On the use of clinical based infection data for pandemic case studies

Giuseppe Tradigo
Ecampus University
Novedrate, Italy
giuseppe.tradigo@uniecampus.it

Patrizia Vizza
Bioinformatic Unit
University Magna Graecia of Catanzaro
Catanzaro, Italy
vizzap@unicz.it

Gabriel Gabriele
Infectious Diseases Unit
University Magna Graecia of Catanzaro
Catanzaro, Italy
email address or ORCID

Maria Mazzitelli
Infectious Diseases Unit
University Magna Graecia of Catanzaro
Catanzaro, Italy
email address or ORCID

Carlo Torti
Infectious Diseases Unit
University Magna Graecia of Catanzaro
Catanzaro, Italy
email address or ORCID

Mattia Prosperi
Epidemiology Department
University of Florida
Florida, USA
email address or ORCID

Pietro Hiram Guzzi
Bioinformatic Unit
University Magna Graecia of Catanzaro
Catanzaro, Italy
hguzzi@unicz.it

Pierangelo Veltri
Bioinformatic Unit
University Magna Graecia of Catanzaro
Catanzaro, Italy
veltri@unicz.it

2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 978-1-7281-6215-7/20/$31.00 ©2020 IEEE DOI: 10.1109/BIBM49941.2020.9313469

***Abstract*—Epidemiological models are relevant to study and analyze clinical as well as environmental and behavioural data, useful to support health studies. The target is to perform epidemiological analysis producing fast and reliable data access useful to guide prevention and curing processes. This is currently true in pandemic emergency as the current Covid-19 context. Epidemiological models should support in the early identification of pandemic phenomena and in making available data set for studying more accurate drug-based strategy for vaccines or virus containment.**

In this contribution we present an epidemiology database which integrates different types of clinical data to support research, follow-up and patient monitoring. The idea starts from an hospital databases cooperation integration where virus available data have been integrated to support statistical based studies. Starting from an available database containing 5 years data of infection related viruses (such as HPC, hepatitis) and patient anonymous data, the proposed system provide an integrated data access able to (i) extracting data filtered by means of clinical hypothesis based on patient profiles, environment and drugs and (ii) allowing to build large scale geographical data mappings in order to study correlations among chronic infection diseases and their relations with upcoming pandemic phenomena. Even if the application is in its infancy, the application is relevant with high very important applications.

***Index Terms*—Epidemiology, pandemia, data integration.**

I. Introduction

Recent Covid-19 (COronaVIrus Disease 19), known also as SARS-CoV-2, pandemic virus diffusion showed the weakness of containment measurements and health structures. Fast diffusion and non-rapid identification of coronavirus have resulted in huge numbers of infected people and deaths as well as in the saturation of available beds in Intensive Care Units (ICUs), caused by a non-predictable events related to virus spread. Epidemiology models represent valid solutions to support physicians, epidemiologist and government in management of healthy related events in preventing and controlling infectious disease and possible critical situations [1]. Risk factors (e.g. environmental consequences, lifestyles and hereditary characteristics) play an important role in these models because their analysis and correlation with clinical data allow to extract relevant and useful information to identify risk class and plan the correct interventions [2], [3]. In this context, Geographical Information System (GIS) integrating infection data with geographical information represents an added value in studying and controlling pandemic events [4], [5]. Many web based geographic systems have been proposed to map Covid-19 data containment as well as web based graphical visualization of virus diffusion [6]. Figure 1 represents an example of data mapping and diffusion around the world gathered from [7].

The collection and integration of environmental data with epidemiological, clinical and treatment data (see Figure 2) can be used by health workers to monitor infection diffusion. For instance, by studying correlation among clinical data, blood analytes and to treatments results or containment measures should be designed.

Epidemic data may be used with (anonymized) clinical data to perform online analysis and monitoring [8], [9]. Using of environmental data with virus diffusion has been treated in many studies [10], [11]. For instance, authors in

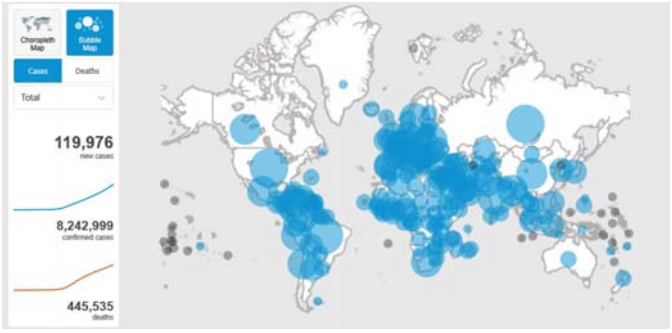

Fig. 1. Coronavirus disease COVID19 map in the world from WHO (World Health Organization) website.

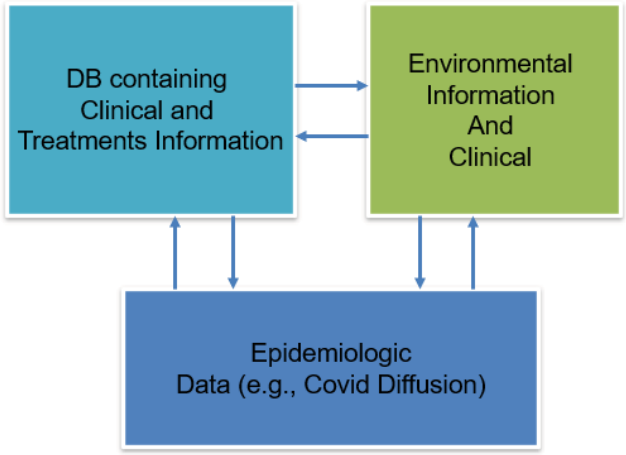

Fig. 2. Clinical, treatment, environmental and epidemiological data integration.

