Biophysical Reviews (2019) 11:83–87
https://doi.org/10.1007/s12551-018-0495-3

REVIEW

A review of big data applications of physiological signal data

Christina Orphanidou [1]

Received: 11 November 2018 / Accepted: 18 December 2018 / Published online: 9 January 2019

Abstract

The proliferation of smart physiological signal monitoring sensors, combined with the advancement of telemetry and intelligent communication systems, has led to an explosion in healthcare data in the past few years. Additionally, access to cheaper and more effective power and storage mechanisms has significantly increased the availability of healthcare data for the development of big data applications. Big data applications in healthcare are concerned with the analysis of datasets which are too big, too fast, and too complex for healthcare providers to process and interpret with existing tools. The driver for the development of such systems is the continuing effort in making healthcare services more efficient and sustainable. In this paper, we provide a review of current big data applications which utilize physiological waveforms or derived measurements in order to provide medical decision support, often in real time, in the clinical and home environment. We focus mainly on systems developed for continuous patient monitoring in critical care and discuss the challenges that need to be overcome such that these systems can be incorporated into clinical practice. Once these challenges are overcome, big data systems have the potential to transform healthcare management in the hospital of the future.

Keywords Physiological signal data · Smart sensors · Healthcare data · Medical decision support · Big data

Introduction

The proliferation of smart physiological signal monitoring sensors, combined with the advancement of telemetry and intelligent communication systems, have led to an explosion in healthcare data in the past few years. While in the past, continuous physiological data generated from wearable sensors have not been typically stored for long periods of time, in the recent past, access to cheaper and more effective power and storage mechanisms has significantly increased the availability of such data. Healthcare data are complex, diverse, and rich in context, creating the grounds for the development of big data applications for health (Roski et al. 2014; Krumholz 2014; Orphanidou and Wong 2017). Big data applications in healthcare are concerned with the analysis of datasets which are too big, too fast, and too complex for healthcare providers to process and interpret with existing tools (Andreu-Perez et al. 2015). The driver for the development of such systems is the continuing effort in making healthcare services more efficient and sustainable in the face of a continuously aging world population. In order to tackle these socioeconomic challenges, scientists are working towards the development of more intelligent healthcare systems, aiming to transform healthcare management by cutting down costs while improving patient outcomes.

Big data and physiological signals

In 2012, healthcare data worldwide amounted to approximately 500 petabytes and by 2020 the amount is projected to be 25,000 petabytes (Roski et al. 2014). The large quantities of physiological data, generated in hospital and at home, have presented the opportunity to develop data-driven approaches for improving diagnoses, delivering best practice, and improving patient outcomes. An important challenge is how the available data can be interpreted in order to provide better and faster (sometimes real-time) decision making and to improve the standard of patient care and, consequently, patient health outcomes (Orphanidou and Wong 2017). Via the use of machine learning approaches, aggregate

This article is part of a Special Issue on 'Big Data' edited by Joshua WK Ho and Eleni Giannoulatou.

✉ Christina Orphanidou
c.orphanidou@gmail.com

[1] Oxygen Research Ltd, 8 Vassileos Constantinou Street, 3075 Limassol, Cyprus

physiological data may reveal new information leading to improved ways of medical diagnosis and may facilitate dynamic online monitoring in the clinical environment or at home. Medical decision support systems based on Big Data/Deep Learning approaches offer the potential of quick and effective decision making at a reduced cost while circumventing the confirmation bias of a clinical expert (Krumholz 2014). However, healthcare datasets, often comprising of structured (e.g., laboratory results), semi-structured (e.g., sensor data), and unstructured (e.g., handwritten patient notes) data (Palanisamy and Thirunavukarasu 2017), are often multidimensional, noisy, and dynamically changing. This complexity makes interpretation very difficult and requires stringent validation. In addition, incorporating these systems into clinical practice will require not only new data sources, but also new training and the establishment of new escalation strategies (Krumholz 2014).

