Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 220 (2023) 826–831

Procedia Computer Science

www.elsevier.com/locate/procedia

The 4th International Workshop on Hospital 4.0 (Hospital)
March 15-17, 2023, Leuven, Belgium

Data Platforms for Real-time Insights in Healthcare: Systematic Review

Rui Miranda[a,*], Carlos Alves[a], António Abelha[a], José Machado[a]

[a] Centro ALGORITMI, Escola de Engenharia – Universidade do Minho, Campus Azurém, 4800-058 Guimarães, Portugal

Abstract

The ever-growing usage and popularity of Internet of Things devices, coupled with Big Data technologies and machine learning algorithms, have allowed for data engineers to explore new opportunities in healthcare and continuous care. Furthermore, there is a need to reduce the gap on time from when information is created to when actions and insights can be offered. However, a challenge in implementing a large-scale data processing architecture is deciding which tools are appropriate, and how to apply them in the best way possible. For example, streaming systems are now mature enough that hospitals worldwide can use their extremely large datasets, along with data producers, to predict and influence future events. Thus, the main objective of this systematic review is to identify the state-of-the-art in data platforms on healthcare that allow the creation of metrics and actions in real-time. The PRISMA guideline for reporting systematic reviews was implemented to deliver a transparent and consistent report, validating the technological advances in a critical sector. Multiple pertinent articles and papers were retrieved from the SCOPUS abstract and citation database on May 13, 2022, using several relevant keywords to identify potentially relevant documents published from January 2020 onward. These documents must have already been published in English and been already published, and accessible through the B-ON consortium that allows Portuguese students to legally download from most publishers. Over seven studies have been selected for deeper discussion based on their relevance and impact for this review, showcasing their main objectives, data sources, and tools used, as well as their approaches for interoperability and support of machine learning algorithms for decision support. In closing, the collected articles have shown that while Big Data is currently in use at health institutions of all sizes, the ability of processing large amounts of data from sensors and events, and notifying stakeholders as quickly as possible is still in its infancy.

Keywords: Data Engineering; Business Intelligence; Big Data; Machine Learning; Streaming Systems; Data Lake; Data Mining Event-Driven Microservices

* Corresponding author.
E-mail address: rui.miranda@algoritmi.uminho.pt

10.1016/j.procs.2023.03.110

1. Introductiom

Healthcare institutions are now understanding the value and impact data can have in the patient's quality of care. This extremely important asset is mostly used for operational record and clinical decision making. Consecutively, the next generation of healthcare services will employ IoT devices, equipped with sensors and wireless connections, gathering large amounts of data [10]. However, due to their format and volume, traditional RDBMs are not able to store in a timely manner. Modern developments allow for Big Data approaches, thus enabling institutions to extract insights from data that was previously thought to be useless or computationally expensive. The term Big Data can be applied to information that can't be processed or analyzed using traditional processes or tools [12]. Currently healthcare institutions have access to a large wealth of data, but can't extract value from them due to its format and quantity. Thus, approaches and technologies can be adopted to help identify patterns and extract valuable knowledge from these sources.

Some of these technologies include stateful stream processing, an application design pattern for processing unbounded streams of events, with the ability to store and access intermediate data; event-driven applications, an evolution of microservices, where they can perform arbitrary computations that involve reading data from or writing data to the state; and the publisher/subscriber messaging pattern, where publishers send data on named topics, and subscribers create named subscriptions to pull data from these topics [2, 6].

However, there are multiple challenges in implementing Big Data platforms in healthcare institutions. The biggest difficulty is, as healthcare units keep growing, so does the complexity of their medical and information technologies, increasing the possibility of medical errors [9]. Thus, interoperability is an imperative when developing and integrating software solutions, allowing the exchange of important data, and initiate actions between different products without any additional effort on the user [4]. Furthermore, data privacy must be assured when processing and storing medical data, and special care must be taken care when implementing machine learning algorithms in healthcare, to ensure accuracy, precision, and avoiding biases.

Therefore, the main objective of this systematic review is to ascertain the current state-of-the-art in the implementation of technologies that enable the processing of information in real-time in healthcare, their main objectives, the adoption of Big Data platforms, and support for the adoption of Deep Learning algorithms.

