseases, infections or injuries once they have already occurred or have clear symptoms noticed by patients. Time-lapse in diagnosis and treatment relies on the initiative taken by the victims in visiting a healthcare facility. Such a time lapse is a critical factor in the treatment of the disease. In many cases, just the delayed diagnosis leads to chronic diseases, advanced stages of cancers or even deaths. The severity of the issue can be realized by the fact that chronic diseases account for 59% annual deaths, and 46% of the burden of total diseases worldwide [1]. Whereas, cancer is the second leading cause of death globally as mentioned by world health organization [2]. Late-stage presentation and inaccessibility to in-time diagnosis facilities are common reasons.

An earlier detection and prompt intervention can increase chances of treatment and cure in case of any diseases, even cancer, for example, as highlighted by ministry of health Ontario, mammography screening of 70% of women between the ages of 50 and 69 would have reduced deaths due to breast cancer, by one-third, over a ten-year period [1]. So there is a requirement of proactive rather than a reactive healthcare that can not only detect possible disease, infections or injuries as soon as they happen but, ideally, even before they start. Futuristic healthcare is moving towards that predictive, participatory, preventive and personalized (P4) [3] paradigm with

an aim to provide patient-specific diagnosis and treatment services in a seamless and proactive manner.

Big data analytics and nano-technology independently have emerged as key players for the realization of such a smart healthcare system. There is a continuous progress on both fronts simultaneously. New nano-technology based miscellaneous smart devices are being invented to perform healthcare tasks [4] of monitoring, diagnosis, and treatment in P4 manner. In future, a huge influx of real-time heterogeneous data is expected from the adoption of devices that rely on body-centric nano-networks. On the other end, an overwhelming amount of big data already exists in the healthcare sector [5] primarily from conventional databases comprising electronic medical records. This data can be exploited to gather intelligence required for the provision of omnipresent P4 Healthcare.

Gathering intelligence form such data is a challenging task. Because the data from conventional and non-conventional sources is expected to possess the characteristics like high volume, velocity, veracity, and value associated with big data. Traditional data analytics tools cannot manage this big data to extract practical knowledge out of it. Therefore, innovative, efficient, interoperable, and scalable big data analytics solutions are required to process and analyze such big data from both sources. But the idea of leveraging from conventional medical data using big data analytics tools to address different healthcare challenges is not a new topic [6]. There are also many examples in the literature where applications [7], potential role [8], possible implementation frameworks [9] and steps involved in analysis are discussed thoroughly for big data from conventional sources.

Reason for the popularity of this concept is that the healthcare data from conventional sources itself is enormous in volume, heterogeneous in nature and possess valuable information if harnessed properly. For instance, conventional data for the United States (US) healthcare have already exceeded 150 Exabyte in 2010 [5]. Similarly, healthcare databases of other countries like China and India are expected to cross zettabyte and yottabyte soon [5]. Beside that an increasing complexity is also seen in the data generated. For example, only in the fields of neuroimaging and genetics, petabytes of new data are generated every year and complexity level is increased 8 to 9 folds comparing to the complexity level in 1985 as a benchmark [10]. According to [5] the size and complexity of healthcare, biomedical and social research information almost double every 12-14 months. These all factors make the use of big data analytics tools an optimal choice for knowledge extraction from conventional data.

On the other end, with the advances in the use of nano-technology in smart healthcare devices and facilities, the challenge to be addressed is the exploitation of exponentially growing continuous data generated by millions of nano-sensors [11]. Nano-sensors communicating internally and with central nodes or macro devices on the body, via body area networks which subsequently communicate on the internet with the central healthcare system is a futuristic paradigm (Internet of bio nano-things) [12] for smart healthcare. Currently, such arrangements facilitate basic health monitoring and reporting for offline diagnostics. Besides, that deployment options are limited because of the underlying safety concerns and precautions required.

But the fact is that a continuous progress can be seen in nano-networks and communications [11], [13], [14] and subsequent cited work is promising, that inspires for a networking paradigms that facilitates seamless deployment of nano-sensors in different healthcare contexts; in the environment, on the body, and inside the body [12]. In the future, living beings, including patients and healthy ones, are expected to carry or be surrounded by numerous nano-sensors continuously sensing and transferring information about their health status. Besides that miscellaneous smart healthcare devices like wearables or regular sensors will also be contributing towards the generation of non-conventional continuous data. According to Statista 414.1m users of wearables are expected by 2022 [15], and ABI researchers report, though, today hub devices like smartphones, laptops, tablets, and wearables hold the major share of smart devices connected to internet, but, by the year 2020, 60% of the estimated 30 billion devices are expected to be nodes or sensors [16].

In such scenario, analysis of data, for educated decision making, in real-time can play a vital role, particularly for in-time provision of crucial healthcare services. Big data analytics tools for stream processing can help to address the challenges of processing continuous data streams efficiently. The impact of big data analytics for knowledge extraction and proactive decision making can be further enhanced by using conventional healthcare data in combination with continuous data from smart devices.

The contribution of this paper is that it highlights the significance and need of interdisciplinary research for nano-sensors, nano-communication and big-data analytics in the healthcare sector. It provides a combined review of existing research work from the areas of nano-sensors, nano-communication, and big data from the healthcare perspective. Nano-sensors are classified on the basis of their applications, implementation methods, and architectural layout in healthcare setup. In conjunction, the state-of-the-art nano-communication approaches introduced in the healthcare space are classified, and underlying challenges are elaborated. Besides that, a big data analytics based knowledge extraction framework is proposed to get an insight from conventional and non-conventional data sources including the applications of nano-sensors and nano-communication. Terms like smart healthcare devices and wearables are used frequently throughout this paper, in general, they mean arrangements that rely on nano-sensors, or specified otherwise.

The rest of the paper is organized as follows: Section II presents a discussion on nano-sensors, their applications, and architecture and implementation methods. Section III covers nano-communication and its enabling technologies.

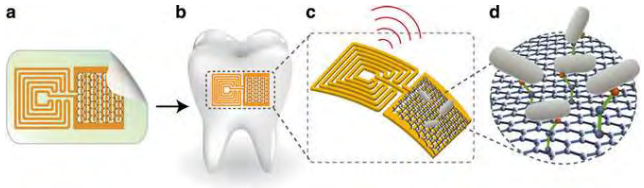

FIGURE 1. Bio-transferrable graphene wireless nano-sensor [18].

Section IV explains a unified big data analytics framework for extracting knowledge from healthcare big data. Some open challenges for the interdisciplinary research are discussed in section V. Section VI concludes the whole discussion on the interdisciplinary study.

II. NANO-SENSORS AND NETWORKS

Nano-sensors are extremely small integrated devices engineered from nanomaterials or biological materials. They are used to detect and respond to a physical property from the environment. Nano-sensors work the same way as conventional sensors, but their size is extremely small - billionths of a meter. They can perform set of simple functions to manipulate signals for detecting, modifying and recording measurements. Nano-sensors come in a variety of sizes and shapes ranging from the size of a macromolecule to that of a bio-cell (i.e., dimensions of 1-100 nm) [17]. In biomedical applications, the size of the nano-sensors used for taking invasive measurements is extremely small compared to the one used to record noninvasive measurements. The application area, measurement site, the end goal, and safety constraints play a critical role in deciding the material and size of a nano-sensor.

In healthcare domain, nano-sensors can be used for different purposes including monitoring, detection, and treatment. For example, nano-sensors can detect chemical compounds in concentrations as low as one part per billion, or the presence of different infectious agents such as virus or harmful bacteria [17]. An example of such a nano-sensor is bio-transferrable graphene wireless nano-sensor [18] illustrated in Fig. 1. The proposed architecture of bio-transferrable nano-sensor has a satisfactory response in sensing extremely sensitive chemicals and biological compounds up to single bacterium. It also has wireless remote powering and readout functionality.

The data measured by nano-sensors needs to be processed and shared with other nano-sensors and nano-devices. Therefore, nano-sensors with thier computation, communication, and action components are miniaturized and fabricated into a single box called nano-machine [19]. Several nano-machines can be connected together through nano-routers that rout measured data to other nano-devices or external devices such as mobile phones [12]. The interconnected cluster of such nano-machines is called a nano-network. An example of such nano-network is given in Fig. 2.

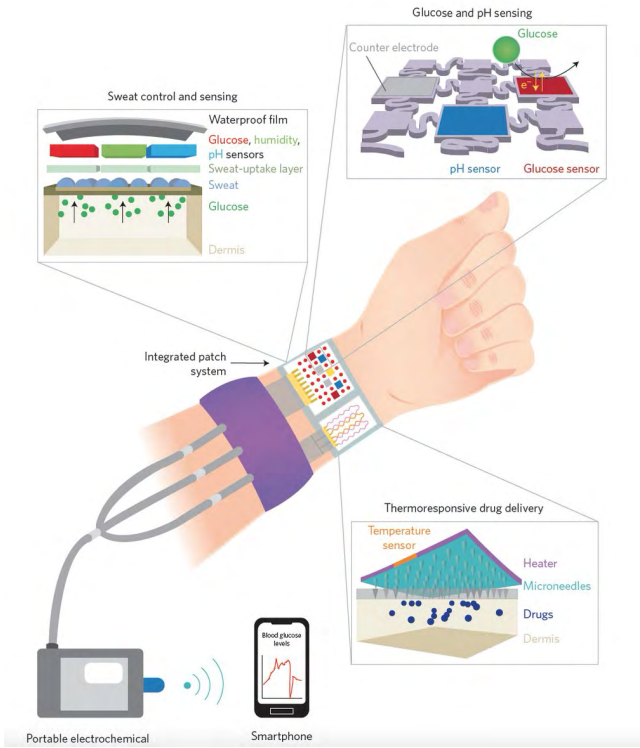

FIGURE 2. Nano-network: glucose graphene skin sweat sensor and drug delivery chip [20].

A. TYPES OF NANO-SENSORS

Nano-sensors are mainly characterized by their material, size, and functionality [21]. They can be physical, chemical, electrical, or magnetic sensors, and can be used to detect minuscule quantities, minute particles, or monitor physical parameters. Some prominent types [17] of nano-sensors are as follow:

Physical Nano-sensors: these types of sensors measure magnitudes of parameters such as pressure, mass, force, or displacement. These nano-sensors usually employ the electronic properties of both nanotubes and nanoribbons that change when they are bent or deformed. A range of physical nano-sensors such as force nano-sensors, displacement nano-sensors, and pressure nano-sensors, are available nowadays [17]. A carbon nanotube (CNT) based physical force nano-sensor is shown in Fig. 3.

Chemical Nano-sensors: these sensors normally measure the gas concentration or the molecular composition of a substance, or they detect specific type of molecules. Chemical nano-sensors operate on the principals that are based on the change in electronic properties of carbon nanotubes (CNTs) and graphene nanoribbons (GNRs), when different kinds of molecules are adsorbed on top of them, it changes the number of electrons that can pass through the carbon lattice [23]. A new type of sensor is proposed in [24], as shown in Fig. 4, which has the potential to replace sniffer dogs when it comes to detecting explosives.

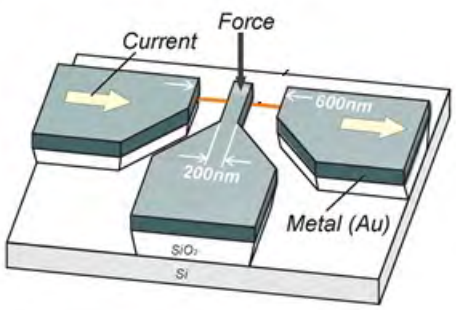

FIGURE 3. Schematic illustration of the fabricated CNT-based nanoelectromechanical force sensor [22].

