

INSTAMED: an Integrated Platform for the Advanced Automation of Diagnosis in Precision Medicine

Tommaso Di Noia*, Corrado Fasciano*[†], Domenico Lofù*[†], Giulio Mallardi*[†],
Graziano Pappadà[†], Corrado Tatulli[†] and Felice Vitulano[†]

*Dept. of Electrical and Information Engineering (DEI), Politecnico di Bari, Bari (Italy),
e-mail: {firstname.lastname}@poliba.it

[†]Innovation Lab, Exprivia S.p.A. – Via A. Olivetti 11, Molfetta (I-70056), Italy,
e-mail: {firstname.lastname}@exprivia.com

Abstract— In recent years, Precision Medicine has allowed the identification of specific therapies and cures for a single patient based mainly on the genetic and phenotypic characteristics that distinguish him from other patients but also taking into consideration elements such as lifestyle, the environment where they live and any other element that may affect the state of health and the effectiveness of the therapy. The INSTAMED project aims to create an integrated hardware and software platform for the standardization and advanced automation of the processes of diagnosis and treatment in a personalized medical perspective.

Index Terms—Healthcare, Recommender System, Big Data, Machine Learning, Iot, Precision Medicine

I. INTRODUCTION

The study of the phenotypic characteristics of a patient plays a fundamental role in Precision Medicine. Thanks to the clinical phenotype, it is possible to acquire intrinsic knowledge that is not always identifiable even by experts such as doctors. The phenotypic characteristics of a single individual and related analyzes allow us to verify how patients with the same pathology but different phenotypic characteristics can react differently to a given therapy or treatment. From a medical point of view, the phenotype is understood as the variation of the morphological and physiological characteristics concerning a behavior considered normal. The aggregation of both genomic and environmental data allows doctors to monitor, over time, these changes through tests or diagnoses [1]. The collection and analysis of patient clinical information to identify common patterns and characteristics from a cohort of patients take the name of Deep Phenotyping. The inclusion of "Deep" terminology derives from the level of depth that you want to adopt for data analysis. The greater the level of quantity and complexity of the data collected, the greater the probability of achieving the presupposed research objective [2].

II. ENHANCING PRECISION MEDICINE APPROACH

A. EHRs in Deep Phenotyping

EHRs have become the primary data source for healthcare applications whose technologies exploit methodologies and approaches in the fields of Artificial Intelligence, Internet

of Medical Things (IoTM) [3] and much more. As previously indicated, the complexity of Deep Phenotyping depends mainly on which data from an EHR to use and how to manage and analyze it. Every single clinical feature that is to be exploited has peculiarities that increase the difficulty of implementing an algorithm in the health sector. The following are the main difficulties and precautions when using EHRs in a Data Analytics contest:

- 1) *Privacy*: EHRs contain medical data that are considered sensitive and, as a result, are either unavailable or subject to restrictions.
- 2) *Availability*: the protection of sensitive data makes it difficult to find useful data for the creation of Precision Medicine algorithms that are highly dependent on the data in question.
- 3) *Knowledge of the sector*: to implement Precision Medicine algorithms based on Deep Phenotyping, different professionals are required who can recognize the application business in which the problem arises and who can reconcile and relate the clinical, biological and IT aspects studied by a population of reference.
- 4) *Missing data and incomplete information*: the data on the EHRs may be scattered and not always considered mandatory, therefore completely absent in the medical record. An important step in the life cycle of an application in Data Science is the Preprocessing phase and in this context, the evaluation of missing data.
- 5) *Variety of data*: the presence of structured and unstructured data makes the Preprocessing phase, specifically the Data Cleaning phase, difficult. In medical applications, the possibility of encountering textual data is frequent. Therefore, the use of Text Mining and Natural Language Processing techniques is necessary.
- 6) *Information dependent on time series*: a patient's data is collected in time to allow doctors to constantly monitor the progress of a given pathology. Time-Series Analysis increases the complexity of a predictive model as different methodologies are used for them compared to traditional approaches.

The problems analyzed constitute a fascinating challenge that, if carried out successfully, increases the quality of a model equipped with Artificial Intelligence.

B. Clinical decision support system for personalized therapy

By exploiting the enormous amount of information obtainable from the health system, computer systems have been developed in recent years that can assist doctors during the clinical decision-making process. These systems are called Clinical Decision Support Systems (CDSS). CDSS can be classified into various main categories of CDSS and the main ones are based on the methodology of approach and the application context. In fact, the decision-making power of CDSS is not limited to the use of personalized therapy but extends to numerous health care areas and, at the same time, linked to precision medicine such as for example for decision support for diagnoses, prediction of conditions of patients [4]. The most popular approaches are Knowledge-Based, Data-Driven, and Recommender Systems.