2. Methods

In this section, the methodology for this systemic review is presented and justified. A careful analysis of past papers and articles allows for an understanding of what others have previously done, what challenges they had to face, and how they were able to overcome them. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist was implemented to provide an accurate account of the main objectives of this review, its methodology, and results [11]. The data sources used, namely online databases, the criteria used to include and exclude documents, and the process to identify and extract information are documented. By adopting explicit, rigorous, and accountable methods, it is expected the interpretation of research findings and research knowledge that can improve society [6].

2.1. Information Sources

As a data source, the SCOPUS[1] abstract and citation database was used, due to its size, wide coverage in terms of publication areas, and quality assurance basis. This database was last searched on the 1st of September of 2022. Furthermore, to retrieve the impact of journals on the showcased publications, the SCImago Journal & Country Rank[2] was used. This source was consulted on September 15 of 2022.

[1] https://www.scopus.com
[2] https://www.scimagojr.com

2.2. Search Strategies

To proceed to the literature search several keywords were identified as a starting point, as shown in figure 1. These keywords will be applied to the title, abstract, and keywords fields (TITLE-ABS-KEY). To further organise the search resources, the keywords were separated into two groups, which are related by the **AND** operator. The keywords in each group are related with the **OR** operator. This reasoning for this decision is made so that each group selects all documents that include at least one of its keywords, and then ensures that only documents that have at least one term from each of the groups are retrieved. The first group is relative to the technical areas and subjects directly related to the research topic, while the second group aims to filter by broader areas of technological scope in order to focus the results in the context of healthcare and continuous care. This review is limited to the Computer Science, Engineering, and Medicine subject areas, and includes articles published from 2019 to 2022.

2.3. Inclusion and Exclusion Criteria

After retrieving the initial set of documents, it's necessary to define criteria that allow for a more precise filtering of the results obtained, depending on what is intended, whether they are inclusion or exclusion criteria. Therefore, the selection criteria that were applied throughout the research, in the order of their application is now presented:

- **Accessibility:** Documents must be available at B-ON, the consortium that allows Portuguese universities to access to a majority of publishers;
- **Timeliness:** Documents must have been published from 2019 onward;
- **Research Area:** Papers should be within the areas of Computer Science and Engineering in a healthcare environment, which are the context areas of this review.
- **Document Type**: The documents must be already published, be either an Article or Paper, and written in English.

This screening process was conducted by Rui Miranda and Carlos Alves, working independently, by reading the titles and abstracts provided by SCOPUS during the search process. When the article seemed relevant for further evaluation, the full text was downloaded and read. Although the B-ON consortium allows Portuguese universities and institutions to access most paid articles, some were excluded due to their cost.

Thus, the initial set resulting from the previously detailed search strategy resulted in more than 3503 documents. Figure 1 presents a diagram illustrating the performed search process.

2.4. Synthesis Methods

The data collection process was conducted by Rui Miranda and Carlos Alves independently. To explore the panorama of available technologies and approaches in processing information in real-time on a healthcare environment, several categories were defined and used to showcase the selected documents. These categories were designed to answer the proposed research questions:

1. **Main Objectives:** The document's main objectives and goals are identified and collected
2. **Data Sources:** These can come from IoT devices, manually created by medical professionals, or synthetic data
3. **Tools Used:** The adopted or proposed frameworks and programming languages for the development of the proposed solution are identified
4. **Semantic Interoperability:** All steps taken to ensure semantic interoperability, including standards and protocols are collected
5. **Machine Learning Support:** Whether the proposed platform supports the development and implementation of machine learning algorithms for decision support in healthcare
6. **Limitations:** In this category all weak points are presented, either recognised by the authors or identified by the reviewers, as well as possible future steps for improvement

This work was developed as part of a research project regarding the application of real-time information processing approaches in multiple industries. Therefore, with this compilation of the state-of-the-art in a healthcare environment, future work would include the study and development of a platform to integrate these concepts, and is able to address the main limitations elucidated, thus bringing additional value to the healthcare industry, and improve the quality of care.

3. Results and Discussion

3503 documents were retrieved from SCOPUS with the research query before mentioned, with the most relevant ones selected, resulting in a subset of 291 documents. All documents from this subset were screened in more detail by analysing the abstract, results and conclusion, resulting in a set of 7 documents that correspond to the intended research themes and context and, as such, constitute the final set of documents and papers that will be used for the literature review and state of the art.

