International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Elastic Data Analytics in Healthcare: Enhancing Patient Outcomes

Shobha Y.[1], Dr. Jyothi D. G.[2], Yamini Sahukar P.[*3]

Submitted:13/03/2024 **Revised**: 28/04/2024 **Accepted**: 05/05/2024

Abstract: Effective management and study of Electronic Health Record (EHR) statistics are essential in transforming healthcare delivery and decision-making. This paper explores the amalgamation of Elastic Data Analytics (EDA) into EHR systems as an approach to address the tasks of managing vast volumes of heterogeneous healthcare data. EDA enables healthcare groups to dynamically scale data resources and adapt analytics workflows to changing requirements, facilitating real-time access to information and data-driven decision-making. The application of EDA in specific healthcare domains, such as cardiovascular disease (CVD) monitoring, highlighting its prospective to reform disease prediction, population health management, and personalized medicine approaches. However, successful implementation requires addressing tasks such as data interoperability, privacy and security concerns, and scalability of healthcare infrastructure. Cooperative efforts between healthcare services providers, data scientists, and policymakers are essential to harness the full potential of EDA and drive positive outcomes in healthcare delivery. This paper underscores the transformative impact of EDA in healthcare and provides visions into its future suggestions for improving patient care and advancing healthcare innovation.

Keywords: *Elastic data analytics, Electronic health records (EHR), Real-time data analysis, Cardiovascular disease (CVD) monitoring.*

1. Introduction

It is compelling to adopt electronic health records (EHR) systems across healthcare settings. These Computer based records provide an important participation in digitizing patient health information and facilitating seamless access to medical records. However, the growing volume and complexity of data within EHR systems pose challenges in terms of data management and analysis.

One challenge stems from the need to pre-process and to assimilate generated on different hardware within the EHR system, such as patient medical history, diagnostic tests, and treatment plans. Combining these disparate sources of data is essential for gaining comprehensive analysis of present health condition along with optimizing care delivery [1]. However, the sheer volume of records and the heterogeneity of data sources can make this integration process complex and resource-intensive.

Another challenge arises from the increasing demand for real-time access to EHR data and timely analysis to support clinical decision-making. Healthcare providers require access to up-to-date patient information to deliver timely and effective care. However, ensuring timely access to EHR data and performing real-time analysis can strain existing data processing infrastructure and require efficient management of computing resources.

To address these challenges, there is a growing need for elastic data analytics solutions tailored for EHR systems. Elastic data analytics refers to the ability to dynamically adjust the processing and analysis of data based on changing requirements and available resources. By deploying elastic data analytics frameworks, healthcare organizations can optimize the administration and analysis of EHR data, ensuring timely access to analysed data and enabling data-driven decision-making [2].

These frameworks leverage scalable computing resources and adaptive analytics algorithms to handle the fluctuating demands of EHR data processing. By dynamically allocating resources based on workload demands and quality-of-service requirements, elastic data analytics frameworks can optimize performance and efficiency while minimizing resource wastage.

2. Data analytics of EHR wrt elastic

Data analytics is an important step in enabling healthcare organizations to manage and process vast amounts of data in EHR systems. These analytics facilitate the identification of trends, elimination of redundant information, and aggregation of relevant data, thereby supporting efficient information management. However, the use of data analytics in EHR systems necessitates robust data storage and processing capabilities, often leveraging resources from the cloud-to-things continuum nodes.

Elastic data analytics solutions must be customized to meet the specific requirements of each healthcare institution, confirming that the results align with their objectives [3].

[1] Associate Professor, Dept. of Artificial Intelligence and Machine Learning, Bangalore Institute of Technology, Bangalore, Karnataka, India. ORCID ID:
[2] Professor & HOD, Dept. of Artificial Intelligence and Machine Learning, Bangalore Institute of Technology, Bangalore, Karnataka, India. ORCID ID: 0000-0002-1859-6733
[3] Asst. Professor, Dept. of Artificial Intelligence and Machine Learning, Bangalore Institute of Technology Bangalore, Karnataka, India. ORCD ID: 0009-0005-8535-4880
** Corresponding Author Email: yaminisahukar@gmail.com*

Controlling the circulation of information is essential to increase the competence of both the system and the data.