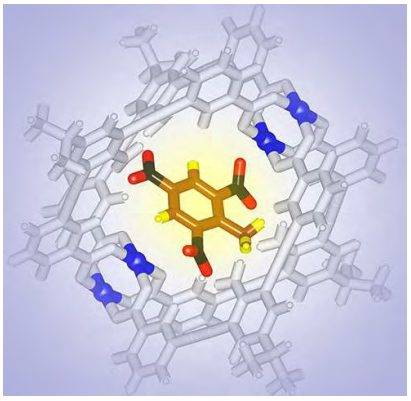

FIGURE 4. Nano-sensors for the detection of TNT [24].

Bio Nano-sensors: Bio-nano-sensors employ biological materials and mechanisms to obtain a measurable biochemical and biophysical signal linked to a specific disease at the level of a single molecule or cell signal, which can be used to detect information regarding a physiological change or the presence of various chemical or biological materials in the environment [25]. Bio-nano-sensors are normally fabricated by incorporating a biological component (e.g., a whole bacterium), a biological product (e.g., an enzyme or antibody), or biomaterials (e.g., biological cells, nucleic acids, proteins, lipids) with or without non-bio-materials so that they can integrate into human body easily. They are mostly used to monitor biomolecular processes such as DNA interactions, antibody, and enzymatic interactions, or cellular communication processes, amongst others [21]

Examples of bio-nano-sensors include genetically modified cells and artificially engineered cells. Bio-nano-sensors are able to detect asthma attacks, lung cancer, common virus. Fig. 5 shows, a Sandwich Assay that combines mechanical and optoplasmonic transduction which can detect cancer biomarkers in serum at ultralow concentrations [27].

B. ARRANGEMENTS

In healthcare applications, two approaches i.e., invasive or non-invasive are used for the implementation of nano-sensors. In invasive approach, inspired by living

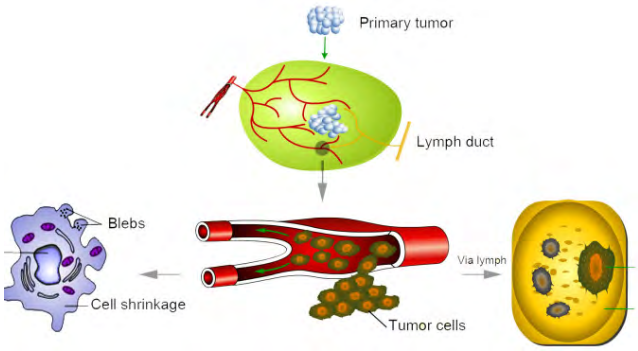

FIGURE 5. Bio-nano-sensor for cancer detection [27].

body natural biological interactions, nano-sensors can be injected/introduced into the body. These sensors can intervene with cells and organs, and communicate with each other in order to exchange information about sensed molecules or chemical concentrations [11]. Whereas in non-invasive approach, sensors (or sensing devices) are placed on or near the surface of the living tissue(s) of the subject. Wearables are an example of non-invasive biosensors used to measure physical and biochemical parameters from the subject without changing their routine and lifestyle.

C. ARCHITECTURAL LAYOUT

With limited communication range and processing capacity, nano-machines use short-range methods to communicate between them. But, by increasing coordination and communication range among several nano-machines, their capabilities and potential applications can be expanded in nanodevices [17], [28]. This will also extend the nano-networks coverage area to reach unprecedented locations. The next natural step is to connect nano-devices with the conventional networks that will define a new networking paradigm referred as the Internet of Nano-Things (IoNT) [12], [28]. Over the last few years, a number of layered architecture and communication approaches have been proposed [12], [17], [29], [30]. In this paper, two most popular nano-communication approaches i.e., molecular communication (MC) and electromagnetic communication (EMC) has been discussed in section III (Nano-Communication). A wireless nano-network architecture is presented in Fig. 6

D. APPLICATIONS OF NANO-SENSORS

Nano-sensors have great potential and incredible applications in all domains of life including, healthcare, environmental monitoring, consumer products, robotics, transportation, security, surveillance, defense, and agriculture etc. Currently, biomedical and healthcare are rapidly growing sectors for nano-sensors due to increasing demand for rapid, compact, accurate and portable diagnostic sensing systems. In biomedical and healthcare area, these devices can offer revolutionary personal healthcare solutions by providing continuous monitoring. The applications of nano-sensors can be

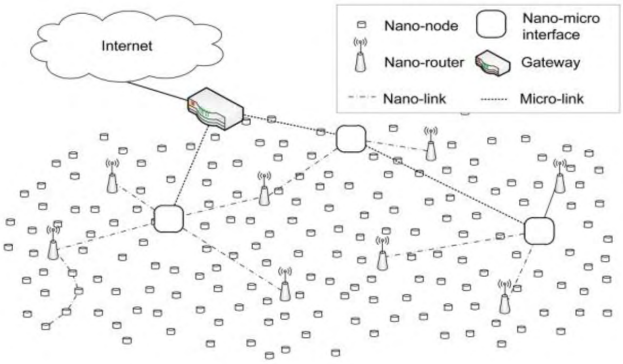

FIGURE 6. Wireless nano-network architecture [30].

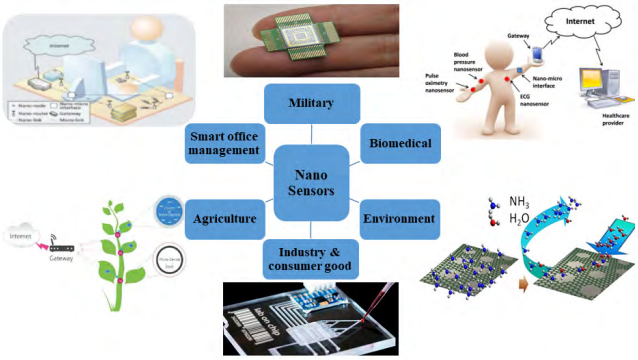

FIGURE 7. Application fields of nano-sensor networks.

divided into the following broad groups: biomedical, environmental, industrial, smart office management, agricultural, and military applications as shown in Fig 7.

E. CHALLENGES/CONSTRAINTS

Improvement in design, size, and biocompatibility of nano-sensors particularly in invasive implementation scheme are challenging requirements for data sensing in an efficient and secure way. The biocompatibility issue is addressed by using the material extracted from natural biomaterials, such as the cell membrane. In invasive approach, metrological considerations have significant importance in general but can become challenging when the biomedical measurement is involved. Therefore, sensors and the instrumentation linked to them need to be calibrated periodically, especially when used for critical diagnosis or therapeutic monitoring.

Inserting engineered sensing devices directly into a subject has certain disadvantages and additional design challenges. In the lab environment, it may cause infection either because of the insertion procedure or discordancy between the organ and the sensing devices. Further complications can arise if the invasive sensing devices are used within the subject for an extended period of time. Therefore safe solutions are required for invasive methods.

III. NANO-COMMUNICATION

By means of communication, nano-sensors will be able to autonomously transmit their sensing information to take actions when needed. In recent years, telemedicine and e-health activities have produced a large number of successful applications in healthcare through different communication technologies. Furthermore, thanks to the big data processing techniques, the health information inside the patient can be collected over a longer period of time, and physicians can perform more reliable analysis rather than relying on the data recorded in short hospital visits [12], [30].

A. METHODS OF NANO-COMMUNICATION AND THEIR CHARACTERISTICS

Nano-communication, the exchange of information at nano-scale, is the only feature that enables nano-machines to work in a synchronous, supervised and cooperative manner to pursue a common objective. However, for the time being, it is still an unsolved challenge to enable the communication among nano-devices. Four main communication paradigms are proposed for nano-networks depending on the technology used to manufacture the nano-machines and the targeted application, namely, nano-electromagnetic, molecular, acoustic, and nano-mechanical [31]. Here we briefly discuss two popular communication methods i.e., electromagnetic and molecular communication two popular methods, and share some examples from the literature.

1) MOLECULAR COMMUNICATION

Molecular communication, the transmission, and reception of information encoded in molecules is a new and interdisciplinary field [32]. In molecular communication, a nano-transmitter releases small particles such as molecules or lipid vesicles into a fluidic or gaseous medium, where the particles propagate until they arrive at a receiver. The nano-receiver then detects and decodes the information encoded in these particles. Messages can be encoded in different properties such as concentration, number, type, release timing, and/or a ratio of molecules. A summary of the design and engineering of components for molecular communication systems from biology, chemistry, and nanotechnology is provided in [25] and [33]. The first approach for engineering bio-nano-sensor has been used and demonstrated in synthetic biology [35]–[37] by modifying a metabolic pathway of a biological cell, which then synthesizes and releases specific signal molecules to send information. Another approach to engineer sender and receiver bio-nano-machines is to create simplified cell-like structures using biological materials (e.g., by embedding proteins in a vesicle) [38], [39].

2) ELECTROMAGNETIC COMMUNICATION

Nano-electromagnetic communication is defined as the transmission and reception of EM radiation from components based on novel nano-materials [32]. The latest advancements in graphene-based electronics have opened the door to EM

communication among nano-devices in the terahertz (THz) band. The refinement of existing architectures and the utilization of new technologies brought THz communication paradigm closer to the reality. In order to enable communication among devices at the nano-scale, THz transmitters should be compact where their area size should reach to hundreds of square nanometers. It is found that electronic sources are capable of providing high levels of average output power at lower THz frequencies [40] and they can be feasible for THz biological research studies. Advancements in microelectronics led to miniature electronic components suitable for intra-body communication [41]. For example, novel miniaturized transistors that adopt non-planar architectures, such as the FinFET [42] and the 3D Tri-gate transistor [43], have been manufactured. Besides their compact size, these architectures mitigate the undesirable behavior of the short channel effect and increase the transistor channel dimension. Nano-antenna can be made of either conventional material or novel materials like carbon nanotube and graphene. A metallic plasmonic nano-antenna is proposed for intra-body nano-networks in [44]. A beam reconfigurable multiple input multiple output (MIMO) antenna system based on graphene nano-patch antenna is proposed in [45], the radiation directions of which can be programmed dynamically, leading to different channel state matrices.

B. CHALLENGES

The integration of in-vivo nano-communication with big data health will expand the potential services that can help in healthcare. Many open challenges [46] are still being investigated, with no mature solutions achieved, robust and large-scale nano-networks [12], [30]. One issue in the physical layer concerns the propagation of signals in various media and environments. Channel models that incorporate path loss, noise and channel capacity for both molecular and electromagnetic nano-communication are needed. In addition, channel characteristics for intra-body nano-networks may vary with health conditions and from person to person, so it is not clear how this variation can be incorporated into channel modeling. One important issue in the link layer is concerned with error handling to improve the reliability of transmission. In light of the capabilities of nano-machines, new coding schemes and error correction mechanisms will have to be developed [31]. The size of nano-machines makes it impractical to have individual network addresses for the individual nodes. Hence, addressing can be cluster-based instead of node-based. This makes it possible to address a group of nodes based on the health functionality they perform or the biological organ or phenomena they monitor [33]. The transmission range is extremely limited, which makes multi-hop communication and routing crucial aspects for nano-networks.

Considering the mobility of nano-machines and indeterministic direction of a communication route, a dynamic routing system is required. Nano-machines suffer from unreliable transmission due to the high path loss and molecular absorption noise [34], which requires the dense deployment of nano-machines. Congestion control is the main challenge, especially in dense nano-networks. The real-time or near-real-time operation is a fundamental requirement in the healthcare application design. However, the unpredictable transmission medium and very short transmission range of intra-body nano-communication bring random delays. Moreover, the heterogeneous properties of nano-machines targeted for various medical purposes will result in different data representations and formats. Therefore, data fusion needs to be optimal, dynamic, and delay-tolerant [30] for applications that rely on the integration of diverse data sources.