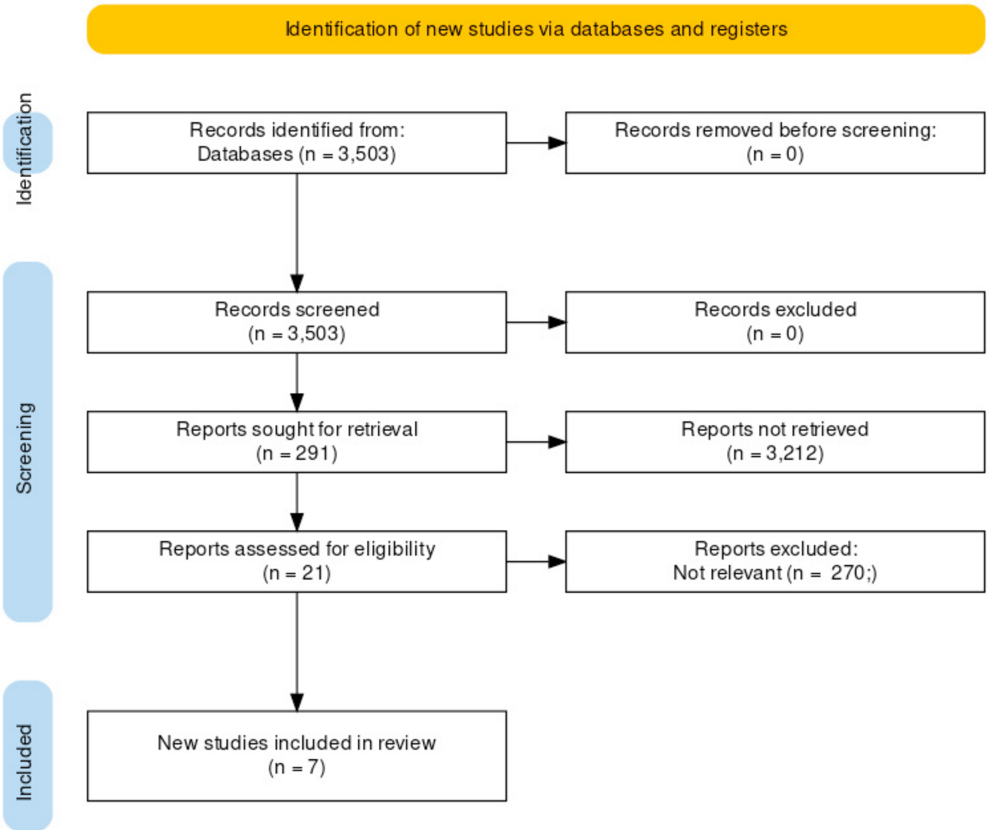

Fig. 1. Research Progress Flow Diagram (Created using [7])

The main objective of this review is to identify the state-of-the-art in data platforms on healthcare that allow the creation of metrics and actions in real-time, allowing for rapid and effective patient care. Additional key aspects include the usage for machine learning approaches, detecting patterns in medical data for further knowledge, as well as considerations for data semantics and standardised formats, enabling further interoperability and validating the platform's viability.

The first research question (RQ1) aimed to elicitate the main benefits of developing and implementing data platforms for real-time insights for the healthcare community and patients themselves.

Of the twenty-two articles analysed, the authors address a variety of topics of which are presented and discussed.

El-Ganainy et al. propose a system that's capable of providing real-time clinical decision support without the need for an offline system for training with large data sets, as compared to machine learning (ML) systems. This ML system was divided into two stages, where one uses hierarchical temporal memory (HTM) and the second to enable real-time flow processing and provide unsupervised predictions. The authors concluded that the proposed system outperforms logistic regression (LR) in terms of classification accuracy, recall, precision, and area under the receiver operating curve (AUROC).

Shehab et al. addressed a framework for improving health monitoring systems through IoT-enabled devices used in multiple environments, such as smart homes and smart hospitals. These systems need a dynamic analysis of critical patient flows (e.g., ECG flow). Their findings indicate that a per-user scheduler overcomes the problems of lack of resources of Fog Computing nodes and that it fully utilises the Fog Computing infrastructure and thus allowing the system to safely expand to double its capacity.

