Deep Science Publishing
https://doi.org/10.70593/978-81-988918-5-3_6

Chapter 6: Utilizing big data analytics for real-time patient monitoring and risk prediction

6.1. Introduction

Some of the most important tasks required of healthcare systems is to provide a high level of acute care services and to ensure that the workforce is productive. Sickness and injury in the workforce not only lowers the productivity of individual workers, but also causes further resource and financial losses for businesses. Hence, it is important for enterprises to reduce any adverse effects of illness and injury on workforce productivity through effective health monitoring and promotion interventions. Advances in computing, communication, and sensing technologies enable healthcare systems to tap into the various streams of data available from workplaces and seamlessly monitor the health of occupational settings and of individual employees. The capability for continuous real-time and data-driven monitoring of health is now possible with devices such as smartwatches, smartphone apps, special sensors for heart rate, blood sugar levels, etc. Transfer of the enormous amounts of collected data to processing units allows for the application of big data analytics techniques to derive information, knowledge, and insights from the data (Dilsizian & Siegel, 2014; Krittanawong et al., 2017; Lee et al., 2017).

The field of healthcare that deals with timeline-based data for the monitoring and prediction of health/disease events is known as health chronemics. Examples of health and disease conditions with recognized diurnal or circadian timelines include blood glucose levels of diabetics, heart rate variability, heart arrhythmias, blood pressure levels, sleep, eating, exercise, etc. Some of these features have been recognized to follow patterns such as deviance from their normal levels, breaches beyond predictive biosignal thresholds, irregularity in periodicity, etc. These techniques have been successfully used for the remote monitoring and early detection of conditions such as cardiac events, sleep apnea, diabetes, metabolic syndrome, hypertension, and mental disorders.

Concerns with the flow of services, particularly when patients do not follow a standard pathway, are becoming more relevant as throughput becomes a concern. Informatics enables the predictive ability of work systems to ensure that capacity is always available, but also provides the mechanism to transport patients in a most efficient manner. Adding time-based information to formal multi-dimensional models enables throughput, classification, and throughput financial planning tools to yield policy planning in patient management systems. The excitement of a whole industry developing around health systems reflects the impact of data and information on real-time patient management and risk assessment (Ristevski & Chen, 2018; Mehta et al., 2019).

Fig 6.1: Big Data Analytics for Real-Time Patient Monitoring and Risk Prediction

6.1.1. Background and Significance

Advances in information technologies have been used for patient monitoring and risk assessment over the last four decades. Various sensors, complex algorithms, and multimedia techniques are now available for clinicians, who require a sophisticated approach to real-time patient assessment that captures the whole of individual patient complexity over the course of their illness. Informatics tools enable early detection of disease, allow at-risk populations to be targeted, and support ongoing assessment within the management of chronic diseases. Real-time patient monitoring has enabled more effective patient management, particularly when focusing on risk, using limited resources. Health managers need the ability to measure progress over time and to act early in terms of access to increased care for abnormal patients. Formal systems are required to absorb data from clinical flow and to assess need on a large scale.

Health resource managers are increasingly focused on business requirements, using informatics tools to ensure limited resources are allocated according to patient need and that risk is contained and visible.

6.2. Overview of Big Data in Healthcare

Big Data generally refers to the datasets which due to their volume, velocity and variety are usually difficult to be handled with traditional computational resources. Recently, there is a growing interest in usage of big data analytics in health and healthcare for real-time patient monitoring and risk prediction. The aim of this study is providing an overview of existing literature which utilizes big data analytics to effectively deal with real-time patient monitoring and risk prediction problems. This study can be beneficial to researchers as to the healthcare professionals who want to adopt big data analytics in their daily practices. A qualitative research approach is utilized to solve the research design. Firstly, the keywords are identified and then keyword searching is applied which produces 890 publications. While filtering the theme, the final sample containing 107 publications is acquired. Content analysis is applied to the 107 publications to produce ten categories under four relevant themes. We gathered actually relevant publications based on title, abstract and keywords ranked according to relevance score, and further analyzed the appropriate papers.

The analysis of these papers revealed that use of supervised learning algorithms in diagnosis and prognosis is presented as a significant characteristic of the domain. Additionally, performance assessment of different algorithms in terms of evaluating metrics has been another distinguishing feature of the contributions. Finally, unique data sources are specifically utilized with dedicated risk prediction and diagnosis studies for specific diseases and symptoms. Therefore, it seems that existing works generally use supervised learning algorithms in a diagnostic approach for predicting risks associated with domains by using dedicated data sources.

