Health Mining: a new data fusion and integration paradigm

Alessandro Di Stefano[(1)(2)], Aurelio La Corte[(1)(3)], Marialisa Scatá[(1)(4)]

(1) Department of Electrical, Electronics and Computer Science Engineering
University of Catania, viale A. Doria 6, 95125 Catania (Italy)
(2) alessandro.distefano@dieei.unict.it
(3) lacorte@dieei.unict.it
(4) lisa.scata@dieei.unict.it

Keywords: health management, data fusion, data integration, bio-inspired approach, data mining, social networks, internet of things.

Abstract. One of the future challenges of information and communication technology (ICT) is particularly targeted at public health and clinical interventions management. The aim is to make smart the management and the processes that characterize this context to optimize the procedures and provide technological support. In this paper we propose the starting point of a new approach through architecture to introduce methodologies for fusing, integrating and drawing inference from a plurality of clinical, social and Internet of Things data, extracted from different sources. Health Mining is the central node of the proposed architecture and serves as a knowledge base for bio-inspired systems to support and perform the diagnostic process; this approach is better than the one applied with a plurality of disgregated data.

1 Scientific Background

Medical statistics, clinical epidemiology, decision analysis and health economics, are different disciplines that, in the smart health context, are becoming more increasingly connected in a fruitful and compelling way. Nowadays, the problem is that health care should be rationed from different points of view. The topic of innovation in analyzing health data, how to classify data from clinical and biomedical observations and studies, sensors measures, social behavioral dynamics, needs for a new analysis and methodologies to get a complete and detailed overview of health context. The plurality of informative resources available does not allow for fast processing, and therefore it would be useful to define a single paradigm able to simplify the procedures optimizing efficiency of the subsequent inference analysis, decision-making and management processes. With the advance of information and communication technologies (ICT) in healthcare context [1], multiple types of data provide many complementary perspectives of a single or different correlated aspects, and highlight the need for algorithms able to unify these heterogeneous resources. Many approaches have been developed over the last several years for improving health management, and in particular robustness, accuracy and support process to diagnostic thecniques [2]. The main issue is to combine relevant information in the most efficient way and obtain a mining process. The aim is infact to mine awareness and knowledge, that is, a comprehensive understanding of the healthcare issues, social dynamics, clinical and biomedical observations and experiments, all included in a single-complex-data. The following features represent a background of the different aspects, treated in the proposal. Clinical Data, statistics, classifications and databases [3] provide and maintain the traditional approach to the collection of bio- medical information. An IoT model is playing a key role in several scenarios [4], one of these is healthcare and wellness. There are several domains which will be impacted by emerging Internet of Things. The sensor information collected are the key for the ubiquitous healthcare.With body area sensors, home and social moni-

toring we gather context-aware data about patients and thereby reducing hospitalization costs through early and fast interventions and treatments. Social Networks are populated by nodes which interacts and share social objects; the ties and the structure of this kind of network is ruled by behavioral dinamics that influences and acts in the social contagion process [5], useful to mine knwoledge about nodes and to predict health phenomena [6]. Traditional Data Mining Process represents techniques used to explore database and detect unknown patterns useful to build predictive models [7]. In our proposal we change the mining process, introducing a layered approach, capturing different aspects of the collected data from several sources. The three data layers, defined in the architecture, create a reality mining in health context, considering not only the clinical data, but also social and IoT information. Finally, the Bio-Inspired approach [8] provides methodologies and algorithms to redifine and optimize analysis process and improve ICT design and management systems.