In the hospital, the electronic patient health record (EHR) is a source of big data containing information regarding socio-demographics, pre-existing and current medical conditions, genetics, biochemical tests, and treatments (Andreu-Perez et al. 2015). While the information contained is rich, this vast and information-rich database cannot be processed by a human in a high throughput manner for medical decision making and is thus largely underused. Data analytics systems, often utilizing machine learning, are required which can help clinical staff organize the data, identify patterns, interpret results, and set thresholds for actions. Natural language processing is also a big data–powered tool which can significantly contribute to the analysis of these records for semantic context. Examples of big data analytics for new knowledge generation, improved clinical care, and streamlined public health surveillance have already been put into action in hospitals (Andreu-Perez et al. 2015). Leveraging big data and building machine learning systems providing a continuous learning mechanism with real-time knowledge production can lead to the development of intelligent systems based on the P4 medicine principle: predictive, preventive, personalized, and participatory (Hood and Flores 2012; Leff and Yang 2015). The HVITAL (Hospital surveillance, monitoring and Alert) system, functioning in the Sao Paolo Hospital since 2012 (Almeida 2016), is an example of a novel analytics platform which leverages all the big data that is stored in the EHRs in real time and applies advanced and fast analytics on top of this information in order to provide decision support to the clinicians. The system calculates and produces over 600 key performance indicators (KPIs) which are not only clinically related but also provide management-level insight for improving hospital operations. By being able to search through millions of patient records quickly, the HVITAL system provides real-time insight to the hospital staff through desktop dashboards and goes even further by constantly monitoring a series of physiological signs from every patient throughout their stay and applying intelligent predictive algorithms for the early identification of patients-at-risk, presented as yellow and red alerts and leading to interventions. An approach often utilized for providing such alerts is that of novelty detection (Pimentel et al. 2014) which is a machine learning approach used in situations where a lot of data is available of the "normal" state and very few of the "abnormal." Often modeled as a one-class classification approach, novelty detection techniques have been applied in the clinical context for providing early warning of deterioration in the emergency department (Wilson et al. 2016), the ICU (Ghassemi et al. 2015), and for post-operative patients (Clifton et al. 2014; Pimentel et al. 2013). Gaussian processes have also been extensively used for dealing with noisy and unreliable physiological data (Dunitz et al. 2015) including in the ICU (Pimentel et al. 2016). As reviewed in Leff and Yang (2015), in addition to applications in the early warning of deterioration in continuous monitoring, medical decision support systems have also been proposed for many other applications including to assess and improve protocol adherence (Klann et al. 2013), for medication reminders (Nair et al. 2010), to improve screening (Wagholikar et al. 2012), and to predict hospital readmission (Futoma et al. 2015).

Applications in continuous patient monitoring

The overarching principle of big data analytics applications in patient monitoring systems is the use of either continuous physiological waveform data, derived vital sign information from different and disparate sources, or combinations of the two, in order to provide early warning of deterioration. Most patient monitoring systems used until now have relied on a single source of information (such as a single vital sign) and have resulted to the phenomenon of "alarm fatigue" whereby the clinical staff becomes desensitized to and ultimately ignore alerts from the monitoring systems (Orphanidou et al. 2015). Big data applications have created the opportunity to develop improved and more comprehensive early warning mechanisms by fusing different sources of information such as multiple simultaneously collected vital signs, as well other clinical information (of different phenotypes), the aggregation of which provide a more comprehensive overview of a patient's health status. In addition to improving the accuracy of alerting, this has created the opportunity of studying the interactions and correlations among multimodal clinical time series data or waveforms (Belle et al. 2015) for gaining clinical knowledge.

In the context of in-hospital monitoring, a vast amount of physiological data in short periods of time is produced in

intensive care units (ICU) where patients are continuously wired into multiparameter monitors which record waveforms and provide periodic measurements of vital signs such as heart rate (HR), respiratory rate (RR), peripheral arterial oxygen saturation (SpO2), arterial blood pressure (ABP), and temperature (T). In addition to data from bedside monitors, data from EHRs, and medication pumps, as well as ventilation data, can be exploited for the development of big data applications for continuous patient monitoring and medical decision support. The potential for developing such systems in the ICU environment has been widely recognized and a number of studies have appeared in the literature proposing big data approaches. The main thesis of most studies is that the complexities characterizing critical illness (often with comorbidities) cannot be adequately addressed using traditional approaches (e.g., single drug interventions), and perhaps the use of big data–driven data fusion approaches is more suitable (Johnson et al. 2016; Sanchez-Pinto et al. 2018). Examples of big data applications in critical care include a scalable infrastructure for developing a patient care management system which combines static and continuous data monitored from critically ill patients in the ICU for data mining and alerting medical staff of critical events in real time (Han et al. 2006), a system to predict increased intracranial pressure in the ICU (Güiza et al. 2013), as well as a system developed for a neonatal ICU which utilized streaming data from EEG monitors, infusion pumps, cerebral oxygenation monitors, etc. to provide medical decision support (Bressan et al. 2012). One of the first studies involving ICU patients utilized data from 250,000 hospital admissions and built predictive models based on logistic regression to estimate ICU transfer, cardiac arrest, or death in ward patients (Churpek et al. 2014). A follow-up study by the same authors replaced logistic regression with more sophisticated machine learning approaches such as random forests and gradient boosting techniques to improve the predictive performance for early detection of deterioration (Churpek et al. 2016). Mortality rate was also accurately predicted by Joshi and Szolovits (2012) using data from the Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) database (Johnson et al. 2016) on Physionet (Goldberger et al. 2000). To achieve an AUC of 0.91, the authors employed clustering techniques, dividing patients into organ-specific patient state. Using the same database, Pirracchio et al. (2014) compared the performance of 12 different mortality prediction algorithms (parametric and non-parametric) to accurately predict mortality and concluded that an ensemble technique using a weighted output of several machine learning algorithms (the *Superlearner*) gave the best predictive performance with an AUB of 0.88.