1) SIMULATORS FOR NANO-COMMUNICATION

Besides the research advancements in communication and hardware devices, a modular and freely available simulation platform is also highly required to enable research activities to achieve the nano-communication. NanoNS [45] and N3Sim [47] are simulators for diffusion-based molecular communication among immobile nano-devices. A simulator of Brownian motion was investigated in [48], where a dual time-step approach was adopted to manage the runtime complexity brought by a large number of particles. A High-Level Architecture-based simulator design was proposed to provide a comprehensive and scalable platform for molecular communication performance evaluation [49]. A generic simulation platform is proposed to fit in with multiple nano-devices, channel models, molecular propagation models and nano-devices mobility models [50].

A simulation tool for bacteria-based communication, BNSim, is introduced in [51]. Recently, a new network simulator nanoNS3 for modeling bacterial molecular communication is developed atop Network Simulator 3 (NS-3) in [52]. The features of existing molecular communication simulators NanoNS, N3Sim, Nano-sim, and nanoNS3 can be summarized in [52]. Meanwhile, with respect to the EM-based nano-networks, Nano-Sim is developed in [53] and [54] as an open source simulator implemented on the top of the platform NS-3 [55]. The simulator models the three basic types of nodes in a nano-sensor network: nano-machines, nano-routers, and nano-interfaces.

Pulse-based communication protocols are implemented in the PHY/MAC layer, and a routing protocol based on selective forwarding is implemented on the network layer. The properties of ultrasound communication for nano-networks were evaluated via a simulation study on detailed channel modeling and network protocols in [56]. Existing simulators use a simplified approximation of the receiver response thus affecting the accuracy of the simulation. For the time being, none of these simulators could capture all features characterizing the nano-communication as one completed platform. A key challenge is, therefore, to integrate a large number of tools into a single package and to make it possible to consistently compare and evaluate various designs for nano-communication. More research is required to develop a mature, flexible and robust platform.

IV. BIG DATA ANALYTICS FOR HEALTHCARE DATA

The advancement of nanotechnology and rapid emergence of nano-sensors would ultimately lead to a network of millions of interconnected devices on a nanoscale. Consequently, such a network would generate a vast amount of data which inspires for the envisioned predictive, preventive, personalized and participatory health-care framework. The knowledge extracted from the large volume of information can induce a reduction in healthcare cost by enabling early diagnosis. From a patient's perceptive, continuous assessment of the biomarkers obtained via IoBN could either altogether prevent or detect early onset of fatal diseases, which is beyond what is possible today. Likewise, the synergy between nanotechnology and big data paradigm strive to address the challenges of collecting long-term volumes of complex and diverse health information, integrating data from sensor arrays, and analyzing massive amount information that would drive the clinical research forward.

In this section first, we describe the characteristics of big data in the context of healthcare systems. Subsequently, aligned with the vision of connected nano-sensors we discuss the related knowledge extraction methods equally applicable to our scenario. While we explain the framework, we categorize and describe sources that contribute to big data in healthcare. We also discuss salient enabling technologies for big data-driven analytics framework.

Doug [57] tried to characterize the future data challenges in three dimensions: *Volume*, *Velocity*, and *Variety* which emerged later as major characteristics of big data. In addition to the three V's, other dimensions of big data have also been mentioned like Veracity, Variability (complexity) and Value, terms are introduced by IBM, SAS and Oracle respectively [58]. However, the general consensus to describe big data is based on four major attributes (4Vs) [59], which are discussed as follows:

Volume: Consider an implementation scheme for ubiquitous sensing with millions of nano-sensors in and off body networks. A frequent communication is expected between sensors and the central nodes or routers resulting in the generation of a huge amount of continuous data. The emergence of bio-nano sensor paradigm would result in a step change, the way we measure and sense. A clear change would be a transition from non-continuous to a continuous source of information since on-body sensors are accepted to transmit information continuously. The merger of such data from multiple networks (internet of nano-things) will increase the data size manifolds. Beside that existing medical databases also contain an abundance of data. It will require big data processing and analytics tools to handle such scale of information.

Velocity: As discussed above, smart devices generate a continuous stream of information as opposed to offline data sources such as health records etc. This demands specialized technologies for collecting, merging and processing the information in real-time while maintaining the data integrity. Nano-sensors have limited energy and data storage capacity, so, for the transfer of information, data can be sent in many small packets with high speed to keep the integrity of the data intact. Therefore, in a bio-nano-sensor implementation scheme, frequent communication is a kind of requirement [60]. Besides the speed of data generation due to frequent communication, another factor contributing towards the velocity is the requirement of real-time analysis for timely and informed decision making. The speed of the data generation and requirement of real-time analysis makes it a big data analytics problem.

Variety: Heterogeneity of smart nano-sensors and diversity in underlying communication technology will lead to a variety of data types produced. In addition, variety can arise by combining heterogeneous data from multiple sources like sensors, clinical examination, medical reports, lab tests etc. This data can be structured like medical databases, semi-structured like digital logs of sensor communication and unstructured like text reports and visual aids.

Value: The value of the data gathered from single source e.g. nano-sensors depends on the factors like age of data and end goals. Fresh data may have more value than the data generated long ago, similarly for a particular problem where data generated from nano-sensors can be enough for decision making the value of the data generated is much more than the other cases where it plays only a partial role. However, a fusion of data from nano-sensors with the data from other sources like electronic medical records (EMRs), can enhance the value of data. In general, it is observed that in healthcare value of single data chunk from a certain time period decreases over the time, but the value of fused or aggregated data increases over the time [61].

A. BIG DATA ANALYTICS FRAMEWORK

In this section, we have presented an analytics framework as shown in Fig.8 that is in-line with the vision of exploiting nano-sensors network data in conjunction with data from conventional sources. The proposed framework comprises three layers: Layer 1 refers to data source layer whereas Layer 2 consists of enabling technologies for data collection and processing. Finally, layer 3 refers to the analytics engine that makes use of state-of-the-art machine learning algorithms to convert raw data into actionable information. These layers are further discussed in detail in the subsequent subsections. Besides that, some use cases from healthcare are presented in Table 1 for each knowledge extraction step.

1) LAYER 1: DATA SOURCES

The transformation of data to knowledge starts from the data collection stage. In our proposed big data analytics pipeline, layer one represents data sources. Today, main sources of healthcare data are miscellaneous smart healthcare devices and electronic records stored in conventional databases. We have categorized these data sources into two broad categories named as Conventional and Non-Conventional sources which are further discussed below:

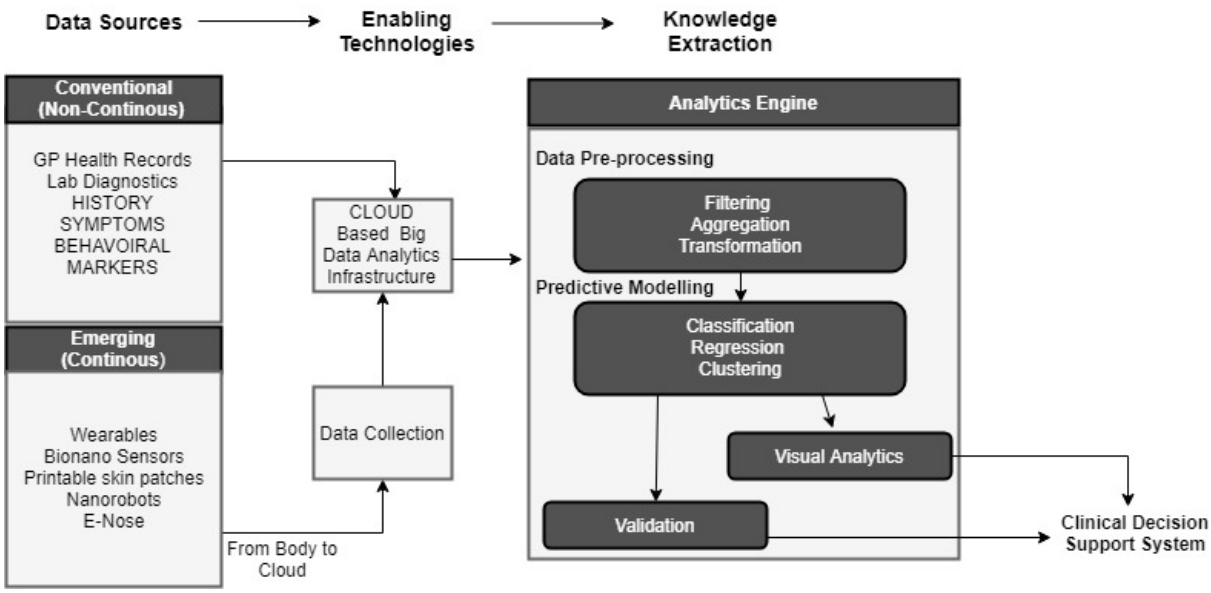

FIGURE 8. Envisaged Big data analytics framework for extracting intelligence from healthcare data.

a: CONVENTIONAL DATA SOURCES

As reviewed by Steinhubl *et al.* [4], the most common type of information from the human body using the current technology includes temperature, pressure, glucose concentration, cardiac heartbeats etc. These biomarkers are analyzed using various data processing techniques to establish patterns that lead to specific diseases. These patterns are further correlated with conventional health-screening technologies such as X-ray tomography, magnetic resonance, ultrasound etc. to provide an accurate diagnosis of a possible medical condition.

Likewise, another source of information with regards to patients' data comes from the text records such as Electronic Health Records (EHR). The EHR contains textual reports on symptoms and corresponding diseases. In the current literature, the knowledge extracted from these sources relies heavily on a myriad of methods from signal processing, image recognition, and text analytics domain.

The amount of information obtained from these conventional sources are huge in volume and also heterogeneous in nature The non-standardization of data format and presence of multiple players in the market offering a variety of hardware and software make data management even more challenging. Similarly, the presence of heterogeneous data makes it impossible for the conventional data storage and processing techniques to extract any meaningful information.

With the help of nonconventional knowledge extraction tools like big data analytics, a world of value can be generated out of this data to improve the overall healthcare system. There already exists an ongoing debate on the possible role of big data analytics in healthcare [5], [6], [8], [61], [62], and its practical implications [63]–[66] that is very promising.

b: NON-CONVENTIONAL CONTINUOUS DATA

We are entering a new era of ubiquitous healthcare services which entitles real-time diagnosis, monitoring, and treatment services anywhere. Emerging technologies for smart healthcare devices are very promising for the realization of this dream. These smart devices vary in their size, capacities, and functions. Example of smart medical devices today vary from surgery assistant robots [67] to wearable and implantable smart devices [68] for continuous monitoring [69]. Due to the continuity of progress in underlying technologies smaller, efficient and cheap devices like nano-sensors are becoming the norm. Numerous nano-sensors working together in the form of a network can be used in-off body for sensing. These smart devices and body-centric sensor-networks can perform personalized healthcare operations like monitoring and drug delivery, in real-time. Such an always-on system is characterized by constant communication of devices among themselves. This machine to machine communication will lead to the generation of nonconventional high speed, high volume streams of data.

The implantable biosensors offer a possibility of continuous assessment of health conditions as opposed to traditional diagnostics methods i.e., laboratory testing of blood and urine. The real-time sensing on a molecular level such as metabolites [70] can help proactively detect health abnormalities in advance. Matzeu *et al.* [71] have shown that the presence of drugs and events of gastroesophageal reflux disease can be detected by continuous monitoring of body fluids such as saliva, tears, and sweats. Likewise, advancement in nanoscale technology has made it possible to sense the level of glucose in the tears via contact lenses [71].

The skin-like stretchable devices which are able to attach to the human skin, known as Biostamps [72], have been

TABLE 1. Knowledge extraction steps and use cases for healthcare.