2 Materials and Methods

One of the targets of health mining consists of expressing, organizing and storing all the tipologies of data and information linked to health, in order to get an exchange and sharing of information and promote the fusion and integration mechanisms. This leads to an improved decision support system in diagnostics, able to support different types of knowledge at distinct layers and to simplify and optimize the management in a smart health context. This health mining paradigm allows to improve the quality of health services and optimize the diagnosis through both a fusion and an integration of information from different sources, solving the problem related to heterogeneity and creating a new type complex data type. Considering this new integrated data in smart health, ICT interventions can be applied so as to produce an increasingly improvement and calibration of data. The various data sources produce different types of data, including clinical, such as the history of disease, information derived from different inspections and exams, diagnosis, treatment planes, medical data, e.g. spectral analysis of genetic mutations for the key genes of an individual, genomic analysis related to drug sensitivity. All this information have to be progressively changed according to the updated medical knowledge, derived also from other sources different from the medical and clinical ones, taking account of data derived from social networks and IoT devices (e.g. on-body health monitoring). Therefore, it is difficult to model this fusion and integration mechanisms and create a unique complex data that contains all these types of information. In order to get this data mining model which accounts for all these different aspects, it is crucial to think about a unique multi-dimensional and multi-layer paradigm, providing a description for each type of data, applying a data fusion at each layer and an integration related to the different dimensions, getting a unique overall knowledge. What makes even more complex these fusion and integration techniques is the correlation and the statistical dependence of these different informative units, since e.g. the social networks creates links between individuals and a social contagion related to some of the risk factors (or morbidities), such as obesity spreading [5], or more the intra-body sensors shed light on some aspects at a microscopic level which affect the social behaviors and clinical analysis of the individual at a macroscopic layer. If from one hand, these dependencies make the analysis more complex, as we have to consider different dimensions in the integration, on the other hand allow to enrich the diagnosis process and the subsequent treatment and prognosis with useful information. Therefore, these latter processes represent another data source, further populating the data set. Data fusion, applied at each layer, is a process by which data from several sources (e.g sensors if we consider the IoT layer), probably of different types, is aggregated in a way that allows more efficient or more accurate inferences; data fusion may entail some reduction, in fact fused data set is smaller, with less noise, and more predictive capacity

compared to any single entity [9]. Data integration is a process in which heterogeneous data is retrieved and combined as an incorporated form and structure. Data integration allows different data types (such as data sets, documents and tables) to be merged, so that it represents an inter-layer integration, which exploits the fusion mechanism applied at each layer [10]. Therefore, the fusion mechanism at each layer, and the integration mechanism between layers allow to reduce the complexity linked to the diversity of data, and provides more flexibility to the data mining system.

3 Results

The following scheme summarizes the proposed paradigm:

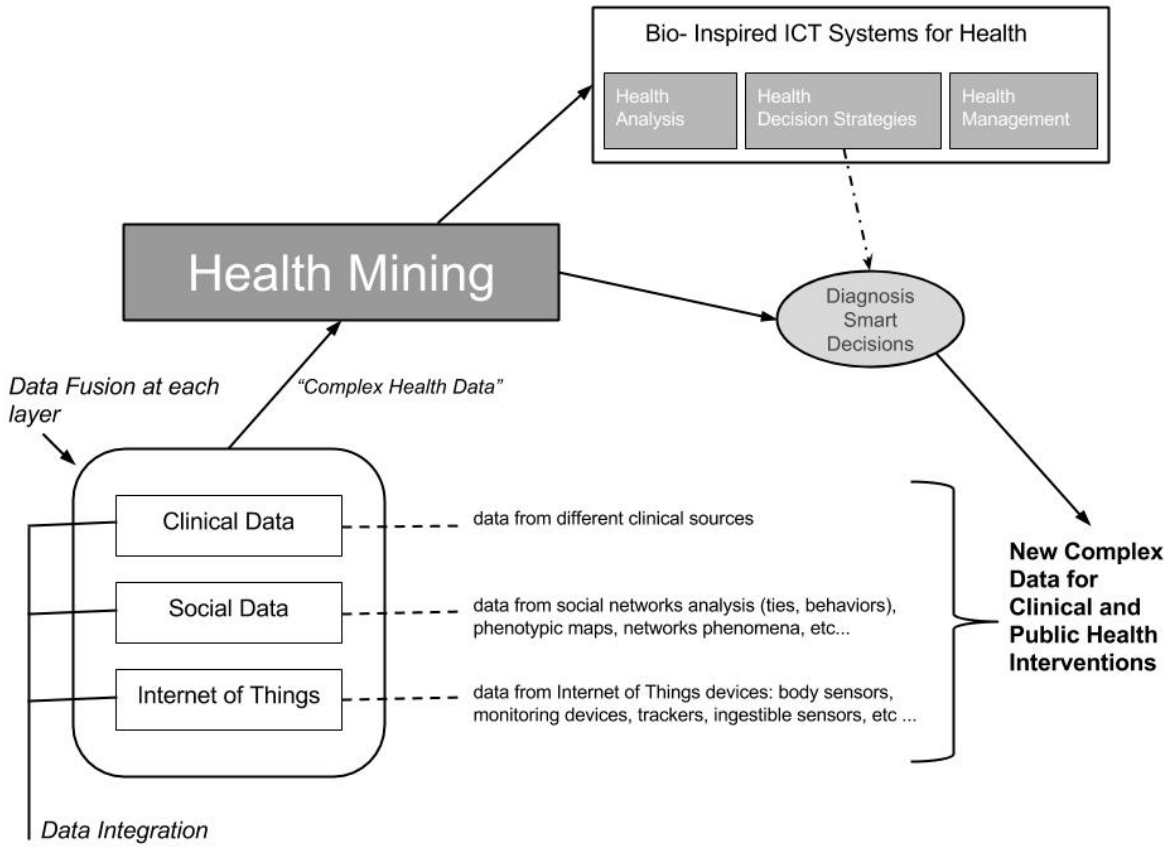

Figure 1: Data Fusion and Data Integration for Health Mining.

In Health Mining Architecture, the idea is to integrate in a unique structure, called "Complex Health Data", data from different sources and of different nature, fused at each layer. In particular, we consider three different macro-types of data: clinical data, social data and IoT data. This structure represents the basic block on which it is possible to get an overall mining process on health context. The Health Mining Process constitutes the smart knowledge that can be exploited from the Bio-inspired ICT systems for Health. These systems includes three different aspects represented by three different blocks in the scheme, that are: Health Analysis, all the procedures of analysis of data in health context, Health Decision Strategies, all the processes that support the decision-making for healthcare and the introduction and tests of actions and strategies in the same context, Health Management is the final block that considers all the features, medical, social and economic aspects observed and estimated. Diagnosis Smart Decisions are a building feature of the Health Decision Strategies, and it can gather the knowledge directly from mining process, to make the procedure smart and at the same time to improve and speed up the steps that lead to diagnosis.

4 Conclusion

The proposed architecture represents a first step to design a new mining model to improve health management. One of the major issues related to health context is the plurality and the heterogeneity of data collected about patients on different health aspects. Considering different kinds of data, the proposed solution introduces intra-layer fusion mechanisms, on data related to the same type of sources, and inter-layer integration mechanisms, on data related to different types of sources. The resulting Complex Health Data populate and represent the basis on which the Health Mining process is builded. The Health Mining process is then exploited from the different Bio-Inspired Systems for Health, which allow to speed up, make smart and improve the overall diagnosis process, considering also medical, social and economic aspects.

References

[1] Amina Jama Mahmud, Ewy Olander, Sara Eriksn and Bo JA Haglund."Health communication in primary health care - A case study of ICT development for health promotion". *BMC Medical Informatics and Decision Making*, vol.13, no. 17, 2013.

[2] Margaret Kubick. "An Uncertain Future: The Impact of Medical Process and Diagnostic Method Patents on Healthcare in the United States". *Northwestern Journal of Technology and Intellectual Property*, vol.9, no.3, 2010.

[3] ICD9DATA Medical Coding Reference. *website: www.icd9data.com* , available at 2014.

[4] D. Miorandi, S. Sicari, F. De Pellegrini, I. Chlamtac. "Internet of Things: Vision, Applications and Research Challenges". *AD HOC Networks* , vol.10, no.7, pp. 1497-1516, 2012.

[5] Nicholas A. Christakis, James H. Fowlerc. "Social contagion theory: examining dynamic social networks and human behavior". *Statistics in Medicine*, vol.32, no. 4, 2013.

[6] S. Kitchovitch, P. Lió. "Community Structure in Social Networks: Applications for Epidemiological Modelling". *PLos One* , vol.6, no.7, 2011.

[7] Hian Chye Koh and Gerald Tan. " Data Mining Applications in Healthcare". *Journal of Healthcare Information Management*, vol.19, no.2, 2005.

[8] Falko Dressler, Ozgur B. Akan. " A survey on bio-inspired networking". *Computer Networks Elsevier* , 2010.

[9] Georgia Tsiliki, Sophia Kossida. "Fusion methodologies for biomedical data". *Journal of Proteomics* , vol.74, 2011.

[10] D. Teodoro, R. Choquet, D.Schober, G. Mels, E. Pasche, P. Ruch, C. Lovis. "Interoperability driven integration of biomedical data sources". *Studies in Health Technologies and Informatics*, vol.169, pp. 185-189, 2011.