Databases, such as the MIMIC III (Multiparameter Intelligent Monitoring in Intensive Care) (Johnson et al. 2016), containing physiologic signals and vital signs time series captured from patient monitors, as well as comprehensive clinical data obtained from hospital medical information systems, for tens of thousands of intensive care unit (ICU) patients, have created the opportunity for researchers to work towards the development of big data applications for knowledge discovery. The MIMIC database has thus been used for applications such as finding similarities among patients within the selected cohorts (Lee and Mark 2010), or to develop systems that fuse multiple waveform information to develop early predictors of cardiovascular instability in patients (Sun et al. 2010). A combination of multiple waveform information available in the MIMIC II database was also utilized to develop early detection of hemodynamic instability in patients (Cao et al. 2008). Other big data applications using physiological signals include a system estimating cardiac output using pulse contour analysis (Attin et al. 2015), a system detecting hypovolemia using photoplethysmography data (Roederer et al. 2015), and a system for predicting hyperlactemia using combined physiological data (Dunitz et al. 2015).

Outside the clinical environment, there is a rise of consumer-grade wearable sensors that provides the opportunity for the development of big data applications using physiological sensor data. These sensors make traditional physiological monitoring devices smaller, cheaper, and more portable. Systems have been proposed to combine ECG parameters from telemetry with demographic information including medical history, ejection fraction, laboratory values, and medications to provide an in-hospital early detection system for cardiac arrest (Sun et al. 2010). However, similar to clinical applications, combining information simultaneously collected from multiple portable devices can become challenging. The variety of fixed as well as mobile sensors available for data mining in the healthcare sector and how such data can be leveraged for developing patient care technologies are surveyed by Sow et al. (2013).

Challenges

Despite the big momentum experienced by the application of big data in healthcare, before big data systems can be applied to real-life clinical problems, a number of challenges need to be overcome. Technical challenges that need to be overcome relate to data quality and analysis and while not unique to healthcare, the acquisition of data from human patients brings additional challenges that do not occur in many other fields (Orphanidou and Wong 2017). One challenge is the analysis of datasets with missing or corrupted data which is a frequent occurrence when dealing with data collected from humans and which, if left unrecognized, may skew or corrupt the analysis and lead to inaccurate decision making. In the context of data collected from wearable sensors, these issues may simply result from an incorrectly attached sensor or movement. To reduce corrupt data, sensors need to be improved in terms of

their ergonomics and attachment strategies. Non-contact vital sign monitoring (Tarassenko et al. 2014) may provide a solution to this problem. If corrupt data has been collected, and cannot be discarded, techniques such as data imputation (Tarassenko et al. 2006) and quality assessment (Orphanidou 2018) may provide a solution. Another challenge in the utilization of physiological data in big data healthcare applications is the issue of integrating heterogeneous data sources. While data fusion, i.e., the processing of information from several different source of data, has been widely used for providing a more holistic view of the problem and for corroborating measurements when quality issues exist, the process of integrating data sources in big data applications is very difficult. Feature selection techniques (reviewed by Saeys et al. 2007) and data fusion techniques such as Kalman filters (Durrant-Whyte and Henderson 2008) and Multiple Kernel Learning (MKL) (Hu et al. 2009) have been proposed in the literature as ways of enabling the integration of multiple data sources simultaneously, but more work is needed such that maximum value can be extracted from simultaneously collected physiological datasets. Techniques based on Multi-Task Gaussian Processes (MTGP) have also been proposed for dealing with datasets which are noisy, incomplete, sparse, heterogeneous, and unevenly-sampled (Ghassemi et al. 2015; Dürichen et al. 2015). In addition to this, sophisticated and low-cost storage and processing mechanisms need to be put into place, which facilitate rapid data pull and commits based on analytics demands (Belle et al. 2015). A number of governance challenges additionally need to be overcome, such as issues relating to the establishment of appropriate data protocols and standards and data privacy issues (Andreu-Perez et al. 2015). Data compartmentalization is another issue which concerns the fact that data collected from multiple devices are often stored in different databases which need to be linked for the development of holistic big data applications (Johnson et al. 2016). A variety of legal and ethical concerns need to be addressed for this linkage to be done effectively and this often includes the need for co-operation between competitive system manufacturers (Orphanidou and Wong 2017).