Paper	Task	Data	Method	Problem Addressed
[84]	Data Pre-processing	Eight different bio datasets	Locally auto-weighted least squares imputation	Missing values estimation
[83]		Medical Records	latent factor model	Reconstruct missing values
[109]	Data Fusion	Personal health record(PHR)	ontology and NLP	Interoperability & integration
[108]		Five gene expression datasets	p-norm singular value decomposition	Dimension reduction
[110]		Simulation data two gene expression datasets	Non-convex proximal P-norm-graph-Laplacian PCA	Dimension reduction
[96]	Feature Extraction	Demographic features & Heart performance indicators	Particle Swarm Optimization	Feature Selection
[97]		Simulated genotypes data	Boosting	Overcome bias/imbalance
[83]		Medical Records	CNN	Feature extraction from text
[111]		ECG	User-defined threshold level	Efficiency and low latency for fog computing
[112]	Visualization	DNA microarrays data	Point clouds, solar systems, and treemaps	Visualization of genes sequential patterns
[87]	Fuzzy Logic	Simulation	Fuzzy logic	Decision-making systems
[100]		ECG, heart rate, 3-axes human body acceleration, temperature	Fuzzy logic	Recognizing mental or physical stress
[102]		3-axis acceleration, heart rate, breathing rate	Fuzzified neural network	The unusual physical condition detection
[101]		Body mass index, waist circumference, waist-to-height ratio	Fuzzy logic	Abdominal obesity & cardio-metabolic risks assessment
[103]		lab tests, historical data, sensor data	Self-learning fuzzy logic scheme	Optimized recommendation system
[97]	Supervised Learning	Gastric cancer dataset	Coarse-to-fine learning	Detection of suspicious loci
[83]		Medical Records, Structured and unstructured(text)	Convolutional neural network (CNN)	Disease Risk Prediction
[113]		Subjective data: A measure of fatigue, stress, anger, depression, environment	Multi-layer perceptron, SVR, GRNN, and KNN	Predict psychological wellness
[96]		Demographic features an& Heart performance indicators	Bagged Tree, Random Forest, and AdaBoost	Detect the heart disease
[110]		Theoretical framework	Attentive vision-based algorithm	Pill Detection & Identification via image recognition
[105]		Microscopic images of biopsy	SVM, CNN	Classification of breast cancer
[108]	Unsupervised Learning	Five gene expression datasets	K-Means	Tumor clustering

manufactured with an aim to monitor body vitals. These devices harvest energy from radio signals relayed to them by a wearable device. Likewise, there are several studies [73] that have reported the use of printable sensor array for measuring electrical impedance on the patients' skin.

Nano-sensors are expected to have diversity because of different manufacturing materials, applications, implementation requirements and underlying technologies. Assuming huge variety in nano-sensors, generation of diverse data types is quite obvious besides the speed. This heterogeneous data form conglomerate of smart devices would comprise structured, semi-structured and unstructured data.

To summarize data generated from smart healthcare devices is characterized by all 4 Vs of big data. The advancement in data processing techniques such as real-time stream processing solutions like Spark or Storm engine on top of Hadoop would allow us to envisage proactive disease prediction and prevention solutions.

2) LAYER 2: ENABLING TECHNOLOGIES

To envision and develop nano-sensors based comprehensive big data analytics solutions, an equally challenging task is data integration. This requires collecting information from conventional and non-conventional data sources and storing

this information in a unified space in order to be processed later for knowledge extraction. Economically or technically it is not feasible for all entities to create such facilities in-house to sotre and process that huge amount of data. Besides that extending of storage and computational capacity according to the demands is also a challenging task. Thanks to robust flexible cloud computing facilities which are becoming the backbone of current big data analytics frameworks that can help to transfer the data from the body to the cloud. The availability of the data in a single space can enable developing a patient-centered model allowing practitioners to quickly access and study long-term patient information.

These cloud-based solutions not only reduce healthcare cost by enabling early diagnosis but from a physician perspective, it allows them to share successful treatments with colleagues or researchers. In recent times one such example of a clinical database is established by Harvar's project entitled informatics for integrating biology and bedside (i2b2) [74]. This platform is currently adopted by over 100 medical institutions worldwide. Likewise, other cloud-based clinical initiatives include Health Cloud Exchange [75], e-Health Cloud [76], [77], and Husky Healthcare social Cloud [78]. The cloud supports the integration of a system that is heterogeneous and geographically separated and becomes an obvious choice for healthcare conglomerates.

Amazon Web Services, Google Cloud, IBM Cloud, Microsoft Azure, Cloudera, Hortonworks, and MapR, are the few of the popular names who provide cloud service commercially. They also support batch processing with Hadoop suitable for conventional data and stream processing facilities like Spark for continuous data, along with many other big data analytics tools. Rallapalli *et al.* [66] also provide a general framework for Big Data analytics in health care.

Fog computing and edge computing [111] are also emerging concepts for data storage and processing for the urgent decision making. In this case another layer of data storage and processing is added between cloud and the data source. Data related to urgent decisions can be processed at this layer and actions necessary can be taken with low recall time.

Effective and efficient machine learning algorithm for gathering intelligence from distributed data from miscellaneous sources is also challenging requirement. Advances on the front of deep learning for IoT environment are promising. Similar deep learning approaches can be replicated for healthcare use cases of nano-sensors applications.

3) LAYER 3: KNOWLEDGE EXTRACTION

A pile of data itself has no value until unless it is converted into knowledge for logical conclusions and decision making. In parallel, advances in data mining, machine learning, computational technologies, biomedical testing and statistical techniques altogether have revolutionized the knowledge extraction process. The Knowledge Extraction (KE) process can be divided into following phases: preprocessing, data fusion, feature extraction, prediction and visualization, which are subsequently discussed in detail.

a: PREPROCESSING

Healthcare big data can be diverse in formats, may contain missing values, noise, errors and inconsistencies. Such gross data may affect the performance of the analysis and lead to incorrect results. So it needs to be preprocessed, i.e. formatted and cleaned, before any prediction model or data analytics technique is applied. It is a common understanding among data scientists that preprocessing takes more time than the model implementation. Data preprocessing is a very laborious but important step to overcome problems discussed below:

Missing values: In preprocessing one of the biggest challenges is to deal with missing values. Medical records can contain missing values due to multiple reasons like data not provided by patients, not entered in records by hospitals, machine or human error etc. Stockdale and Royal [79] and Wells *et al.* [80] categorize missing data in the context of healthcare as missing not at random (MNAR), missing at random (MAR), and missing completely at random (MCAR). Stockdale and Royal [79] discuss three popular strategies to handle missing values. Easiest and common approach is to ignore data with missing values and take available complete cases. Instances or features with missing value can be ignored simply in this approach. It can be opted where the rest of data can lead to meaningful conclusions anyway. Other is the use of single value imputation method, in this popular approach missing values are replaced by some alternative value like zeros or mean, mod etc. [81], [82] of the relevant data. Most advance approach is, model-based imputation where some prediction method can be used to attain the values close to possible actual data with the help of available data [83]. Yu *et al.* [84] take model-based approach and use locally auto-weighted least squares imputation (LAW-LSimpute) for the estimation of missing values in microarray data, it automatically weights the neighboring genes based on their importance. Another example of imputation method is the use of latent factor model by [83] to replace the missing values in the patient data using the latent factor matrix developed by the combination of patients features in particular scenario. White *et al.* [85] discuss in detail a method of multiple imputation using chained equations (MICE) and [80] recommends it as one of very effective approach for handling missing values.

Erroneous Data and Noise: They are the factors that also highly influence the quality of analysis, [86] explores four approaches for noise removal. Three of these methods are basically outlier detection techniques based on distance calculation, clustering, and local outlier factor (LOF). The fourth one is a hyper clique-based data cleaning technique. George and Singh [87] recommend the use of Fuzzy logic as a resilient and consistent approach to handle noise.

b: DATA FUSION

Data aggregation from heterogeneous sources is required at different levels like routers, local data processing

units (LDPUs) and cloud. Data aggregation schemes for nano-sensors also depend on the network topology used for nano-networks.

Alam *et al.* [88] discuss three machine learning data fusion techniques for IoT data, named as probabilistic methods, artificial intelligence, and theory of belief. Whereas Misra and Chatterjee [89] categorize data aggregation techniques as cluster-based, tree-based, or structure-free algorithms. They define these categories for sensors network data based on networking approaches and consider clustering based approach most popular one. Bagaa [90] think the tree-based approach is adopted commonly for the flow of data in wireless sensor networks (WSN). Different data aggregation schemes for tree-based networks are also discussed in the literature [91]–[94]. These schemes or their modified versions can also be opted for data aggregation in nano-sensors particularly electromagnetic ones to overcome latency.

Aggregations of data from routers, sinks or LDPUs to upload it to the cloud is a separate task from the data aggregation from multiple sensors discussed above. Misra and Chatterjee [89] propose a solution comprising Body Area Network Data Aggregation Algorithm (Banag) for data aggregation and an Optimal Channelization Algorithm (OCA) for data transfer to cloud. This solution prioritizes data aggregation and transportation to cloud on the basis of urgency and criticality.

c: FEATURES EXTRACTION

Healthcare big data gathered from different sources and comprising a variety of data records can have numerous data features at the end. Use of all the features may not be feasible and can lead to the curse of dimensionality. Each subgroup from the complete feature space can lead to a different type of information, and findings of varying quality as the result of the analysis performed. So the selection of the right features become very important to reach desired information with a certain accuracy.

Measuring correlation of features with the predicate and among themselves can be a good approach to find relevant features that can lead to meaningful results. Existing feature selection methods can be categorized as wrappers, filters and embedded [95]. Wrappers are algorithms that choose a subset of features for best performance on the basis of an iterative process, being exhaustive in nature they may not be feasible in case of high dimensions. Filtering is simple and fast as it takes feature characteristics into consideration for ranking and uses this ranking score for feature selection instead of relying on some algorithm. Embedding is a blend of wrapper and filter. It carries advantages of the both while overcoming the iterative search bottleneck of wrapping. Yekkala *et al.* [96] have used a filtering technique called Particle Swarm optimization technique, a computationally inexpensive method in terms of memory and speed. It is applied to the sensors data set to select a subset of features discarding irrelevant data to detect heart disease using different enabling algorithms. Bao *et al.* [97] use a coarse based wrapper approach for genomic level feature selection, on multiple balanced samples subgroup selected randomly.

Chen *et al.* [83] use expert knowledge and Pearson correlation to extract 79 features, related to patient's demographic information, cerebral infarction and living habits (e.g. smoking), from structured medical data to predict the risk of cerebral infarction disease. They use keyBword selection method to extract important features from text data and then use convolutional neural networks CNN to extract 100 features from them using multilayer approach. At each layer, the combination of words are selected and passed to the next layer from the text data with an assigned weight, in the end, features from text and structured data are combined. Similarly, in case of nano-sensor networks a combinatorial approach can be adopted to jointly exploit features from conventional and non-conventional data sources to achieve higher intelligence.

d: DATA VISUALIZATION

Data visualization techniques are used to make sense of the vast amount of data graphically allowing the users to interact with the data for knowledge discovery. Data visualization techniques can be broadly divided into Scientific Visualization (SciViz) and Information Visualization (InfoViz). The data obtained from computerized tomography (CT Scan) and magnetic resonance imaging (MRI) are examples of SciViz. InfoViz, on the other hand, is a representation of complex models which do not necessarily has a physical meaning. The visualization of data obtained from the sensors or developing a visual representation of textual health records of the patient comes under the realm of InfoViz. Lifeline systems [98] and cube techniques [99] are examples of InfoViz that allows examining patients visually based on their health records attributes. Preprocessing techniques such as principal component analysis PCA, and singular value decomposition (SVD) are often applied to reduce the dimensionality of the data that further aids in the data visualization. We foresee that similar approaches will be adopted for the visualization of data obtained from bio-sensor nodes.

e: PREDICTIVE MODELING

If data is the backbone of knowledge extraction process then machine learning based predictive modeling plays a role of the brain. Robust machine learning algorithms on big data platforms can not only find hidden patterns and correlations in the data to provide an insight about existing realities but it can also exploit them to predict possible future outcomes very accurately. Here we try to highlight the role of machine learning in knowledge extraction by discussing some common techniques used in the healthcare sector.