Elastic data analytics impact several critical dimensions that must be cautiously taken care in a sustainable computing and information environment. Data quality, resource consumption, and cost are among the key considerations. Accuracy and freshness are particularly important aspects of data quality, with accuracy representing the level to which data reflect real-world scenarios and freshness indicating the timeliness of the data.

Different types of analytics processes may be executed considering the accuracy and freshness necessities of the data. Small-scale analytics processes are suitable for scenarios where accuracy and freshness are low, giving important information from atomic data. As accuracy and freshness increase, larger-scale analytics processes are deployed to take care of big amount of data for business decision-making.

The resource consumption of nodes within the cloud-to-things continuum is directly influenced by the accurateness of the analytics processes. Greater accuracy may involve more IoT devices in sending information, leading to increased resource consumption. Edge and fog nodes can help distribute and process analytics tasks, but their dispersion may increase resource consumption.

Apart from data costs, the handling of EHR data entails infrastructure costs related with the use of edge and fog nodes. The higher the precision and frequency at which information must be processed, the greater the need for processing this information in edge or fog nodes, resulting in higher infrastructure costs.

In environments where multiple healthcare applications are consuming similar information and utilizing the same computing resources, elastic data analytics solutions should be able to adjust the provided quality based on available resources and consumer requirements. An efficient framework is needed to manage the dissemination of analytics processes within the cloud-to-things continuum, ensuring optimal resource utilization while maximizing the worth of service provided [4].

Healthcare is undergoing a profound transformation powered by technological advancements and the growing availability of health-related data. In this era of digital healthcare, the effective management and study of data play an acute part in improving patient outcomes, optimizing clinical workflows, and driving innovation. Elastic data analytics, which enables dynamic scaling and adaptation of data resources in response to changing requirements, has emerged as a powerful tool for addressing the complex challenges of healthcare delivery and decision-making.

At its core, elastic data analytics in healthcare involves the presentation of scalable and flexible data management techniques to healthcare data, encompassing rich data types, sources, and use cases. This includes structured data such as electronic health records (EHRs), clinical databases, and administrative claims data, and also unstructured data such as medical imaging, genomic data, and data that is generated by wearable devises worn by the patient and mobile applications.

The ability to handle huge volumes of data efficiently and cost-effectively. Healthcare industry also faces the challenge of managing ever-growing volumes of data which is regularly produced from different physical places, including traditional clinical systems, remote monitoring devices, and patient engagement platforms [5]. By leveraging elastic data management techniques, such as cloud computing, distributed databases, and parallel processing frameworks, healthcare providers can scale their data infrastructure dynamically to accommodate fluctuations in data volume and processing demands, ensuring that critical healthcare insights are available when and where they are needed.

The healthcare industry is provided with actionable insights from their data in real-time, facilitating more timely and informed decision-making at all levels of the healthcare system by Elastic analytics. In our research we could use the data for predictive analytics models powered by elastic data analytics can help identify patients at risk of developing chronic conditions or experiencing adverse events, enabling proactive interventions to prevent complications and improve outcomes. Similarly, real-time monitoring and analysis of clinical workflows can identify inefficiencies and bottlenecks, enabling healthcare providers to optimize resource allocation and improve the delivery of care.

However, the implementation of elastic data analytics in healthcare has its own challenges. Privacy and security concerns, regulatory compliance, data interoperability, and technical complexity are among the key issues that need to be considered to realize the full potential of elastic data analytics in healthcare. Additionally, the effective integration of elastic data analytics into clinical workflows and decision-making processes requires collaboration and coordination across multidisciplinary teams, including healthcare providers, data scientists, IT professionals, and policymakers [6].

3. Literature review

The proliferation of Internet-connected devices has resulted in a surge of data collection, which companies leverage to enhance their decision-making processes. This trend places increasing demands on cloud and communications infrastructure. Given limited resources and the need for resource sharing, efficient management becomes critical.

The cloud-to-things continuum offers a solution by enabling analytics closer to data sources. In this article, we propose various dimensions for achieving elastic analytics and introduce a dynamic framework for modifying their actions [10].

