Personalized Decision Support System to Enhance Evidence Based Medicine through Big Data Analytics

Yelena Yesha[1], Vandana P. Janeja[2], Naphtali Rishe[3], Yaacov Yesha[1]
[1]Computer Science and Electrical Engineering,
[2]Information Systems,
University of Maryland Baltimore County
[3] School of Computing and Information Sciences, Florida International University

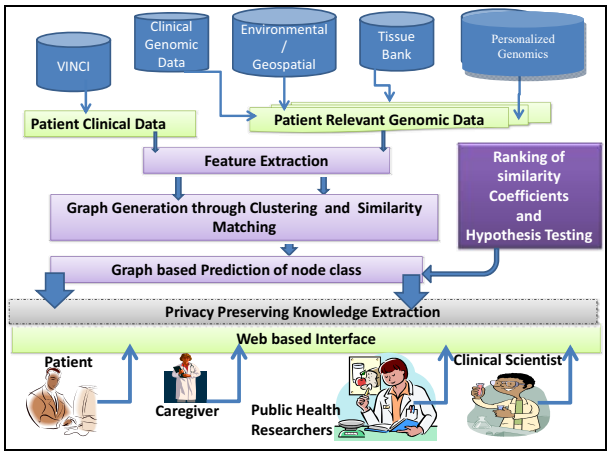
***Abstract*—NG Health IT and UMBC are collaborating together to advance the fields of healthcare and bioinformatics, with an emphasis targeting Personalized Medicine. As progress in biomedical analysis and personalized medicine provide new sources and levels of information about genomics and other 'omics, problems of sorting through, integrating, and presenting this information to clinical practitioners to allow them to reach sound, actionable, conclusions represents a significant challenge given the nature of the data which is heterogeneous and large. This project will provide significant value for the Veterans Administration (VA) and MHS systems by providing improved access to care and improved clinical decision making utilizing clinical, genomic, proteomic, and other 'omic information. The work being done focuses on the data access and analysis of multiple sources of healthcare data.**

I. BIG DATA ANALYTICS FOR PDSS

Patient data can be present in Clinical notes, genomic data sources, environmental and geospatial data sources, tissue banks and personal genetic data. A complete view of the patient health can be achieved when relevant data from all these sources are extracted and modelled in a personalized manner. We propose a Personalized Decision Support System (PDSS) to enhance personalized medicine though big data analytics. This requires addressing several challenges:

- Extracting feature vectors defining a patient not just from the structured Electronic Health records but also the unstructured clinical notes associated with them. This should address issues such as Negation, conditional, Uncertain statements and also medical and family history statements which make the clinical notes processing challenging for machine learning algorithms.
- Extracting patient relevant data from Clinical Genomic data such that links can be established between patient clinical factors and the genomic factors.
- In several cases, the patient health is impacted by the environmental factors, which are present in the patients geospatial proximity. These correlations provide key insight into the patient health outcomes but are generally ignored in a traditional healthcare setting.
- Several specific analyses that are done on the patient tissue samples or individual genotype may also play a key role in understanding the personalized medicine outcomes.

To address these challenges we propose a framework as shown in the figure, which encompasses several of these components in a privacy-preserving manner, such that the knowledge can be provided to individual patients, caregivers, public health researchers and clinical scientists.



The framework engineered in a high performance computing environment cuts across several distinct datasets. The components of the framework include feature extraction from clinical data, features from patient relevant genomic data along with the environmental factors. This helps extract a relevant set of features, which define the patient traits across multiple datasets. These feature vectors are used to generate a similarity based graph using clustering and similarity matching. A query from the users such as patient or caregiver can identify other similar patients matching the query. In addition, prediction of a link between two patient nodes may potentially facilitate in disease identification, which may not have been diagnosed based on just one set of attributes.

978-1-4799-5701-9/14 $31.00 © 2014 IEEE
DOI 10.1109/ICHI.2014.71