Alharbi et al. The authors propose a real-time heart rate prediction system, which will help doctors and patients to anticipate heart rate risk in real time. The proposed system consists of two phases, an offline phase, which aims to develop the model using different prediction techniques to find the smallest root mean square error, and an online phase, where Apache Kafka and Apache Spark were used to predict heart rate in advance based on the best developed model.

Saleh et al. have tackled a system for predicting systolic blood pressure, in order to predict it on the fly. This system works either by an offline mode model using several deep learning models to obtain the smallest root square error, or in online mode using Apache Kafka and Apache Spark to predict the systolic blood pressure in real time using the best deep learning model and streaming time series data from the systolic blood pressure system. It was concluded that with this study that the BI-LSTM model achieved the best performance using three hidden layers.

Thanh et al. have presented their IoHT Platform for Healthcare environment that's designed by a brokerless, microservices architecture, focusing heavily on data collection, users and device management, and remote device control. In addition, the IoHT platform addresses the limited processing capacity of devices, power savings to the device, speed and accuracy of data collection, security mechanisms, and scalability of the system. Thus, to make the IoHT platform suitable for the health tracking domain, real-time alerts for the medical team were also added. The authors concluded the effectiveness of the proposed IoHT platform (i.e., the proof-of-concept) in performance, with no errors and unaffected by geographical distance.

Afreen Banu and Rajamani have designed an IoT-based multisensory online vital monitor, named VITALS, to detect four physiological parameters at the bedside, which include heart rate, body temperature, blood pressure, and peripheral oxygen saturation. As such, this system extracts human vital signs every 30 minutes, and sends them to a Big Data analytics system over Wi-Fi for further analysis. For this purpose, several technologies already used by other authors were used as well, like Apache Kafka (to collect live data streams from connected sensors), Apache Spark (to categorize patient's vital signs and notify medical professionals when identifying abnormalities in physiological parameters) the Hadoop Distributed File System (HDFS) (to archive data streams for further analysis and long-term storage), Spark SQL, Hive and Matplotlib (to help caregivers access/visualize appropriate information from the collected data streams and explore/understand the health status of individuals). A mobile application was also developed by the authors. It was concluded that the proposed system provides enhanced care solutions, especially for those whose access to care services is limited.

Hassan et al. propose a system that focuses on applying machine learning streaming models (i.e., streaming linear regression with SGD) on ingested streaming integrity data events to bootstrap streaming via Kafka threads. The experimental results are done on historical medical datasets (i.e., a diabetes dataset, a heart disease dataset, and a breast cancer dataset) and generated datasets that are simulated for wearable medical sensors. The experimental results have proven that the online prediction system can learn online and update the model according to the arrival of new data and window size.

4. Conclusions and Future Work

Healthcare environments are known for being complex and stressful, especially when people's lives are at stake. Thus, technological advancements in identifying potential events and scenarios, and notifying stakeholders as fast as possible are crucial, including in hospitals and health institutions. While technologies like Big Data and machine learning allow for using the data collected from past patients for knowledge and insights, when combined with real-time information processing approaches, medical professionals can be notified as quickly as possible when new events occur.

In this paper, a systematic review was conducted to ascertain the current state-of-the-art in the implementation of technologies that enable the processing of information in real-time in healthcare. The review implements the PRISMA 2020 statement for reporting systematic reviews, and uses the SCOPUS database as the primary source for documents.

This work was developed as part of a research project regarding the application of real-time information processing approaches in multiple industries. Therefore, with this compilation of the state-of-the-art in a healthcare environment, future work would include the study and development of a platform to integrate these concepts, and is able to address the main limitations elucidated, thus bringing additional value to the healthcare industry, and improve the quality of care.

References

[1] Afreen Banu, E., Rajamani, V., 2023. Design of Online Vitals Monitor by Integrating Big Data and IoT. Computer Systems Science and Engineering 44, 2469–2487. URL: https://www.techscience.com/csse/v44n3/49124, doi:10.32604/csse.2023.021332.

[2] Akidau, T., Chernyak, S., Lax, R., 2018. Streaming Systems. First edition ed., O'Reilly Media, Inc.