The Internet of Things (IoT) paradigm promises automation of real-world processes. However, when applying IoT to resource-intensive domains, stringent Quality-of-Service (QoS) requirements, particularly ultra-short response times, become critical. To address this challenge, we explore the distribution of computational workloads across infrastructure layers (edge, fog, and cloud). Additionally, the leverage software-defined networks (SDNs) to enhance QoS by leveraging the global network view provided by SDN controllers and executing optimization algorithms.

The paper focus lies in identifying optimal placements for both computation elements and SDN controllers to achieve superior QoS. While separate optimization of computing and networking dimensions is possible, it often leads to suboptimal outcomes. Therefore, we propose a unified approach to address this problem. Specifically, we analyze the impact of both dimensions on response times within fog computing environments powered by SDNs.

DADO, a novel framework for identifying optimal deployments of distributed applications. Leveraging mixed-integer linear programming, DADO dynamically determines the most efficient placement strategy is introduced. An evaluation using an Industrial Internet of Things (IIoT) case study demonstrates that our proposed framework achieves scalable deployments across topologies of varying sizes and user bases. Remarkably, our approach yields response times up to 37.89% lower than alternative solutions and up to 15.42% shorter than state-of-the-art benchmarks [2].

Artificial Intelligence (AI) systems heavily reliant on Cloud computing face challenges related to transmission latency and bandwidth consumption. These limitations hinder real-time monitoring of physical objects, especially within the Internet of Things (IoT) context. Edge systems, positioned closer to end devices, cater to time-sensitive applications. However, their computational constraints pose difficulties when handling state-of-the-art Deep Neural Networks (DNNs).

A novel technology framework that merges the Edge-Cloud architecture with the advantages of BranchyNet. This combination aims to achieve fault-tolerant and low-latency AI predictions. Our implementation and evaluation demonstrate the benefits of deploying Distributed DNNs (DDNNs) in the Cloud-to-Things continuum. Notably, compared to Cloud-only deployments, our approach yields a remarkable 45.34% improvement in response time. Additionally, the paper has an extension for Kafka-ML, enhancing flexibility and reducing rigidity in managing and deploying DDNNs across the Cloud-to-Things continuum [6].

The rapid evolution of Internet of Things (IoT) computing and technologies has catalyzed the decentralization of Cloud-based systems. Directly transmitting all IoT device data to the Cloud, however, is impractical for applications with stringent demands on real-time response, low latency, energy-efficient processing, and robust security.

This decentralization trend has given rise to a new computing layer situated between the Cloud and IoT—the Edge computing layer. Comprising a spectrum of devices, from small computing units (e.g., Raspberry Pi) to larger nodes (such as Gateways, Road Side Units, Mini Clouds, MEC Servers, and Fog nodes), this layer aims to address the limitations of centralized Cloud architectures.

The research deals with the intricacies of processing IoT data streams within the Edge computing layer. Leveraging a real-world dataset derived from car telemetry and an actual infrastructure featuring Raspberry Pi and Node-Red servers, we illuminate the challenges associated with meeting real-time requirements for IoT stream processing applications. [3]

Enabling Elastic Data Analysis for CVD Prediction Using Electronic Medical Records or Prescriptions:

Elastic data analysis has an important role in predicting cardiovascular disease (CVD) risk and improving patient outcomes by leveraging electronic medical records (EMRs) or prescription data. This approach enables healthcare organizations to utilize scalable and flexible data management techniques to get more usable data from large and complex datasets, ultimately enhancing the correctness and productivity of CVD prediction models.

Data Integration and pre-processing:

The first step in enabling elastic data analysis for CVD prediction involves integrating and pre-processing electronic medical records or prescription data from diverse sources [7]. This may include aggregating patient demographic information, clinical biomarkers, medication histories, laboratory test results, and other relevant data elements from EMRs, pharmacy databases, and health information exchanges. Data pre-processing techniques, such as data cleaning, normalization, and feature engineering, are then applied to ensure data quality and consistency before analysis.

Feature Selection and Extraction:

Next, elastic data analysis techniques are used to identify and extract relevant features from the integrated dataset for CVD prediction. This may involve applying ML algorithms, such as feature selection methods and dimensionality reduction techniques, to prioritize informative variables and reduce the computational complexity of the prediction model. Features such as age, gender, blood pressure, cholesterol levels, smoking status, medication adherence, and comorbidities are commonly used in CVD prediction models to capture the multifactorial type of the disease [8].