In addition to analysis and quality-related challenges, the adoption of big data techniques in real-life clinical applications poses a number of operational challenges. It requires the development of new clinical practices and ways of thinking. Big data approaches relying on machine learning and data mining require the acceptance of searching for patterns without knowing what might emerge (Krumholz 2014). This approach differs greatly from the classic scientific approach of starting with a specific research question and requires stringent methods for validating findings to ensure credibility and statistical significance.

Conclusion

For big data applications to find their route to clinical practice, the biggest challenge that lies ahead is the need for the development of new skills, new ways of thinking, and new ways of inferring knowledge, as well as new reaction and escalation mechanisms within the clinical setting. Research activity in the area of big data applications in healthcare should therefore not only be limited to the technical implementation of such systems but also to the appropriate mechanisms for integration into clinical practice. Once this challenge is overcome, big data applications have the potential to have a huge impact not only on clinical research but also on the health outcomes of the general population.

Compliance with ethical standards

Conflict of interest Christina Orphanidou declares that she has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by the author.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Almeida JP (2016) A disruptive big data approach to leverage the efficiency in management and clinical decision support in a hospital. Porto Biomed J 1(1):40–42

Andreu-Perez J, Poon CCY, Merrifield RD, Wong STC, Yang GZ (2015) Big data for health. IEEE J Biomed Health Inf 19(4):1193–1208

Attin M, Feld G, Lemus H et al (2015) Electrocardiogram characteristics prior to in-hospital cardiac arrest. J Clin Monit Comput 29(3):385–392

Belle A, Thiagarajan A, Reza SM et al (2015) Big data analytics in healthcare. Biomed Res Int 370194:16

Bressan N, James A, McGregor C (2012) Trends and opportunities for integrated real time neonatal clinical decision support. Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI '12), pp 687–690

Cao H, Eshelman L, Chbat N, Nielsen L, Gross B, Saeed M(2008) Predicting ICU hemodynamic instability using continuous multiparameter trends, EMBC '08, pp 3803–3806

Churpek MM, Yuen TC, Winslow C et al (2014) Multicenter development and validation of a risk stratification tool for ward patients. Am J Respir Crit Care Med 190(6):649–655

Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP (2016) Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. Crit Care Med 44(2):368–374

Clifton L, Clifton DA, Pimentel MAF, Watkinson PJ, Tarassenko L (2014) Predictive monitoring of mobile patients by combining clinical observations with data from wearable sensors. IEEE J Biomed Health Inf 18(3):722–730

Dunitz M, Verghese G, Heldt T (2015) Predicting hyperlactatemia in the MIMIC II database. Proc. EMBC '15, pp 985–988

Dürichen R, Pimentel MAF, Clifton L, Schweikard A, Clifton DA (2015) Multitask Gaussian processes for multivariate physiological time-series analysis. IEEE Trans Biomed Eng 62(1):314–322

Durrant-Whyte H, Henderson TC (2008). Multisensor data fusion. Springer Handbook of Robotics, pp 585–610

Futoma J, Morris J, Lucas J (2015) A comparison of models for predicting early hospital readmissions. J Biomed Inform 56: 229–238

Ghassemi M, Pimentel MA, Naumann T, et al. (2015) A multivariate timeseries modeling approach to severity of illness assessment and forecasting in ICU with sparse. Heterogeneous clinical data. Proc Conf AAAI Artif Intell 2015:446–453

Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE (2000) PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 101(23):e215–e220

Güiza F, Depreitere B, Piper I, Van den Berghe G, Meyfroidt G (2013) Novel methods to predict increased intracranial pressure during intensive care and long-term neurologic outcome after traumatic brain injury: development and validation in a multicenter dataset. Crit Care 41(2):554–564

Han H, Ryoo HC, Patrick H (2006) An infrastructure of stream data mining, fusion and management for monitored patients. In Proceedings of the 19th IEEE International Symposium on Computer-Based Medical Systems (CBMS '06), pp 461–468