Fuzzy Logic, a rules-based method, is a very popular technique in health care [87], [100]–[103] for decision making according to a predefined instructions set. Reference [103] is a good example for understanding how fuzzy logic can work in the context of healthcare. Authors propose a self-learning Fuzzy Expert System that takes as input the blood pressure (BP), heart rate (HR), blood sugar (BS) data, collected from

continuous monitoring. Then it fuzzifies (compares) the data against predefined fuzzy rules and prescribes on the basis of those rules. The salient characteristic of the proposed system is a loopback feature that improves fuzzy rules, knowledge base, and recommendation set continuously.

Apart from fuzzy logic prediction and approaches can be divded into supervised or unsupervised categories.

i) SUPERVISED LEARNING PROBLEMS

They are the problems where the predicate, the outcome parameter is known. In supervised learning, problems can be further subcategorized as classification problems or regression problems. Cases, where the goal is to separate different instances from a data into distinct groups on the basis of some attributes, is called classification, for example, categorizing different patients on the basis of their medical stats.

The cases where we are interested to estimate something in terms of continuous numeric values they are called regression problems, for example, calculating chances of a patient to get a disease or predicting expected pulse rate at a certain time and particular circumstances on the basis of previous data. There are plentiful algorithms for each type of problem.

In healthcare frequent problems discussed are related to classification like genome or cancer type classification. There are many machine learning algorithms available for classification, some popular conventional algorithms include Naive Bayesian (NB), K-nearest Neighbour (KNN), Decision Tree (DT), Hidden Markov Model(HMM), Neural Networks(NN) and Support Vector Machine(SVM). In addition to that advance deep learning methods are also getting popular. Chen *et al.* [83] have used three conventional machine learning algorithms Naive Bayesian (NB), K-nearest Neighbor (KNN), and Decision Tree (DT) to predict the risk of cerebral infarction disease in patients using the patients' demographic and cerebral infarction data.

One of the popular application of classification algorithms is cancer detection in healthcare research. Huda *et al.* [104] have tried to classify the subtype of Oligodendroglioma tumor by analyzing images of Pathology samples. They have proposed (GANNIGMA-ensemble) model comprising globally optimized Artificial Neural Network Input Gain Measurement Approximation (GANNIGMA) combined with an ensemble classification (GANNIGMA-ensemble) technique to generate the diagnostic decision rule. The GANNIGMA hybrid feature selection finds the significant features and ensemble applies a combination of algorithms. Bardou *et al.* [105] have used a very popular classification algorithm SVM in combination with convolutional neural networks to classify breast cancer subcategory by performing analysis on microscopic images of biopsy. A Hidden Markov model-based method is proposed by [106] for behavioral profiling of dementia patients using data collected by continuous monitoring with the help of in-home sensors. Bao *et al.* [97] have applied a unique coarse-to-fine learning approach on genome data to identify whether a loci is suspicious or not as a carrier of gastric cancer.

ii) UNSUPERVISED LEARNING PROBLEMS

In healthcare, another common category of problems fall in unsupervised learning, where we have a pile of data but the predicate or outcome variable is not defined. This approach is commonly used to explore and group the data on the basis of its characteristics using some clustering method.

There are different clustering approaches like Partitional, Density-based, Hierarchical, Spectral, Grid-based, Gravitational, Correlation, and Herd clustering [107]. There also exist many clustering algorithms but some prominent of them are Kmeans, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Mean-Shift Clustering, Expectation–Maximization (EM), Gaussian Mixture Models (GMM) etc. Clustering algorithms are very important for the analysis of healthcare data as in many cases the ground truth is not known and all we have is bulk of data and we have to extract knowledge out of it. For example, we have miscellaneous heart rate, blood pressure, glucose level data of men and women patients and we do not want to estimate anything or categorize patients into any categories. All we want is to explore data for any possible hidden meaning in it or want to group together patients on the basis of similarities data itself reflects. Kong *et al.* [108] apply K-means clustering method for tumor clustering based on the low-rank approximation matrix.

V. OPEN RESEARCH CHALLENGES

Application of nano-technology and big data analytics in healthcare is emerging as a new integrated area of study. Growth in bio-nano-technology and big data healthcare is likely to attract more interest of researchers in this area. Already a lot of progress can be seen in the area of bio-nano-technology and big data healthcare independently, but their interdisciplinary study is at nascent stages. Therefore there exist many open challenges that seek intention of the researcher, some important challenges are given below:

A. STANDARDIZATION OF DATA FORMATS AND PROTOCOLS

There are so many vendors who deliver a plethora of miscellaneous equipment, technologies, and services for diagnosis, monitoring, and treatment in healthcare. This diversity of equipment and technology lead to the generation of numerous data formats and transfer protocols, for example, several GB of raw data generated by ECG in a day can be transferred in XML format and skin images taken by a camera can be multimedia files [114]. Handling such a variety of data is a challenging task. It becomes even more crucial for the future smart healthcare system appraised for the efficiency of real-time services. In such scenario, it becomes important that stakeholders step up for standardization of data formats and protocols to reach some unified solution. In addition, attention should be paid to the fact that data formats and protocols should be light and efficient to support nano-communication [115].

B. UNIFIED DATA SCHEMA

As mentioned earlier there are numerous stakeholders involved in data generation and management. They use different schema for the data storage which makes data integration and interoperability difficult. There also exists some efforts for the standardization of different aspects of data exchange and interoperability for conventional health care data, for example in March 2003, the Consolidated Health Informatics (CHI) set requirement that all federal health care services agencies adopt the primary clinical messaging format standards (i.e., the Health Level Seven [HL7] Version 2.x [V2.x] series for clinical data messaging, Digital Imaging and Communications in Medicine [DICOM] for medical images, National Council for Prescription Drug Programs [NCPDP] Script for retail pharmacy messaging, Institute of Electrical and Electronics Engineers [IEEE] standards for medical devices, and Logical Observation Identifiers, Names and Codes [LOINC] for reporting of laboratory results) [116]. Fast Healthcare Interoperability Resources (FHIR) by HL7 is another standard for exchanging healthcare information electronically [117]. Similar standards are also required for non-conventional data sources. Data storage schema may vary due to different placement arrangements of nano-sensors as well. It is desired that a unified data storage schema be adopted by all stakeholders so exchange and integration of data could be easy and efficient for the provision of seamless smart healthcare services.

C. DEVELOPMENT OF PARAMETERS AND METHODS TO VALIDATE THE QUALITY OF DATA

In conventional data analytics, sample data was taken into account, and a refined sample, free of errors and missing values was opted for analytics. But big data analytics has shifted the paradigm from sample to whole data analytics. This whole data approach has also raised a concern about quantity vs quality [118]. Increase in data amount also increase the chances of missing and erroneous data. It is common to have missing values [79], noise or other parameters that may affect the quality of healthcare big data. This bad quality input data can affect the performance of analytics and may lead to incorrect findings. It is the demand of time that new robust parameters to compare the quality of data be defined and new data validation methods should be developed to assess and improve the quality of raw data.

D. THE MECHANISM FOR DEFINING LEVEL AND SCALE OF AGGREGATION

Data fusion or aggregation in big data from continuous and conventional sources in itself is a very challenging task not only because of volume, heterogeneous nature and velocity of data but also because of the purpose of aggregation. In nano-sensor-network, whether the data be aggregated at the router level, at any interim smart hop or in the cloud, it all depends on the end goals and utility of data. Similarly, the decision about the scale of the data aggregation also depends on the subject and purpose of data aggregation. Nevertheless, keeping all these aspects in consideration, some generalized rules should be defined for level and scale of data aggregation. It will make not only the data sharing, processing and analysis easy but it can also help to improve the privacy and security of data.

E. ANALYTICS TOOLS

Analytics has shifted from the sample-based approach to whole data based methodology. But most of the existing algorithms are designed to extract knowledge from sample data and they are not efficient for large-scale data and there is a scarcity of big data analytics techniques. Therefore new scalable techniques and algorithms are required to perform analytics on big data. For example in healthcare a very common task is clustering which requires robust clustering algorithm scalable to big data problems, in addition, metrics to validate the clustering results are also required. Besides that such algorithms are required which can also perform for real-time analysis on distributed data.

F. PRIVACY

Where the capacity of big data analytics to find a needle from the haystack is its merit, there it also raises a serious question on how to maintain the privacy of individuals and institutions. Though different anonymization approaches exist, still knowledge of few key attributes and application of some data mining can help to trace individuals. For the security and privacy, new anonymization methods should be developed.

G. SECURITY

Continues machine to machine communication particularly at a very small device like nano-sensor which do not have much processing power, data security becomes a challenging tasks against data hacking and malicious adversary attempts. New robust techniques are required to ensure data security and privacy.

H. HYBRID SOLUTIONS

Since there are lots of communication paradigms for nano-communication, the study on the interaction between two different communications paradigm is still missing. It is generally believed that by merging all the communications together the nano-network would be much more flexible and powerful. Hence studies on hybrid communication mechanism and their feasibility are much needed future direction.

VI. CONCLUSION

In future, such ubiquitous and smart healthcare system is desired which is equipped with P4 (i.e. predictive, preventive, personalized and participatory) to perform diagnostics, monitoring and treatment functions in seamless and smart manner. Developments in the field of big data analytics and nano-technologies in the context of healthcare are very promising to realize this dream. Smart devices based on nano-sensor networks can provide technical support to take critical

measurements on real time. Big data analytics can help to gather intelligence from those measurements data to make important and urgent decisions for diagnosis to drug delivery. Therefore, there is a need of thorough research on this interdisciplinary area. In this paper we have covered the role of nano-sensors, nano-communication and big data analytics towards futuristic healthcare system and emphasise the need to open the door to an interdisciplinary research discussing some pressing challenges.

REFERENCES

[1] Ministry of Health and Long-Term Care. *Preventing and Managing Chronic Disease: Ontario's Framework*. Accessed: Jun. 9, 2018. [Online]. Available: http://www.health.gov.on.ca/en/pro/programs/cdpm/pdf/framework_full.pdf

[2] (Feb. 1, 2018). *Cancer*. Accessed: Apr. 4, 2018. [Online]. Available: http://www.who.int/news-room/fact-sheets/detail/cancer

[3] M. Sagner *et al.*, "The P4 health spectrum—A predictive, preventive, personalized and participatory continuum for promoting healthspan," *Prog. Cardiovascular Diseases*, vol. 59, no. 5, pp. 506–521, 2017.

[4] S. R. Steinhubl, E. D. Muse, and E. J. Topol, "The emerging field of mobile health," *Sci.*, vol. 7, no. 283, p. 283rv3, 2015.

[5] J. Andreu-Perez, C. C. Y. Poon, R. D. Merrifield, S. T. C. Wong, and G.-Z. Yang, "Big data for health," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 4, pp. 1193–1208, Jul. 2015.

[6] M. M. Hansen, T. Miron-Shatz, A. Y. S. Lau, and C. Paton, "Big data in science and healthcare: A review of recent literature and perspectives," *Yearbook Med. Inform.*, vol. 23, no. 1, pp. 21–26, 2014.

[7] Y. Liang and A. Kelemen, "Big data science and its applications in health and medical research: Challenges and opportunities," *Austin J. Biometrics Biostatist.*, vol. 3, no. 1, pp. 1–9, 2016.

[8] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: Promise and potential," *Health Inf. Sci. Syst.*, vol. 2, no. 1, pp. 1–10, 2014.

[9] M. A. Alhumud, M. A. Hossain, and M. Masud, "Perspective of health data interoperability on cloud-based medical cyber-physical systems," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2016, pp. 1–6.

[10] I. D. Dinov *et al.*, "The perfect neuroimaging-genetics-computation storm: Collision of petabytes of data, millions of hardware devices and thousands of software tools," *Brain Imag. Behav.*, vol. 8, no. 2, pp. 311–322, 2014.