Model Development and Training:

Elastic data analysis enables the development and training of predictive models for CVD risk assessment using the extracted features. ML algorithms, such as logistic regression (LR), decision trees, random forests, support vector machines, or deep learning neural networks, are trained on historical patient data to learn patterns and relationships between predictor variables and CVD outcomes. Model performance is evaluated using metrics such as accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC) to assess predictive performance and generalizability.

Scalable Computing Infrastructure:

Elastic data analysis requires a scalable computing infrastructure to support the computational demands of model development, training, and deployment. Cloud computing platforms, distributed computing frameworks, or high-performance computing clusters are used to provision and manage computational resources dynamically based on workload requirements. This enables healthcare organizations to scale their data analytics infrastructure up or down as needed to accommodate changing data volumes, processing demands, and user interactions, ensuring that predictive models can be developed and deployed efficiently and cost-effectively.

Real-time Prediction and Decision Support:

Once trained, elastic data analysis models can be deployed to generate real-time predictions of CVD risk for individual patients based on their clinical profiles and medication histories. These predictions can be integrated into clinical decision support systems or electronic health record systems to provide healthcare providers with actionable insights and personalized treatment recommendations at the point of care. By leveraging real-time predictive analytics, healthcare organizations can identify high-risk patients earlier, implement targeted interventions promptly, and ultimately improve patient outcomes and reduce the incidence of CVD-related complications [9].

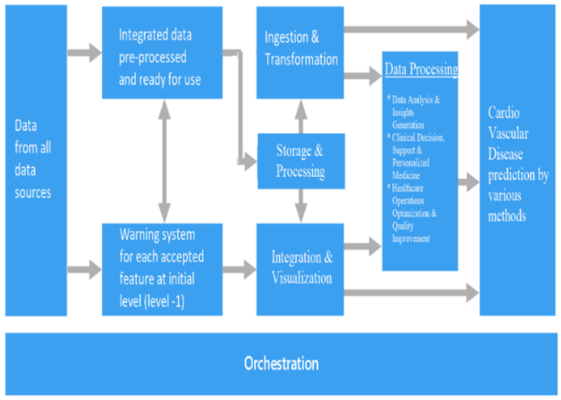

Fig 1: Block diagram of CVD prediction using EHR

In the fig 1. The blocking diagram is shown depicting the storage of health records using elastic data storage.

a. Data from All Sources:

This represents the raw health data collected from various sources, including hospitals, clinics, wearable devices, and other healthcare providers.

Elastic data storage efficiently ingests this diverse data stream.

b. Ingestion & Transformation:

The raw data undergoes initial processing, including cleaning, structuring, and transformation.

Elastic storage ensures smooth handling of this data influx, regardless of volume.

c. Data Processing:

In this stage, insights are generated from the processed data.

Elasticity allows for real-time analysis, clinical decision support, personalized medicine approaches, and optimization of healthcare processes.

d. Storage & Processing:

Elastic data storage accommodates both pre-processed integrated data and the results of data processing.

It dynamically allocates resources based on demand, ensuring scalability and cost-effectiveness.

e. Integration & Visualization:

Processed data integrates seamlessly with other healthcare systems (billing, radiology, etc.).

Elastic storage supports user-friendly interfaces for clinicians and administrators to access and visualize the data.

f. Orchestration:

This overarching block coordinates the entire process, ensuring efficient management of health records.

Elasticity adapts to changing requirements, maintaining system responsiveness.

g. Cardiovascular Disease Prediction:

A specific application of processed EHRs—using insights to predict cardiovascular health.

Elastic storage enables efficient storage and retrieval of relevant data for predictive modeling.

4. Conclusion

The integration of elastic data analytics into electronic health record (EHR) systems presents a transformative opportunity to report the complex challenges facing modern healthcare delivery and decision-making. As healthcare organizations strive to manage and analyze vast volumes of heterogeneous data, elastic data analytics emerges as a powerful solution to optimize data management, processing, and analysis workflows.