[3] Alharbi, A., Alosaimi, W., Sahal, R., Saleh, H., 2021. Real-Time System Prediction for Heart Rate Using Deep Learning and Stream Processing Platforms. Complexity 2021, 1–9. URL: https://www.hindawi.com/journals/complexity/2021/5535734/, doi:10.1155/2021/5535734.

[4] Cardoso, L., Marins, F., Portela, F., Santos, M., Abelha, A., Machado, J., 2014. The next generation of interoperability agents in healthcare. International journal of environmental research and public health 11, 5349–71. URL: http://www.ncbi.nlm.nih.gov/pubmed/24840351http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4053905, doi:10.3390/ijerph110505349.

[5] El-Ganainy, N.O., Balasingham, I., Halvorsen, P.S., Rosseland, L.A., 2020. A New Real Time Clinical Decision Support System Using Machine Learning for Critical Care Units. IEEE Access 8, 185676–185687. URL: https://ieeexplore.ieee.org/document/9220774/, doi:10.1109/ACCESS.2020.3030031.

[6] Gough, D., Oliver, S., Thomas, J., 2012. An Introduction to Systematic Reviews. SAGE Publications Ltd.

[7] Haddaway, N.R., Page, M.J., Pritchard, C.C., McGuinness, L.A., 2022. Prisma2020: An r package and shiny app for producing prisma 2020-compliant flow diagrams, with interactivity for optimised digital transparency and open synthesis. Campbell Systematic Reviews 18, e1230. URL: https://doi.org/10.1002/cl2.1230, doi:https://doi.org/10.1002/cl2.1230. https://doi.org/10.1002/cl2.1230.

[8] Hassan, F., E., M., Sahal, R., 2020. Real-Time Healthcare Monitoring System using Online Machine Learning and Spark Streaming. International Journal of Advanced Computer Science and Applications 11, 650–658. URL: http://thesai.org/Publications/ViewPaper?Volume=11&Issue=9&Code=IJACSA&SerialNo=77, doi:10.14569/IJACSA.2020.0110977.

[9] Miranda, M., Duarte, J., Abelha, A., Machado, J., Neves, J., 2009. Interoperability and healthcare, pp. 205–212. URL: http://hdl.handle.net/1822/11288. done.

[10] Oliveira, D., Miranda, R., Hak, F., Abreu, N., Leuschner, P., Abelha, A., Machado, J., 2021. Steps towards an healthcare information model based on openehr. Procedia Computer Science 184, 893–898. doi:10.1016/j.procs.2021.04.015.

[11] Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., Moher, D., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372. URL: https://www.bmj.com/content/372/bmj.n71https://www.bmj.com/content/372/bmj.n71.abstract, doi:10.1136/BMJ.N71.

[12] Pereira, C., Mesquita, J., Guimarães, D., Santos, F., Almeida, L., Aguiar, A., 2019. Open iot architecture for continuous patient monitoring in emergency wards. Electronics 8, 1074. URL: https://www.mdpi.com/2079-9292/8/10/1074, doi:10.3390/electronics8101074.

[13] Saleh, H., Younis, E.M.G., Sahal, R., Ali, A.A., 2021. Predicting Systolic Blood Pressure in Real-Time Using Streaming Data and Deep Learning. Mobile Networks and Applications 26, 326–335. URL: https://link.springer.com/10.1007/s11036-020-01645-w, doi:10.1007/s11036-020-01645-w.

[14] Shehab, R., Taher, M., Mohamed, H., 2021. Live Big Data Analytics Resource Management Techniques in Fog Computing for TeleHealth Applications. Jordanian Journal of Computers and Information Technology 7, 89. URL: https://www.ejmanager.com/fulltextpdf.php?mno=6525, doi:10.5455/jjcit.71-1605864596.

[15] Thanh, L.N.T., Phien, N.N., Nguyen, T.A., Vo, H.K., Luong, H.H., Anh, T.D., Tuan, K.N.H., Son, H.X., 2021. IoHT-MBA: An Internet of Healthcare Things (IoHT) Platform based on Microservice and Brokerless Architecture. International Journal of Advanced Computer Science and Applications 12, 594–601. URL: http://thesai.org/Publications/ViewPaper?Volume=12&Issue=7&Code=IJACSA&SerialNo=68, doi:10.14569/IJACSA.2021.0120768.