III. INSTAMED ARCHITECTURE

The aim of this work is to illustrate Artificial Intelligence techniques applicable to Precision Medicine, starting from the state of the art up to the creation of the algorithms underlying the *eProfile* and *eTherapy* modules, which will be integrated into the *Clinical Collaboration model*. The *eProfile* module aims to identify patients suffering from oncological or rare diseases, similar to each other on the basis of the genomic and phenotypic profile. The *eTherapy* module is a clinical decision support system (Clinical Decision Support System) which aims to provide support to the doctor for the assignment of personalized and specific therapy for the patient examined, exploiting the correlation between the clinical phenotype identified by *eProfile* with medical-scientific sources such as medical records, best practices such as Diagnostic Therapeutic Care Pathway (DTCP) [5] and articles of scientific literature.

eProfile is an application that aims to identify a patient who, based on his clinical and genomic characteristics, is placed in a specific subgroup containing similar patients. Patient grouping is based on clinical phenotype. The problem that you want to solve with *eProfile* is of a predictive nature and this analysis is carried out by building an architectural pipeline typical of a Data Science project. The methodology chosen to divide all the various phases that lead to the creation of *eProfile* is the CRISP-DM standard (Cross Industry Standard Process for Data Mining). CRISP-DM is an architectural model that allows you to better organize complex projects, dividing them in a simple and accurate way.

eTherapy is an application made available in the clinical collaboration platform in which the doctor is able to prescribe a personalized therapy for the patient having available a series of treatments that can be extracted from:

- Clinical Pathways or best practices
- Similar clinical cases
- Scientific literature
- Targeted Therapy

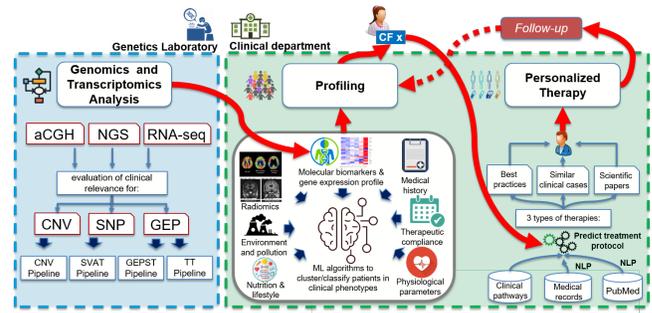


Fig. 1. INSTAMED Architecture.

Drugs deriving from medical macro-areas must be associated and consistent with the clinical phenotype of the patient identified by *eProfile* which, together with the patient profile, is used both as filtering data in drug research and as an input tool for Artificial Intelligence techniques. In fact, even if the doctor has the possibility to manually define the therapy, *eTherapy* automatically provides a summary of the definitive therapy using a system based on Recommender System, capable of recommending the most suitable treatments for the patient’s clinical phenotypic. The following paragraphs will illustrate how *eTherapy* works, in particular how it will use the clinical phenotype as a support tool in the four medical areas and how the Recommender System will be created. In both cases, the CRISP-DM methodology already exposed in *eProfile* will be adopted as a pipeline. *eProfile* and *eTherapy* are two applications integrated into the clinical collaboration platform designed to put into practice the constructs of Precision Medicine, making use of Artificial Intelligence techniques, towards patients suffering from oncological diseases, rare diseases and corneal diseases.

ACKNOWLEDGMENT

This work was partially funded by the European Union, Horizon 2020 research and innovation programme, through the ECHO project (grant agreement no 830943), by the Italian project INSTAMED (PON 2014–2020 FESR, project code F/080015/02/X35).

REFERENCES

- [1] P. N. Robinson, “Deep phenotyping for precision medicine,” *Human Mutation*, vol. 33, no. 5, pp. 777–780, 2012.
- [2] I. R. König, O. Fuchs, G. Hansen, E. von Mutius, and M. V. Kopp, “What is precision medicine?” vol. 50, no. 4, 2017.
- [3] A. Paziienza, R. Anglani, G. Mallardi, C. Fasciano, P. Novello, C. Tatulli, and F. Vitulano, “Adaptive critical care intervention in the internet of medical things,” in *2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, 2020, pp. 1–8.
- [4] H.-J. Lai, T.-H. Tan, C.-S. Lin, Y.-F. Chen, and H.-H. Lin, “Designing a clinical decision support system to predict readmissions for patients admitted with all-cause conditions,” *Journal of Ambient Intelligence and Humanized Computing*, 01 2020.
- [5] C. Ardito, F. Bellifemine, T. Di Noia, D. Lofù, and G. Mallardi, “A proposal of case-based approach to clinical pathway modeling support,” in *2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, 2020, pp. 1–6.