The adoption of elastic data analytics enables healthcare providers to derive actions from diverse sources of healthcare data, including structured EHRs, unstructured medical imaging, genomic data, and patient-generated health data. By leveraging scalable computing resources, adaptive analytics algorithms, and real-time processing capabilities, healthcare organizations can unlock the full potential of their data to improve patient outcomes, optimize clinical workflows, and drive innovation in healthcare delivery.

Furthermore, the application of elastic data analytics in specific domains, such as cardiovascular disease (CVD) monitoring, illustrates its potential to revolutionize disease prediction, population health management, and personalized medicine approaches. By harnessing predictive modeling techniques and real-time analytics, healthcare providers can identify individuals at high risk of developing CVD, track disease trends at the population level, and deliver targeted interventions to improve patient outcomes and reduce the burden of CVD on individuals and communities.

The integration of elastic data analytics into EHR systems represents a paradigm shift in healthcare, empowering healthcare organizations to harness the power of data to drive better outcomes, enhance patient experiences, and advance the delivery of high-quality, personalized care. By embracing elastic data analytics, healthcare providers can navigate the complexities of modern healthcare delivery with confidence and pave the way for a healthier, more connected future.

Author contributions

Shobha Y:

Literature review, performance metrics, leading framework conceptualization and design, developing metric selection guidelines, Contributing to theoretical framework section.

Dr. Jyothi DG:

Gathering empirical data through interviews and surveys, conducting case studies for framework validation, Analysing data using statistical techniques, Documenting results for empirical validation section.

Yamini Sahukar P:

Developing adaptability mechanisms for the framework, designing guidelines for framework tailoring, ensuring scalability and integration with existing processes, Contributing to conclusion section.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Ciffolilli, A. and A. Muscio. 2018. “Industry 4.0: national and regional comparative advantages in key enabling technologies”. European Planning Studies, 26(12):2323–2343. doi:10.1080/09654313.2018.1529145.

[2] J. L. Herrera, J. Galán-Jiménez, J. Berrocal and J. M. Murillo, "Optimizing the Response Time in SDN-Fog Environments for Time-Strict IoT Applications," in *IEEE Internet of Things Journal*, vol. 8, no. 23, pp. 17172-17185, 1 Dec.1, 2021, doi: 10.1109/JIOT.2021.3077992.

[3] Fatos Xhafa, Burak Kilic, Paul Krause, Evaluation of IoT stream processing at edge computing layer for semantic data enrichment, Future Generation Computer Systems, Volume 105, 2020, Pages 730-736.

[4] M. Zhaofeng, W. Xiaochang, D. K. Jain, H. Khan, G. Hongmin and W. Zhen, "A blockchain-based trusted data management scheme in edge computing", *IEEE Trans. Industrial Informatics*, vol. 16, no. 3, pp. 2013-2021, 2020

[5] Z. Yang, K. Yang, L. Lei, K. Zheng and V. C. M. Leung, "Blockchain based decentralized trust management in vehicular networks", *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1495-1505, 2019.

[6] Daniel R. Torres, Cristian Martín, Bartolomé Rubio, Manuel Díaz, An open source framework based on Kafka-ML for Distributed DNN inference over the Cloud-to-Things continuum, Journal of Systems Architecture, Volume 118, 2021

[7] M. Villari, M. Fazio, S. Dustdar, O. Rana and R. Ranjan, "Osmotic computing: A new paradigm for edge/cloud integration", *IEEE Cloud Computing*, vol. 3, no. 6, pp. 76-83, 2016.

[8] Lyytinen, K., Y. Yoo and R.J. Boland Jr. 2016. Digital product innovation within four classes of innovation networks. Info Systems J 26:47–75. doi:10.1111/isj.12093

[9] Lasi, H., P. Fettke, H.-G. Kemper, T. Feld, and M. Hoffman. 2014. “Industry 4.0”. Business & Information Systems Engineering 6(4):239–242.

[10] S. Laso *et al*., "Elastic Data Analytics for the Cloud-to-Things Continuum," in *IEEE Internet Computing*, vol. 26, no. 6, pp. 42-49, 1 Nov.-Dec. 2022, doi: 10.1109/MIC.2021.3138153.