[11] N. Agoulmine, K. Kim, S. Kim, T. Rim, J.-S. Lee, and M. Meyyappan, "Enabling communication and cooperation in bio-nanosensor networks: Toward innovative healthcare solutions," *IEEE Wireless Commun.*, vol. 19, no. 5, pp. 42–51, Oct. 2012.

[12] N. A. Ali and M. Abu-Elkheir, "Internet of Nano-Things healthcare applications: Requirements, opportunities, and challenges," in *Proc. IEEE 11th Int. Conf. Wireless Mobile Comput., Netw. Commun. (WiMob)*, Oct. 2015, pp. 9–14.

[13] Y. Okaie, T. Nakano, T. Obuchi, and T. Hara, "Research challenges in bionanosensor networks," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr. 2016, pp. 1035–1036.

[14] F. Dressler and S. Fischer, "Connecting in-body nano communication with body area networks: Challenges and opportunities of the Internet of Nano Things," *Nano Commun. Netw.*, vol. 6, no. 2, pp. 29–38, 2015.

[15] *Wearables*. Accessed: May 10, 2018. [Online]. Available: https://www.statista.com/outlook/319/100/wearables/worldwide#

[16] (May 9, 2013). Accessed: May 10, 2018. [Online]. Available: https://www.abiresearch.com/press/more-than-30-billion-devices-will-wirelessly-conne

[17] I. F. Akyildiz and J. M. Jornet, "Electromagnetic wireless nanosensor networks," *Nano Commun. Netw.*, vol. 1, no. 1, pp. 3–19, 2010.

[18] M. S. Mannoor *et al.*, "Graphene-based wireless bacteria detection on tooth enamel," *Nature Commun.*, vol. 3, Mar. 2012, Art. no. 763.

[19] A. Guney, B. Atakan, and O. B. Akan, "Mobile ad hoc nanonetworks with collision-based molecular communication," *IEEE Trans. Mobile Comput.*, vol. 11, no. 3, pp. 353–366, Mar. 2012.

[20] H. Lee *et al.*, "A graphene-based electrochemical device with thermoresponsive microneedles for diabetes monitoring and therapy," *Nature Nanotechnol.*, vol. 11, pp. 566–572, Mar. 2016.

[21] T. Nakano, T. Suda, Y. Okaie, M. J. Moore, and A. V. Vasilakos, "Molecular communication among biological nanomachines: A layered architecture and research issues," *IEEE Trans. Nanobiosci.*, vol. 13, no. 3, pp. 169–197, Sep. 2014.

[22] C. Stampfer, A. Jungen, and C. Hierold, "Fabrication of discrete nanoscaled force sensors based on single-walled carbon nanotubes," *IEEE Sensors J.*, vol. 6, no. 3, pp. 613–617, Jun. 2006.

[23] S. C. B. Gopinath, T.-H. Tang, Y. Chen, M. Citartan, and T. Lakshmipriya, "Bacterial detection: From microscope to smartphone," *Biosensors Bioelectron.*, vol. 60, pp. 332–342, Oct. 2014.

[24] A. Cao *et al.*, "Metal–organic polyhedra-coated Si nanowires for the sensitive detection of trace explosives," *Nano Lett.*, vol. 17, no. 1, pp. 1–7, 2017.

[25] T. Nakano, M. J. Moore, F. Wei, A. V. Vasilakos, and J. Shuai, "Molecular communication and networking: Opportunities and challenges," *IEEE Trans. Nanobiosci.*, vol. 11, no. 2, pp. 135–148, Jun. 2012.

[26] C. Wang *et al.*, "An aptameric graphene nanosensor for label-free detection of small-molecule biomarkers," *Biosensors Bioelectron.*, vol. 71, pp. 222–229, Sep. 2015.

[27] P. M. Kosaka *et al.*, "Detection of cancer biomarkers in serum using a hybrid mechanical and optoplasmonic nanosensor," *Nature Nanotechnol.*, vol. 9, pp. 1047–1053, Nov. 2014.

[28] I. F. Akyildiz and J. M. Jornet, "The Internet of nano-things," *IEEE Wireless Commun.*, vol. 17, no. 6, pp. 58–63, Dec. 2010.

[29] S. Kumar *et al.*, "Mobile health technology evaluation: The mHealth evidence workshop," *Amer. J. Preventive Med.*, vol. 45, no. 2, pp. 228–236, 2013.

[30] S. Balasubramaniam and J. Kangasharju, "Realizing the Internet of nano Things: Challenges, solutions, and applications," *Computer*, vol. 46, no. 2, pp. 62–68, Feb. 2013.

[31] R. A. Freitas, Jr., "Nanotechnology, nanomedicine and nanosurgery," *Int. J. Surg.*, vol. 3, no. 4, pp. 243–246, 2005.

[32] C. Rutherglen and P. Burke, "Nanoelectromagnetics: Circuit and electromagnetic properties of carbon nanotubes," *Small*, vol. 5, no. 8, pp. 884–906, 2009.

[33] O. B. Akan, H. Ramezani, T. Khan, N. A. Abbasi, and M. Kuscu, "Fundamentals of molecular information and communication science," *Proc. IEEE*, vol. 105, no. 2, pp. 306–318, Feb. 2017.

[34] R. Zhang, K. Yang, Q. H. Abbasi, K. A. Qaraqe, and A. Alomainy, "Analytical modelling of the effect of noise on the terahertz *in-vivo* communication channel for body-centric nano-networks," *Nano Commun. Netw.*, vol. 15, pp. 59–68, Mar. 2017.

[35] L. You, R. S. Cox, R. Weiss, and F. H. Arnold, "Programmed population control by cell–cell communication and regulated killing," *Nature*, vol. 428, pp. 868–871, Apr. 2004.

[36] S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold, and R. Weiss, "A synthetic multicellular system for programmed pattern formation," *Nature*, vol. 434, pp. 1130–1134, Apr. 2005.

[37] M.-T. Chen and R. Weiss, "Artificial cell-cell communication in yeast *saccharomyces cerevisiae* using signaling elements from Arabidopsis thaliana," *Nature Biotechnol.*, vol. 23, pp. 1551–1555, Nov. 2005.

[38] M. J. Doktycz and M. L. Simpson, "Nano-enabled synthetic biology," *Mol. Syst. Biol.*, vol. 3, no. 1, p. 125, 2007.

[39] Y. Sasaki *et al.*, "A nanosensory device fabricated on a liposome for detection of chemical signals," *Biotechnol. Bioeng.*, vol. 105, no. 1, pp. 37–43, 2010.

[40] M. Feiginov, R. Gonzalo, I. Maestrojun, O. Cojocari, M. Hoefle, and E. Limiti, *THz Electronics*. Hoboken, NJ, USA: Wiley, 2015, pp. 254–303.

[41] H. Elayan, R. M. Shubair, J. M. Jornet, and P. Johari, "Terahertz channel model and link budget analysis for intrabody nanoscale communication," *IEEE Trans. Nanobiosci.*, vol. 16, no. 6, pp. 491–503, Sep. 2017.

[42] R. Borkar, M. Bohr, and S. Jourdan, "Advancing Moore's law-the road to 14 nm," Intel Corp, Santa Clara, CA, USA, Tech. Rep., 2014, vol. 14.

[43] U. E. Avci, I. Ban, D. L. Kencke, and P. L. D. Chang, "Floating body cell (FBC) memory for 16-nm technology with low variation on thin silicon and 10-nm BOX," in *Proc. IEEE Int. SOI Conf.*, Oct. 2008, pp. 29–30.

[44] M. Nafari and J. M. Jornet, "Metallic plasmonic nano-antenna for wireless optical communication in intra-body nanonetworks," in *Proc. 10th EAI Int. Conf. Body Area Netw. (BodyNets)*, Sydney, NSW, Australia, 2015, pp. 287–293.

[45] E. Gul, B. Atakan, and O. B. Akan, "NanoNS: A nanoscale network simulator framework for molecular communications," *Nano Commun. Netw.*, vol. 1, no. 2, pp. 138–156, 2010.

[46] Q. H. Abbasi, M. Ur-Rehman, K. Qaraqe, and A. Alomainy, *Advances in Body-Centric Wireless Communication: Applications and State-of-the-Art*. Stevenage, U.K.: IET, 2016.

[47] I. Llatser, D. Demiray, A. Cabellos-Aparicio, D. T. Altilar, and E. Alarcón, "N3Sim: Simulation framework for diffusion-based molecular communication nanonetworks," *Simul. Model. Pract. Theory*, vol. 42, pp. 210–222, 2014.

[48] Á. Toth, D. Banky, and V. Grolmusz, "3-D brownian motion simulator for high-sensitivity nanobiotechnological applications," *IEEE Trans. Nanobiosci.*, vol. 10, no. 4, pp. 248–249, Dec. 2011.

[49] A. Akkaya, G. Genc, and T. Tugcu, "HLA based architecture for molecular communication simulation," *Simul. Model. Pract. Theory*, vol. 42, pp. 163–177, Mar. 2014.

[50] L. Felicetti, M. Femminella, and G. Reali, "A simulation tool for nanoscale biological networks," *Nano Commun. Netw.*, vol. 3, no. 1, pp. 2–18, 2012.

[51] G. Wei, P. Bogdan, and R. Marculescu, "Efficient modeling and simulation of bacteria-based nanonetworks with BNSim," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 12, pp. 868–878, Dec. 2013.

[52] Y. Jian *et al.*, "nanoNS3: A network simulator for bacterial nanonetworks based on molecular communication," *Nano Commun. Netw.*, vol. 12, pp. 1–11, Jun. 2017.

[53] G. Piro, L. A. Grieco, G. Boggia, and P. Camarda, "Nano-Sim: Simulating electromagnetic-based nanonetworks in the network simulator," in *Proc. 6th Int. ICST Conf. Simul. Tools Techn.*, 2013, pp. 203–210.

[54] G. Piro, K. Yang, G. Boggia, N. Chopra, L. A. Grieco, and A. Alomainy, "Terahertz communications in human tissues at the nanoscale for healthcare applications," *IEEE Trans. Nanotechnol.*, vol. 14, no. 3, pp. 404–406, May 2015.

[55] *NS-3 Platform*. Accessed: Apr. 15, 2018. [Online]. Available: https://www.nsnam.org/

[56] G. E. Santagati, T. Melodia, L. Galluccio, and S. Palazzo, "Distributed MAC and rate adaptation for ultrasonically networked implantable sensors," in *Proc. 10th Annu. IEEE Commun. Soc. Conf. Sensor, Mesh Ad Hoc Commun. Netw. (SECON)*, Jun. 2013, pp. 104–112.

[57] D. Laney, "3D data management: Controlling data volume, velocity and variety," META Group, Stamford, CT, USA, Res. Note, 2001, vol. 6, no. 70, p. 1.

[58] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *Int. J. Inf. Manage.*, vol. 35, no. 2, pp. 137–144, 2015.

[59] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile Netw. Appl.*, vol. 19, no. 2, pp. 171–209, 2014.

[60] R. Yu *et al.*, "Smart healthcare: Cloud-enabled body sensor networks," in *Proc. IEEE 14th Int. Conf. Wearable Implantable Body Sensor Netw. (BSN)*, 2017, pp. 99–102.

[61] I. D. Dinov, "Volume and value of big healthcare data," *J. Med. Statist. Inform.*, vol. 4, 2016, Art. no. 3.

[62] H. M. Krumholz, "Big data and new knowledge in medicine: The thinking, training, and tools needed for a learning health system," *Health Affairs*, vol. 33, no. 7, pp. 1163–1170, 2014.

[63] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "Health-CPS: Healthcare cyber-physical system assisted by cloud and big data," *IEEE Syst. J.*, vol. 11, no. 1, pp. 88–95, Mar. 2017.

[64] F. F. Costa, "Big data in biomedicine," *Drug Discovery Today*, vol. 19, no. 4, pp. 433–440, 2014.