6.2.1. Research design

This research highlights both qualitative and quantitative methods. Our mixed-method approach to a meta-analysis seeks to expand the size of presently measured effects by collecting information from a number of prior studies. The study involves general observation of the Group D population. A host institution was identified in a large collection of prior secondary studies that contributed a summary, sample size, magnitude, direction, and significance of sixty patient self-service health monitoring interests and behaviors. We delve into primary data through in-library research. We utilize both primary and secondary data to build a model, and we test that model with simple correlations and t-tests.

Research design is an iterative process, and although it is described as a linear workflow process, when we believe a design might work, we might proceed to the implementation step and run that design through its paces. At that time, we might quickly find issues with it, resolve those issues, and get back to refining the model structure for estimating the design. The mixed-method simple workflow used is an adaptation of that discussed in the sections titled, “Quantitative Research” and “Qualitative Research.” We start with a Problem statement, then gather background research to assist us. After that, we run through the design/prototyping, implementation/testing, evaluation, and publication steps in each of the quantitative and qualitative areas as needed, refining the design with what we find as we go. Finally, we use the results to prove the need for more detailed focus and research on the issues involved with the Problem statement.

6.3. Real-Time Patient Monitoring

Real-time patient monitoring enables tracking a patient's status and location throughout the day using various technologies that automatically collect data. These technologies detect habits and health-related variables of different people in different surroundings and daily routines, with or without their trust. Examples of variables being measured include heart rate, blood pressure, oxygen saturation, temperature, and weight. Monitoring can be generalized into intra-communications, peri-communications, and extra-communications, which means monitoring data communicated inside the body, on the skin surface, or outside the body. Currently, a capability gap exists between two-way communication devices that have high user’s trust and acceptance, and one-way communication devices such as sensors which drastically collect patient-specific information.

Various technologies have been developed to address this challenge, including cardiac telemetry, cardiac magnets, photoplethysmography, pulse oximetry, trans-esophageal echocardiography, bioimpedance pneumography, smart wearable devices, smart clothing, implantable coding chips, implantable smart sensing devices, and ambient sensing technologies. Different patient monitoring applications have different technology options. Within a care model period, the difference in information requirements determines what type of patient monitoring technology is suitable for that model period. A care model is a patient portioning method that divides patients with similar clinical needs into multiple short model periods for effective management. The information requirements determine what questions the patient monitoring answer. The technology options provide the capacity to manage various model periods. Therefore, the combination of information requirements and technology options forms a monitoring question such as "how to monitor heart rate."

The search for the best method for improving risk prediction in modeled populations using risk factor data has not ended but the review considers only the most widely used methods. Improvement comes from collaboration between modelers and data users in creatively addressing specific objectives, whether the purpose of prediction is allocation of scarce healthcare resources or proactive prevention and care based on greater efficiency and reduced cost from increasingly concentrated prevention and treatment of risk factors prior to the locks of the principal doors. Consequently, it is the goal of this review to summarize the risk prediction methods in the literature developed for the area already populated with vulnerable housing/rooming without duels.

Fig 6.2: Real-Time Patient Monitoring

6.3.1. Technologies for Monitoring

Monitoring vital parameters is crucial for real-time diagnosis, prognosis, and treatment decision strategies in different diseases. Patient health data is not static; it changes continuously like the changing pattern of data in the stream. Designing a continuous trend monitoring mechanism from streamed patient data to detect any change and record it for clinical significance is an important task and the core of the concept of real-time patient monitoring. Many pervasive patient monitoring systems have evolved in the last decade, while sensor-based, secure mobile applications for health monitoring are now available. With the development of smart sensing devices and low-cost wireless sensor networks, health monitoring-protocol regulation via a dedicated wireless channel is possible.

Continuous remote monitoring of patients has been shown to reduce emergency department visits, hospitalizations, and healthcare costs. Remote patient monitoring trusts medical professionals to make the right decision following symptom data obtained from a monitoring device. Wireless body area networks employ low-power miniaturized sensor nodes attached to or planted inside the human body to wirelessly monitor physiological signals. These signals, mainly for healthcare applications, originated in Europe, while transborder, intercontinental implementations started slowly a few years later. The need for remote monitoring of patients without requiring constant attention from healthcare professionals, as well as new technologies for creating triggered alarms, increased the research interest in WBANS in recent years. WBANS use wearable, personal, or implantable sensor nodes. Wearable sensors can noninvasively measure ECG, EMG, heart rate, flow sounds, photoplethysmographic signals, pulse transit time, perspiration, skin temperature, blood glucose level, etc. Thus, WBANS also helps both on disease diagnosis and therapy.