Hood L, Flores M (2012) A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized and participatory. New Biotechnol 29(6):613–624

Hu M, Chen Y, Kwok JT (2009) Building sparse multiple-kernel SVM classifiers. IEEE Trans Neural Netw 20(5):827–839

Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, Mark RG (2016) MIMIC-III, a freely accessible critical care database. Scientific Data

Joshi R, Szolovits P (2012) Prognostic physiology: modeling patient severity in intensive care units using radial domain folding. Paper presented at: AMIA Annual Symposium Proceedings

Klann JG, Anand V, Downs SM (2013) Patient tailored prioritization for a pediatric care decision support system through machine learning. J Am Med Inform Assoc 20(e2):e267–e274

Krumholz HM (2014) Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. Health Aff 33(7):1163–1170

Lee J, Mark RG (2010) A hypotensive episode predictor for intensive care based on heart rate and blood pressure time series. IEEE Comput Cardiol 81–84

Leff DR, Yang G-Z (2015) Big data for precision medicine, engineering. 1(3):277–279

Nair BG, Newman SF, Peterson GN, Wu WY, Schwid HA (2010) Feedback mechanisms including real-time electronic alerts to achieve near 100% timely prophylactic antibiotic administration in surgical cases. Anesth Analg 111(5):1293–1300

Orphanidou C (2018) Signal quality assessment in physiological monitoring: state of the art and practical considerations. Springer, Cham

Orphanidou C, Wong D (2017) Machine learning models for multidimensional clinical data. In: Khan SU, Zomaya AY, Assad A (eds) Handbook of large-scale distributed computing in smart healthcare, scalable computing and communications. Springer, Cham, pp 177–216

Orphanidou C, Bonnici T, Charlton P, Clifton D, Valance D, Tarassenko L (2015) Signal-quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring. IEEE J Biomed Health Inform 19(3):832–838

Palanisamy V, Thirunavukarasu R (2017) Implications of big data analytics in developing healthcare frameworks – a review. J King Saud Univ Comput Inf Sci

Pimentel MAF, Clifton DA, Clifton L, Watkinson PJ, Tarassenko L (2013) Modelling physiological deterioration in post-operative patient vital-sign data. Med Biol Eng Comput 51:869–877

Pimentel MAF, Clifton DA, Clifton L, Tarassenko L (2014) A review of novelty detection. Signal Process 99:215–249

Pimentel MAF et al (2016) Outcome prediction for patients with traumatic brain injury with dynamic features from intracranial pressure and arterial blood pressure signals: a Gaussian process approach. Acta Neurochir Suppl 122:85–91

Pirracchio R, Petersen ML, Carone M, Rigon MR, Chevret S, van der Laan MJ (2014) Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. Lancet Respir Med 3(1):42–52

Roederer A, Weimer J, DiMartino J, Gutsche J, Lee I (2015) Robust monitoring of hypovolemia in intensive care patients using photoplethysmogram signals. Proc. EMBC '15, pp 1504–1507

Roski J, Bo-Linn GW, Andrews TA (2014) Creating value in health care through big data: opportunities and policy implications. Health Aff 33(7):1115–1122

Saeys Y, Inza I, Larrañaga P (2007) A review of feature selection techniques in bioinformatics. Bioinformatics 23(19):2507–2517

Sanchez-Pinto LN, Luo Y, Churpek MM (2018) Big data and data science in critical care. Chest 154(5):1239–1248

Sow D, Turaga DS, Schmidt M (2013) Mining of sensor data in healthcare: a survey. Managing and Mining Sensor Data, pp 459–504

Sun J, Sow D, Hu J, Ebadollahi S(2010) A system for mining temporal physiological data streams for advanced prognostic decision support, in Proceedings of the 10th IEEE International Conference on Data Mining (ICDM' 10), pp 1061–1066

Tarassenko L, Hann A, Young D (2006) Integrated monitoring and analysis for early warning of patient deterioration. Br J Anaesth 97(1): 64–68

Tarassenko L, Villarroel M, Guazzi A, Jorge J, Clifton DA, Pugh C (2014) Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. Physiol Meas 35(5):807–831

Wagholikar KB et al (2012) Clinical decision support with automated text processing for cervical cancer screening. J Am Med Inform Assoc 19(5):833–839

Wilson SJ, Wong D, Pullinger RM, Way R, Clifton DA, Tarassenko L (2016) Analysis of a data-fusion system for continuous vital sign monitoring in an emergency department. Eur J Emerg Med 23(1):28–32