[65] Q. Yao, Y. Tian, P.-F. Li, L.-L. Tian, Y.-M. Qian, and J.-S. Li, "Design and development of a medical big data processing system based on Hadoop," *J. Med. Syst.*, vol. 39, no. 3, p. 23, 2015.

[66] S. Rallapalli, R. R. Gondkar, and U. P. K. Ketavarapu, "Impact of processing and analyzing healthcare big data on cloud computing environment by implementing hadoop cluster," in *Proc. Int. Conf. Comput. Modelling Secur. (CMS)*, vol. 85, 2016, pp. 16–22.

[67] K. A. Guru *et al.*, "Current status of robot-assisted surgery in urology: A multi-national survey of 297 urologic surgeons," *Can. J. Urol.*, vol. 16, no. 4, pp. 4736–4741, 2009.

[68] K. D. Wise, "Wireless integrated microsystems: Wearable and implantable devices for improved health care," in *Proc. Int. Solid-State Sens., Actuators Microsyst. Conf. TRANSDUCERS*, Jun. 2009, pp. 1–8.

[69] A. Darwish and A. E. Hassanien, "Wearable and implantable wireless sensor network solutions for healthcare monitoring," *Sensors*, vol. 11, no. 6, pp. 5561–5595, 2011.

[70] S. Vaddiraju, I. Tomazos, D. J. Burgess, F. C. Jain, and F. Papadimitrakopoulos, "Emerging synergy between nanotechnology and implantable biosensors: A review," *Biosensors Bioelectron.*, vol. 25, no. 7, pp. 1553–1565, 2010.

[71] G. Matzeu, L. Florea, and D. Diamond, "Advances in wearable chemical sensor design for monitoring biological fluids," *Sens. Actuators B, Chem.*, vol. 211, pp. 403–418, May 2015.

[72] T. S. Perry. (2015). *A Temporary Tattoo That Senses Through Your Skin*. [Online]. Available: https://spectrum.ieee.org/biomedical/devices/a-temporary-tattoo-that-senses-through-your-skin

[73] S. L. Swisher *et al.*, "Impedance sensing device enables early detection of pressure ulcers *in vivo*," *Nature Commun.*, vol. 6, Mar. 2015, Art. no. 6575.

[74] S. N. Murphy *et al.*, "Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2)," *J. Amer. Med. Inform. Assoc.*, vol. 17, no. 2, pp. 124–130, 2010.

[75] S. Mohammed, D. Servos, and J. Fiaidhi, "HCX: A distributed OSGi based Web interaction system for sharing health records in the cloud," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. Intell. Agent Technol.*, vol. 3, Aug./Sep. 2010, pp. 102–107.

[76] H. Löhr, A.-R. Sadeghi, and M. Winandy, "Securing the e-health cloud," in *Proc. 1st ACM Int. Health Inform. Symp.*, 2010, pp. 220–229.

[77] A. Benharref and M. A. Serhani, "Novel cloud and SOA-based framework for E-health monitoring using wireless biosensors," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 1, pp. 46–55, Jan. 2014.

[78] R. Wooten, R. Klink, F. Sinek, Y. Bai, and M. Sharma, "Design and implementation of a secure healthcare social cloud system," in *Proc. 12th IEEE/ACM Int. Symp. Cluster, Cloud Grid Comput. (CCGRID)*, 2012, pp. 805–810.

[79] M. Stockdale and K. Royal, "Missing data as a validity threat for medical and healthcare education research: Problems and solutions," *Int. J. Healthcare*, vol. 2, no. 2, p. 67, Jun. 2016.

[80] B. J. Wells, K. M. Chagin, A. S. Nowacki, and M. W. Kattan, "Strategies for handling missing data in electronic health record derived data," *eGEMs*, vol. 1, no. 3, p. 1035, 2013, doi: 10.13063/2327-9214.1035.

[81] A. Gelman and J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge, U.K.: Cambridge Univ. Press, 2006, pp. 529–543.

[82] M. S. B. Sehgal, I. Gondal, and L. S. Dooley, "Collateral missing value imputation: A new robust missing value estimation algorithm for microarray data," *Bioinformatics*, vol. 21, no. 10, pp. 2417–2423, 2005.

[83] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," *IEEE Access*, vol. 5, pp. 8869–8879, 2017.

[84] Z. Yu, T. Li, S.-J. Horng, Y. Pan, H. Wang, and Y. Jing, "An iterative locally auto-weighted least squares method for microarray missing value estimation," *IEEE Trans. Nanobiosci.*, vol. 16, no. 1, pp. 21–33, Jan. 2017.

[85] I. R. White, P. Royston, and A. M. Wood, "Multiple imputation using chained equations: Issues and guidance for practice," *Stat. Med.*, vol. 30, no. 4, pp. 377–399, 2011.

[86] H. Xiong, G. Pandey, M. Steinbach, and V. Kumar, "Enhancing data analysis with noise removal," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 3, pp. 304–319, Mar. 2006.

[87] A. K. George and H. Singh, "DNA implementation of fuzzy inference engine: Towards DNA decision-making systems," *IEEE Trans. Nanobiosci.*, vol. 16, no. 8, pp. 773–782, Dec. 2017.

[88] F. Alam, R. Mehmood, I. Katib, N. N. Albogami, and A. Albeshri, "Data fusion and IoT for smart ubiquitous environments: A survey," *IEEE Access*, vol. 5, pp. 9533–9554, 2017.

[89] S. Misra and S. Chatterjee, "Social choice considerations in cloud-assisted WBAN architecture for post-disaster healthcare: Data aggregation and channelization," *Inf. Sci.*, vol. 284, pp. 95–117, Nov. 2014.

[90] M. Bagaa, A. Derhab, N. Lasla, A. Ouadjaout, and N. Badache, "Semi-structured and unstructured data aggregation scheduling in wireless sensor networks," in *Proc. INFOCOM*, Mar. 2012, pp. 2671–2675.

[91] X. Chen, X. Hu, and J. Zhu, "Minimum data aggregation time problem in wireless sensor networks," in *Mobile Ad-hoc and Sensor Networks* (Lecture Notes in Computer Science), vol. 3794, X. Jia, J. Wu, and Y. He, Eds. New York, NY, USA: Springer, 2005, pp. 133–142.

[92] S. C.-H. Huang, P.-J. Wan, C. T. Vu, Y. Li, and F. Yao, "Nearly constant approximation for data aggregation scheduling in wireless sensor networks," in *Proc. 26th IEEE Int. Conf. Comput. Commun. (INFOCOM)*, May 2007, pp. 366–372.

[93] S. K. A. Imon, A. Khan, and S. K. Das, "EFFECT: An energy efficient framework for data compression in tree-based wireless sensor networks," in *Proc. IEEE 15th Int. Symp. World Wireless, Mobile Multimedia Netw. (WoWMoM)*, Jun. 2014, pp. 1–9.
[94] M. Bagaa, M. Younis, A. Ouadjaout, and N. Badache, "Efficient multi-path data aggregation scheduling in wireless sensor networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2013, pp. 1560–1564.
[95] J. Li and H. Liu, "Challenges of feature selection for big data analytics," *IEEE Intell. Syst.*, vol. 32, no. 2, pp. 9–15, Mar./Apr. 2017.
[96] I. Yekkala, S. Dixit, and M. A. Jabbar, "Prediction of heart disease using ensemble learning and particle swarm optimization," in *Proc. Int. Conf. Smart Technol. Smart Nation (SmartTechCon)*, Aug. 2017, pp. 691–698.
[97] F. Bao, Y. Deng, Y. Zhao, J. Suo, and Q. Dai, "Bosco: Boosting corrections for genome-wide association studies with imbalanced samples," *IEEE Trans. Nanobiosci.*, vol. 16, no. 1, pp. 69–77, Jan. 2017.
[98] C. Plaisant, R. Mushlin, A. Snyder, J. Li, D. Heller, and B. Shneiderman, "LifeLines: Using visualization to enhance navigation and analysis of patient records," in *The Craft of Information Visualization*. Aalborg, Denmark: Elsevier, 2003, pp. 308–312.
[99] G. Falkman, "Information visualisation in clinical odontology: Multidimensional analysis and interactive data exploration," *Artif. Intell. Med.*, vol. 22, no. 2, pp. 133–158, 2001.
[100] T. Tanaka *et al.*, "Wearable health monitoring system by using fuzzy logic heart-rate extraction," in *Proc. World Automat. Congr. (WAC)*, Jun. 2012, pp. 1–4.
[101] T. Nawarycz, K. Pytel, W. Drygas, M. Gazicki-Lipman, and L. Ostrowska-Nawarycz, "A fuzzy logic approach to the evaluation of health risks associated with obesity," in *Proc. Federated Conf. Comput. Sci. Inf. Syst. (FedCSIS)*, Sep. 2013, pp. 231–234.
[102] M. Nii, Y. Kakiuchi, K. Takahama, T. Matsuda, Y. Matsumoto, and K. Maenaka, "Fuzzified neural network based human condition monitoring using a small flexible monitoring device," in *Proc. World Automat. Congr. (WAC)*, Aug. 2014, pp. 325–330.
[103] A. Benharref, M. A. Serhani, and A. R. Nujum, "Closing the loop from continuous M-health monitoring to fuzzy logic-based optimized recommendations," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.(EMBC)*, Aug. 2014, pp. 2698–2701.
[104] S. Huda, J. Yearwood, H. F. Jelinek, M. M. Hassan, G. Fortino, and M. Buckland, "A hybrid feature selection with ensemble classification for imbalanced healthcare data: A case study for brain tumor diagnosis," *IEEE Access*, vol. 4, pp. 9145–9154, 2016.
[105] D. Bardou, K. Zhang, and S. M. Ahmad, "Classification of breast cancer based on histology images using convolutional neural networks," *IEEE Access*, vol. 6, pp. 24680–24693, 2018.
[106] S. Enshaeifar *et al.*, "Health management and pattern analysis of daily living activities of people with dementia using in-home sensors and machine learning techniques," *PLoS ONE*, vol. 13, no. 5, p. e0195605, 2018.
[107] K.-C. Wong, "A short survey on data clustering algorithms," in *Proc. 2nd Int. Conf. Soft Comput. Mach. Intell. (ISCMI)*, Nov. 2015, pp. 64–68.
[108] X.-Z. Kong, J.-X. Liu, C.-H. Zheng, M.-X. Hou, and J. Wang, "Robust and efficient biomolecular clustering of tumor based on *p*-norm singular value decomposition," *IEEE Trans. Nanobiosci.*, vol. 16, no. 5, pp. 341–348, Jul. 2017.
[109] A. Roehrs, C. A. da Costa, R. da Rosa Righi, S. J. Rigo, and M. Wichman, "Toward a model for personal health records interoperability," *IEEE J. Biomed. Health Inform.*, to be published, doi: 10.1109/JBHI.2018.2836138.
[110] C.-M. Feng, Y.-L. Gao, J.-X. Liu, C.-H. Zheng, and J. Yu, "PCA based on graph laplacian regularization and P-norm for gene selection and clustering," *IEEE Trans. Nanobiosci.*, vol. 16, no. 4, pp. 257–265, Jun. 2017.
[111] T. N. Gia, M. Jiang, A.-M. Rahmani, T. Westerlund, P. Liljeberg, and H. Tenhunen, "Fog computing in healthcare Internet of Things: A case study on ECG feature extraction," in *Proc. IEEE Int. Conf. Comput. Inf. Technol., Ubiquitous Comput. Commun., Dependable, Autonomic Secure Comput., Pervasive Intell. Comput.*, Oct. 2015, pp. 356–363.
[112] A. Sallaberry, N. Pecheur, S. Bringay, M. Roche, and M. Teisseire, "Sequential patterns mining and gene sequence visualization to discover novelty from microarray data," *J. Biomed. Inform.*, vol. 44, no. 5, pp. 760–774, 2011.
[113] J. Park, K.-Y. Kim, and O. Kwon, "Comparison of machine learning algorithms to predict psychological wellness indices for ubiquitous healthcare system design," in *Proc. Int. Conf. Innov. Design Manuf. (ICIDM)*, Aug. 2014, pp. 263–269.
[114] B. Farahani, F. Firouzi, V. Chang, M. Badaroglu, N. Constant, and K. Mankodiya, "Towards fog-driven IoT eHealth: Promises and challenges of IoT in medicine and healthcare," *Future Generat. Comput. Syst.*, vol. 78, pp. 659–676, Jan. 2018.
[115] A. Z. Shirazi, S. M. Mazinani, and S. K. Eghbal, "Protocol stack for nano networks," in *Proc. Int. Symp. Comput., Consum. Control (IS3C)*, Jun. 2012, pp. 849–853.
[116] S. M. Erickson, J. Wolcott, J. M. Corrigan, and P. Aspden, *Patient Safety: Achieving a New Standard for Care*. Washington, DC, USA: National Academies, 2003.
[117] D. Bender and K. Sartipi, "HL7 FHIR: An agile and RESTful approach to healthcare information exchange," in *Proc. 26th IEEE Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2013, pp. 326–331.
[118] S. Kaisler, F. Armour, J. A. Espinosa, and W. Money, "Big data: Issues and challenges moving forward," in *Proc. 46th Hawaii Int. Conf. Syst. Sci.*, Jan. 2013, pp. 995–1004.