6.3.2. Data Sources and Collection Methods

Wearables like smartwatches and handheld devices are user-friendly, inexpensive, and with lower barriers of adoption and acceptance. Recent advancement in sensor technologies enable these small devices to monitor multiple physiological parameters. These smart devices can be used to proactively monitor patient conditions remotely and avoid long-lasting physical examinations in clinical settings. To compensate for their battery limitations, sensors may be intermittently activated to minimize energy consumption. These intermittent sampling phases depend on user activity events, which may be accurately identified based on raw sensor data with proper machine learning models. However, the battery life of smart devices is still much shorter than long-term health monitoring timelines, especially for chronic diseases that require continuous care. Furthermore, wearable sensors only cover a limited number of parameters. More research efforts focus on introducing lightweight sensor devices to enable advanced disease monitoring. For example, advanced smart sensors are deployed on gloves to allow multimodal physiological signal acquisition, including blood pressure, blood glucose, blood pulse, ECG, and skin temperature. Moreover, by monitoring the ECG signals via smartwear and investigating wavelet-transformed ECG feature patterns, arrhythmia in heart diseases could be predicted.

6.4. Risk Prediction Models

The proliferation of advanced population monitors for detecting threats and conditions relative to the health of nations and the rise of big data relative to populations based on the devices has promoted awareness of the potential for modeling and predicting the risk of consuming costly healthcare resources. The prediction is based on available, continuously updated, and potentially integrated data from countries coming from government or non-governmental organizations, social media such as health-geared information, blogs, search engines, and diagnostic and treatment center information about the occurrence and nature of the health event. However, noting the potential for prediction, the state of the art relative to methods for prediction, a review of the literature shows that traditional risk models are often based on simple statistical prediction based on risk factor data collected years or months prior to a diagnosis.

The potential for improving the models may come from modeling highly populated areas or using a log-normal count data model where counts are normally distributed after log-transformation.

6.4.1. Statistical Approaches

Risk prediction models are an essential part of today's clinical decision-making for chronic disease management. Clinical risk scores are usually created to identify high disease risk groups with the purpose of prompting preventive medication or implementing lifestyle modification interventions. As the number of hospitalized patients increases over time, predicting the future risk of adverse events is becoming crucial in order to deliver timely preventative measures. Statistical techniques are commonly used in the construction of clinical risk scores. Traditional regression modelling approaches, such as logistic regression, Poisson regression, and linear modelling are used for predicting binary or continuous outcomes, while Cox regression is used for survival analysis.

Traditional regression techniques use a small number of explanatory factors, known as covariates, which are used to fit a model to the data in order to gain insight into the dependence structure. The model can then be used for prediction purposes. Understanding and interpreting regression models is quite straightforward and time efficient. However, there are also many limitations to these approaches. The traditional statistical approaches have a number of limitations. Understanding and interpreting regression models is quite straightforward. However, these models can only accommodate a few predictors, are sensitive to outliers, rely on the chosen model, can easily overfit small training datasets, and are biased. Model development and validation strategies are usually influenced by which predictor selection method is used. In order to improve risk scores, researchers proposed variable selection techniques. Clearly, simple logistic and Cox regression models have been criticized for being oversimplified. In response, many statisticians have turned their attention to more sophisticated statistical techniques and machine learning developments.

6.4.2. Machine Learning Techniques

Machine Learning (ML) is defined as a computer algorithm that can learn from the data and make predictions without a set of rules. Recently, ML techniques have shown a substantial ability to outperform the classical statistical models in the context of risk modeling. There are many advantages of ML algorithms. First, they possess high flexibility, being capable of handling both linear and nonlinear relationships between input and output variables, as well as interactions among the predictors. Additionally, they are able to adapt and work well with the large volume and high dimensionality of data, particularly some ensemble learning algorithms. Furthermore, many existing ML algorithms are specifically built to handle imbalanced classes. In particular, in the context of medical predictions, the ratio of adverse events to non-events is typically very low. For example, for a study that uses in-hospital mortality as the outcome, the malignancy cohort had a cohort proportion of less than 10%. Therefore, more ML models may be better suited for capturing the important classifiers in highly imbalanced settings compared to traditional statistical methods that were primarily designed for models with balanced outcomes. Moreover, unlike some classical models that suffer from overfitting, ML algorithms can make very precise predictions, particularly when using ensemble approaches, by combining the results from multiple learners and reaching a final outcome. Finally, the ML approach provides a built-in mechanism for assessing variable importance. Unlike classical methods that provide estimates of variable effects, ML provides a ranking of variable importance, identifying the strongest predictors in the analysis.

