

A Data Management Platform for Early Diagnosis and Follow-Up of Human Diseases: Architecture Design, Development and Analysis

Michele Sansone^{1*}, Jolanda Palmisani^{1*}, Alessia Di Gilio¹, Marirosa Rosaria Nisi¹, Lucia Pastore¹, Miriana Cosma Mazzola¹, Egidio De Benedetto², Gianluigi de Gennaro¹

¹ *Department of Biosciences, Biotechnologies and Environment, University of Bari, via Orabona, 4, 70126, Bari, Italy*

² *Department of Electrical Engineering and Information Technologies, University of Naples Federico II, 80125, Naples, Italy*

**co-first authors*

Corresponding Author's e-mail address: m.sansone24@phd.uniba.it

Summary

The growing application of sensor devices in the healthcare field is leading to the production of large datasets, whose effective management represents a critical challenge for current clinical processes. Although affected by high complexity, the heterogeneity of the collected data may enable healthcare professionals and researchers to extract relevant insights for the diagnosis and follow-up of diseases. Accordingly, this work proposes an integrated digital platform for the acquisition, management and analysis of multi-source clinical data, with the aim of improving healthcare and supporting patient treatment and monitoring.

Keywords: On-line breath analysis, Sensor technologies, Integrated data platform, Early Diagnosis, Telemedicine

Background, Motivation and Objective

The digital revolution in the healthcare system is redefining the way clinical data are collected, managed and interpreted [1]. The widespread adoption of electronic health records (EHRs) and sensor devices to detect and measure biomarkers in human biofluids has generated a large volume of clinical data (e.g., demographic and biometric information, diagnostic data, biomarkers and physiological signals) [2]. The use of target-analyte sensors in breath analysis represent a promising frontier for the diagnosis of several human diseases, due to their capability to detect and monitor the disease-related biomarkers pattern in a non-invasive and cost-effective way [3]. These devices enable the quantitative measurement of specific compounds in the exhaled breath, facilitating the identification of biomarkers associated with diseases such as lung cancer, myocardial infarction, and diabetes mellitus [4]. This wealth of information, combined with advances in computational methods, paves the way for new opportunities to support evidence-based decision-making [5].

A robust architecture is needed for the acquisition, management and analysis of data deriving from a variety of sensor devices and analytical platforms, bridging the gap between raw data and the healthcare professionals who need to interpret it [6]. Moreover, early diagnosis and long-term patient follow-up require scalable data

management solutions and the integration of analytical pipelines that combine statistical methods and machine learning techniques, with the aim of supporting the clustering of diseases and continuous monitoring of healthcare conditions.

Based on these considerations, this work describes the design of a data management platform focused on early diagnosis and clinical follow-up. The clinical platform aims to integrate data from multiple sensor technologies (e.g., target analyte-sensors, multi-sensor array devices) and analytical platforms such as GC/MS platform for the chemical characterization of human breath. The platform is intended to be the 'space' where all the clinical data can be also matched with other physiological parameters (e.g., heart rate, blood pressure, respiratory rate) and environmental data (e.g., air quality data), in the view of providing a comprehensive overview of the patient's health status. In line with the principles of telemedicine and exposomics, the overall goal of this study is to provide healthcare professionals a flexible solution that optimizes diagnostic accuracy, supports personalized patient care, and establishes the foundation for personalized healthcare applications.

Description of the System

The effective governance of healthcare data requires the development and implementation of a dedicated platform capable of the categorization,

integration and treatment of multiple-sources data, through the application of advanced statistical and AI-driven tools, complying at the same time with regulatory requirements, as the assurance of data security and the confidentiality of patient privacy. *Figure 1* illustrates the general architecture of the proposed platform.

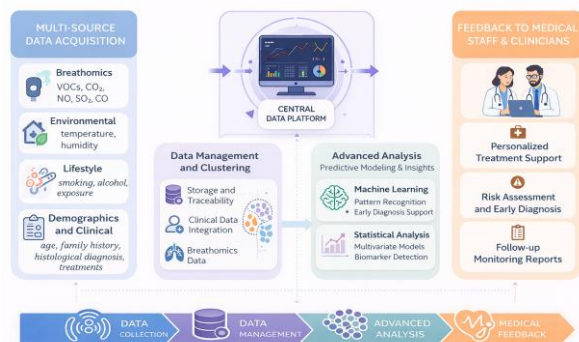


Fig. 1. Architecture of a multimodal clinical data platform for integrated data management, advanced analytics, and clinical decision support.