ALI RIZWAN received the bachelor's degree in applied and theoretical math in 2006 and the MBA-IT degree from Bahauddin Zakariya University, Pakistan, in 2008, and the M.Sc. degree in big data science from the Queen Mary University of London, London, U.K., in 2016. He is currently pursuing the Ph.D. degree with the University of Glasgow, Glasgow, U.K. He is conducting research, in the area of big data analytics for self-organizing wireless networks. His other area of interest is big data analytics for healthcare.

AHMED ZOHA received the M.Sc. degree in communication engineering from the Chalmers University of Technology, Sweden, in 2010, and the Ph.D. degree in electronic engineering from the University of Surrey, Surrey, U.K., in 2014, which was funded by the Engineering and Physical Sciences Research Council (EPSRC). He has contributed to a number of interdisciplinary and multimillion-funded international research projects in the domain of self-organizing networks, smart energy monitoring, and healthcare analytics. As a Lead Researcher on the project Quality-Aware Self-Organizing Networks, he was involved in the development and design of scalable predictive algorithms for real-time network intelligence, scheduling, and optimization using big data analytics technologies. As a Contributor to NHS funded healthcare project TIHM, he was involved in the development of prototype solution for chronic disease management driven by Internet of Things and artificial intelligence. His current research encompasses machine learning at the edge and representation learning using deep neural networks to drive intelligent reasoning. He was a recipient of two Best Paper Awards at the IEEE IS 2012 and the IEEE SSNIP 2013 as part of his contribution to EPSRC funded project Reshaping Energy Demands Using ICT.

RUI ZHANG received the bachelor's degree in electronic information engineering from Northwestern Polytechnical University, China, in 2011. She is currently pursuing the Ph.D. degree in electronics engineering with the School of Electronics Engineering and Computer Science, Queen Mary University of London, funded by the China Scholarship Council. Her research mainly focuses on THz communication, body-centric nano-communication, and THz characterization of skin.

WASIM AHMAD (S'06–M'12) received the Ph.D. degree from the University of Surrey, U.K., in 2011, with a research focus on machine learning, and sound modelling for interactive and multimedia applications. From 2011 to 2015, he was a Lecturer at the School of Engineering, University of Hull, U.K., and at the International Study Centre, University of Surrey. He is currently a Lecturer (Assistant Professor) in signal processing and machine learning at the School of Engineering, University of Glasgow, U.K. Since 2016, he has been a member of the Communication, Sensing and Imaging Group, University of Glasgow, where his research interests span applications of machine learning and data analytics in sound modeling, speech profiling, self-organizing networks, healthcare, nano-sensors and nano-networks, activity and gesture recognition for human–computer interface, and sign language.

KAMRAN ARSHAD (SM'15) is currently the Dean of Graduate Studies and Research and an Associate Professor in electrical engineering at Ajman University, United Arab Emirates. Prior to join Ajman University in 2016, he has been associated with the University of Greenwich, U.K., as a Senior Lecturer and a Program Director of MSc Wireless Mobile Communications Systems Engineering. He led a number of locally and internationally funded research projects encompassing the areas of cognitive radio, LTE/LTE-advanced, 5G, and cognitive machine-to-machine communications. He has contributed to several European and international large-scale research projects. He has 100+ technical peer-reviewed papers in top quality journals and international conferences, received three best paper awards, one best research and development track award, and chaired technical sessions in several leading international conferences. He is a Senior Fellow of the UK Higher Education Academy. He is an Associate Editor of *EURASIP Journal on Wireless Communications and Networking*. He delivered keynote/guest/invited talks at several international forums/conferences/workshops and became a panellist in the IEEE PIMRC 2011 Workshop on Cognitive Radio.

NAJAH ABU ALI received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Jordan and the Ph.D. degree from the Department of Electrical and Computer Engineering, Queen's University in Kingston, Canada. She is currently an Associate Professor at the Faculty of Information Technology, United Arab Emirates University. Her general research interests include modeling wireless communications, resource management in wireless networks, and reducing the energy requirements in wireless sensor networks.

More recently, she has strengthened her focus on the Internet of Things, particularly at the nano-scale communications level, in addition to vehicle-to-vehicle networking. She has also delivered various seminar and tutorials at both esteemed institutions and flagship gatherings. She has also been awarded several research fund grants, particularly from the Emirati NRF/UAEU funds and the Qatar National Research Foundation.

AKRAM ALOMAINY (S'04–M'07–SM'13) received the M.Eng. degree in communication engineering and the Ph.D. degree in electrical and electronic engineering (specialized in antennas and radio propagation) from the Queen Mary University of London (QMUL), U.K., in 2003 and 2007, respectively. He joined the School of Electronic Engineering and Computer Science, QMUL, in 2007, where he is an Associate Professor (Senior Lecturer) with the Antennas and Electromagnetics Research Group. His current research interests include small and compact antennas for wireless body area networks, radio propagation characterization and modelling, antenna interactions with human body, computational electromagnetic, advanced antenna enhancement techniques for mobile and personal wireless communications, and advanced algorithm for smart and intelligent antenna and cognitive radio system. He has managed to secure various research projects funded by research councils, charities, and industrial partners on projects ranging from fundamental electromagnetic to wearable technologies. He is a member of the Institute of Bioengineering and the Centre for Intelligent Sensing, QMUL. He is the Lead of wearable creativity research at the Queen Mary University of London and has been invited to participate at the Wearable Technology Show 2015, Innovate UK 2015, and also in the recent Wearable Challenge organized by Innovate UK IC Tomorrow, as a leading challenge partner to support SMEs and industrial innovation. He has authored and co-authored a book, five book chapters, and over 150 technical papers (2800+ citations and H-index 25) in leading journals and peer-reviewed conferences.

Since 2011, he has been serving as an Associate Professor with the Department of Electrical and Electronics Engineering, Altınbaş University. His current research interests are in the area of wireless communication, digital signal processing, channel coding, channel estimation and equalization techniques in particular multilevel coding, turbo coding, and equalization for wireless communication systems.

Dr. Alomainy is a member of the IET, a fellow of the Higher Education Academy (U.K.), and also a College Member for Engineering and Physical Sciences Research (EPSRC, U.K.) and its ICT prioritization panels. He was a recipient of the Isambard Brunel Kingdom Award, in 2011, for being an Outstanding Young Science and Engineering Communicator. He was selected to deliver a TEDx talk about the science of electromagnetic and also participated in many public engagement initiatives and festivals. He is also a reviewer for many funding agencies around the world, including Expert Swiss National Science Foundation Research, the Engineering and Physical Sciences Research Council, U.K., and the Medical Research Council, U.K. He is an Elected Member of UK URSI (International Union of Radio Science) panel to represent the UK interests of URSI Commission B (2014–2017).

MUHAMMAD ALI IMRAN (M'03–SM'12) received the M.Sc. (Hons.) and Ph.D. degrees from Imperial College London, U.K., in 2002 and 2007, respectively. He is the Vice Dean Glasgow College UESTC and a Professor of communication systems with the School of Engineering, University of Glasgow. He is an Affiliate Professor at the University of Oklahoma, USA, and a Visiting Professor at the 5G Innovation Centre, University of Surrey, U.K. He has over 18 years of combined academic and industry experience, working primarily in the research areas of cellular communication systems. He has been awarded 15 patents, has authored/co-authored over 300 journal and conference publications, and has been principal/co-principal investigator on over £6 million in sponsored research grants and contracts. He has supervised 30+ successful Ph.D. graduates.

He has an award of excellence in recognition of his academic achievements, conferred by the President of Pakistan. He was also awarded the IEEE Comsoc's Fred Ellersick Award 2014, the FEPS Learning and Teaching Award 2014, and the Sentinel of Science Award 2016. He was twice nominated for the Tony Jean's Inspirational Teaching Award. He is a shortlisted finalist for The Wharton-QS Stars Awards 2014, the QS Stars Reimagine Education Award 2016 for innovative teaching, and VC's Learning and Teaching Award from the University of Surrey. He is a Senior Fellow of the Higher Education Academy, U.K. He is the editor/co-editor of two books: *Access, Fronthaul and Backhaul Networks for 5G and Beyond* (IET, ISBN 9781785612138) and *Energy Management in Wireless Cellular and Ad-hoc Networks* (Springer, ISBN 9783319275666).

QAMMER H. ABBASI (SM'16) received the B.Sc. and M.Sc. degrees (Hons.) in electronics and telecommunication engineering from the University of Engineering and Technology (UET), Lahore, Pakistan, and the Ph.D. degree in electronic and electrical engineering from the Queen Mary University of London (QMUL), U.K., in 2012. In 2012, he was a Post-Doctoral Research Assistant with the Antenna and Electromagnetics Group, QMUL. From 2012 to 2013, he was an International Young Scientist with the National Science Foundation China and an Assistant Professor with UET. From 2013 to 2017, he was with the Centre for Remote Healthcare Technology, Wireless Research Group, Department of Electrical and Computer Engineering, Texas A &M University, initially as an Assistant Research Scientist and later was promoted to Associate Research Scientist and Visiting Lecturer, where he received multiple Qatar national research foundation grants (worth U.S. $3 million). He has been mentoring several undergraduate, graduate students, and postdoctoral positions. He is currently a Lecturer (Assistant Professor) with the School of Engineering, University of Glasgow, U.K., a Visiting Research Fellow with QMUL, and a Visiting Associate Research Scientist with TAMUQ. He has contributed to a patent, over 150 leading international technical journal and peer reviewed conference papers, and five books. The most downloaded papers in the IEEE Transactions on Terahertz Science and Technology, Media Coverage by Analog IC Tips, Microwaves & RF Newsletters, and Vertical News. His research interests include nano communication, RF design and radio propagation, biomedical applications of millimeter wave and terahertz communication, antenna interaction with human body, wearables and implants, nano-scale agri-tech, body centric wireless communication issues, wireless body sensor networks, non-invasive health care solutions, cognitive and cooperative network, and multiple-input-multiple-output systems. He has been a member of the technical program committees of several IEEE flagship conferences. He was a member of IET. He received several recognitions for his research, including the University Research Excellence Award from TAMUQ in two consecutive years, and a Reward for excellence from the University of Glasgow, and an UK Exceptional Talent Endorsement by the Royal Academy of Engineering. He is an Associate Editor of the IEEE Access and the IEEE Journal ofElectromagnetics, RF and Microwaves in Medicine and Biology, and a guest editor for numerous special issues in top notch journals. He has been a technical reviewer for several IEEE and top-notch journals. He contributed in organizing several IEEE conferences, workshop, and special sessions, in addition to the European School of Antenna Course.