6.5. Data Analytics Frameworks

In this section, we specifically introduce a family of frameworks for real-time data analytics using Big Data systems. We provide an overview on how frameworks offer a deferred programming model to support online and batch analytics. Our second goal is to show how other frameworks are specifically built for developing highly scalable systems for stream analytics, processing high volume data from various sources. Our contributions in this aspect are two-fold.

First, we provide a detailed overview of the above frameworks within the context of Big Data analytics. Recent pipeline utilization study reveals that about 70% of interactive queries in data warehouses run over a relatively small amount of data. Similarly, it has also been noted that in many enterprise applications in which business insights need to be generated from Big Data Warehouse, batch query and stream analytics are often needed together. Queries in such an analytics pipeline first run an initial batch query over older data currently circulating in the warehouse to get an overview of the result and then they run a stream query that gives continuous updates over the latest, most of the time, relatively small amount of data. For analytics systems that are used for batch processing, the current stream analytics techniques will not be suitable. However, frameworks provide an alleviation to that, by supporting batch and real-time interactive queries in the same system over the same datasets.

6.5.1. Frameworks for Real-Time Analytics

Though big data analytics has long been a passion of researchers, its usage for real-time applications is relatively new and less explored. What also makes it difficult is research being conducted on the basis of distributed computing and database paradigms. Dealing with real-time analytics is associated with various technologies such as Stream Processing Systems, Big Data Management Systems, Hybrid Model and Cloud Systems, Complex Event Processing Engines, and Data Streams Infrastructure Libraries.

Some of the tools being used for real-time analytics are user-friendly and help researchers quickly simulate the proposed algorithm, but they are still abstract provided wrappers.

Conventional models of big data analysis are not sufficient for real-time analytics as they process the data in batches leading to a delay in obtaining results. Also, there is no seamless media of exchange among existing tools, if at all an application is built using multiple of them. These problems prompted us to design a more efficient and better model that would be capable of faster computations. Another reason to reinvent the wheel is that many current tools were designed for a different class of problems and they do not extract maximum and necessary parallelization.

The performance of such systems is no more than a couple of times better than that of a single workstation. Also, all real-time analytics applications require low latency, and as such would need to perform costly computations at sad times. Many challenges have been posed by cloud computing for real-time big data analytics.