The central data platform employs a Role-based Access Control (RBAC) model, in which data access privileges are governed by security policies specific to each role. The architecture of central platform is structured in three sequential layers:

- **Presentation Layer:** the user interface was developed as a web app using Flutter, a cross-platform framework that enables deployment across heterogeneous environments from a single codebase. The system supports multi-source data acquisition, integrating breathomics parameters, environmental variables, and patient demographic and clinical data;
- **Application Layer:** the middle layer is involved in communication and clustering processes. The integration with backend services is handled via Firebase, a Backend-as-a-Service (BaaS) platform that provides real-time data synchronization, authentication, and cloud functions. Additionally, this layer manages data parsing prior to storage, clustering the data based on filters such as age, gender and smoking habits;
- **Data Layer:** data persistence is guaranteed by a NoSQL database, chosen for its flexibility in managing heterogeneous data. Prior to storage, the data is processed through a normalization pipeline, thereby rendering the dataset available for further analysis workflows. The information provided to healthcare professionals includes a longitudinal history of visits, diagnostic reports, laboratory test results, and multi-sensor physiological measurements.

The proposed platform aims to provide an innovative contribution to the spectrum of tools supporting patient diagnosis and follow-up. The modularity of the architecture, the capability to manage large volumes of heterogeneous data, and flexible communication between the operator and healthcare professionals are distinctive features. Further developments in the near future will be focused on the validation of the developed platform in the clinical practice to manage data from populations of individuals affected by chronic and oncologic diseases as well as of subjects in therapeutic follow-up.

Acknowledgements

The authors thank Dr. Giovanni Liso and all the team of the enterprise MAV Reality srl (Italy) for the precious support in the development of the architecture of the digital platform.

This work is contextualized in the framework of the project INNOVA – Italian Network of Excellence for Advanced Diagnosis (PNC-E3-2022-23683266), funded by the Italian Ministry of University and Research under the National Complementary Plan to the PNRR.

References

- [1] Nazir, S., Khan, S., Khan, H. U., Ali, S., Garcia-Magarino, I., Atan, R. B., & Nawaz, M. (2020). A comprehensive analysis of healthcare big data management, analytics and scientific programming. *IEEE Access*, 8, 95714-95733; doi: 10.1109/ACCESS.2020.2995572
- [2] Harb, H., Mansour, A., Nasser, A., Cruz, E. M., & de la Torre Diez, I. (2020). A sensor-based data analytics for patient monitoring in connected healthcare applications. *IEEE Sensors Journal*, 21(2), 974-984; doi: 10.1109/JSEN.2020.2977352
- [3] Hooshmand, S., Kassanos, P., Keshavarz, M., Duru, P., Kayalan, C. I., Kale, I., & Bayazit, M. K. (2023). Wearable nano-based gas sensors for environmental monitoring and encountered challenges in optimization. *Sensors*, 23(20), 8648; doi: doi.org/10.3390/s23208648
- [4] Kaloumenou, M., Skotadis, E., Lagopati, N., Efsthopoulos, E., & Tsoukalas, D. (2022). Breath analysis: a promising tool for disease diagnosis—the role of sensors. *Sensors*, 22(3), 1238; doi: doi.org/10.3390/s22031238
- [5] Lu, S., Yang, J., Gu, Y., He, D., Wu, H., Sun, W., ... & Guo, C. (2024). Advances in machine learning processing of big data from disease diagnosis sensors. *ACS sensors*, 9(3), 1134-1148; doi: 10.1021/acssensors.3c02670
- [6] Emeakaroha, V. C., Fatema, K., Healy, P., & Morrison, J. P. (2015). Contemporary analysis and architecture for a generic cloud-based sensor data management platform. *Sensors & Transducers*, 185(2), 100.