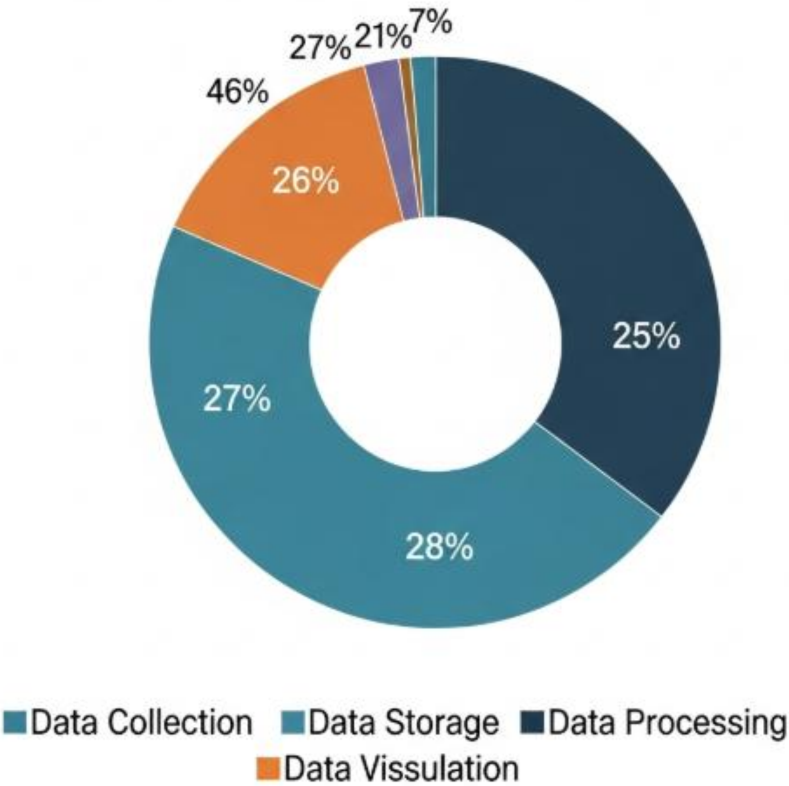

Fig : Utilizing Big Data Analytics for Real-Time Patient Monitoring

6.5.2. Integration with Existing Systems

A primary requirement of the above requirement is a robust integration capability with minimal alterations to existing infrastructures. A certain degree of parallel processing is in-built with MapReduce-style frameworks. Pub/Sub models employ a publish-subscribe mechanism to push sensor readings to multiple services which in-turn could perform some processing before they are written to databases. Pub/Sub models provide high availability, predictable performance, and an automatic scaling ability. A similar model employs a Distributed Messaging System. Anything that is written can be consumed by different consumers and saved to different databases. This is a good choice for high-speed analytics but does not have a failover-to-disk mechanism. A good choice if fault tolerance, high throughput, and low latency is more important than semantic richness of how events are monitored and what filters are used.

An alternative approach to integrate systems is the use of REST APIs. The use of REST APIs for data monitoring, collection, filtering, and persistence can also be compared to microservices architecture. Cloud-based solutions can be used to either write the API that gets called for every sample in the stream or utilize the service to save input datasets automatically in databases. Another solution utilizing microservices is published to container orchestration platforms which have deployed containers on top of the cloud platform underlying computing and storage resources to automatically scale resources based on load and gracefully shred container instances on a regular basis. These solutions are more of a classical approach. They are suitable for use-cases where low badness is not required.

6.6. Conclusion

This concluding section has summarized key parts of the work in a concise format, while providing new insights and information regarding the investigated domain. We have outlined the important role that big data analytics can play for making advances in both the patient-centric and personal monitoring domains, as well as in the development of personalized health-care approaches, with a view to improve quality of care and ultimately medication compliance, particularly for chronic diseases, diabetic patients and patients undergoing chemotherapy for cancer. Complementary to this summary of the major remaining content, we have also described what we see as future development trends in the investigated area. In particular, we see that there is a need for efficient, real-live systems capable of both personalized and active patient risk prediction, as well as real-time patient monitoring traits, limiting the required active patient participation to a minimum. To us, these systems seem essential if we are to arrive at innovative solutions for the personalized and patient-centric management of employers' health insurances and thereby increasing the overall societal productivity. We expect wearable and implantable biosensors technologies to rapidly evolve during the next couple of years, leading to a dramatic increase in available sensor measurements. At the same time, we expect the ongoing technology development in the big data systems and software area to result in real-time big data patient risk evaluation predictive engines supporting the societal goals outlined above. With this perspective, we hope that this work can provide a useful guide for further explorative efforts in the evolving domain of real-time patient monitoring and risk prediction.

6.6.1. Future Trends

For the last few decades, healthcare has witnessed a remarkable transformation in how healthcare professionals work with patient-data with the myriad technologies available today. Apart from routine hospital visits and paper sensation, most diseases now are monitored by Healthcare Analytics. Hospitals have existing equipment around like infusion pumps, mechanical ventilators, ECG machines, multiparameter patient monitors, etc. that gather vast amounts of patient-centric data in addition to the previous laboratory-based methods of assessments. Still, we find that the cost of delivering hospital care has reached an all-time high. We are in dire need for patient-centered secure data models to be used by healthcare Analytics technologies to reduce operating costs, and prolong people's lives with healthy longevity. Major investment decisions are being pushed into the intelligent decision support systems in Healthcare Analytics.

References

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., ... & Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. Stroke and Vascular Neurology, 2(4), 230-243. https://doi.org/10.1136/svn-2017-000101

Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), 44-56. https://doi.org/10.1038/s41591-018-0300-7

Reddy, S., Fox, J., & Purohit, M. P. (2019). Artificial intelligence-enabled healthcare delivery. Journal of the Royal Society of Medicine, 112(1), 22–28. https://doi.org/10.1177/0141076818815510

Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. JAMA, 319(13), 1317-1318. https://doi.org/10.1001/jama.2017.18391

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Sweater, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118. https://doi.org/10.1038/nature21056