[12] developed malaria maps by integrating socio-economic, epidemiology and geographical dimensions of three eastern districts in India through the use of different GIS maps. The results are useful to policy makers to produce cost-effective measures for malaria control. Another example of integrated system for epidemiological studies is proposed in [13] and [**?**]. This system is used for epidemiological research by clarifying the clinical spectrum and natural history of diseases and the effectiveness of treatment. The study in [14] aimed to disclose the epidemiological and clinical characteristics of Severe fever with thrombocytopenia syndrome (SFTSV) infection in China by implementing an integrated clinical database. It contains clinical manifestations, routine laboratory tests of acute infection, hospitalization duration and disease outcome. Authors in [15] proposed a system to monitor in real-time epidemiological indicators, entomological indices, social surveys and environmental variables aiming to identify variables that favor virus transmission. Concerning recent pandemic Covid-19, many studies have been presented to furnish relevant information useful to study and characterize this viral infection [16], [17]. For example, authors [18] in used daily epidemiological data from surveillance about COVID-19 in Honduras to calculate the rates of incidence and its developed at national, departmental, and municipal levels by implementing GIS-based maps.

The proposed integrated platform aims to support research, follow-up and patient monitoring in epidemiological context. An important function of the platform is to analyze the possibility of relating therapy use and correlation with viruses or comorbidities. Starting from an existing raw database used for clinical studies in chronic hepatitis disease, we design and implemented an integrated system for epidemiological analysis of clinical data relating to infectious diseases gathered from different health centers. The main goal is to create a platform able to support targeted epidemiological studies taking into account geographical areas and specific diseases. The presented contribution, even if in a preliminary version, represents an important reference in the Italian clinical status. Mapping resources, virus diffusion and environmental data gathered from external sources (such as WHO) as well as environmental data (such as industries presence or simply weather type) may help in performing: (i) simulation for therapies applications, (ii) testing or supporting hypothesis on viruses diffusion, (iii) applying and tuning statistical experiments on clinical data.

II. Methods

An epidemiological platform has been developed to support clinical centers in studying and evaluating of viral infections and their geographical correlations on a cohort of population. The system consists of a web-based application which interacts with a clinical database and provides statistical support for the patient cohort. The architecture of the proposed system is shown in Figure 3.

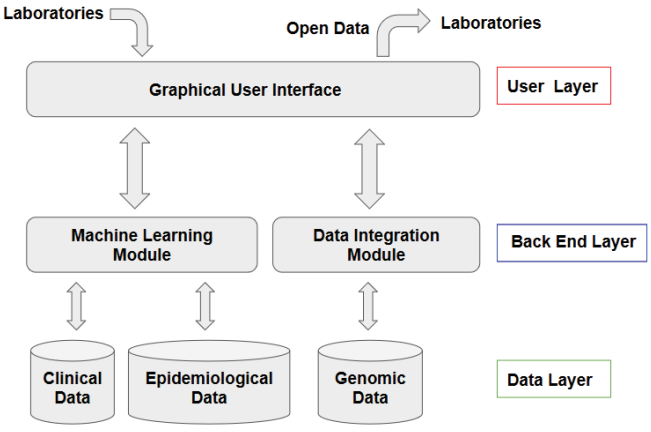

Fig. 3. Architecture of the proposed system.

The proposed architecture has three layers. The first layer (User Layer) is responsible for the data collection by external laboratories and for the extraction of data using open data model. The second layer (Back End Layer) deals with data integration and machine learning. It is responsible for integration of data of different laboratories into a single common model. It also contains a machine learning module for data

analysis to support research. Finally, the layer (Data Layer) is responsible for data storage and it is implemented using a relational database. Data stored are clinical, epidemiological and genomic which are useful to perform epidemiological studies. From a user point of view, the application may be modeled as reported in Figure 4.

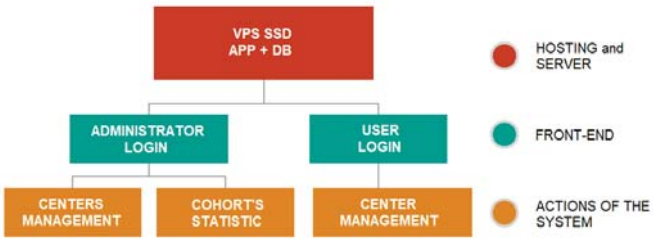

Fig. 4. Block diagram of the platform.

The idea behind the development of the web based system is to define a mechanism to allow analysis of epidemiology data. For instance, gathering data from different clinical and environmental data allows to perform studies on chronic infectious related diseases.

A. Clinical database

A clinical database has been implemented to collect data coming from different medical Italian centers. It contains information relating to patients and their follow up. Clinical information have been integrated with geographical data to evaluate possible correlation between environmental factors and the onset of diseases. Moreover, a comparison between patients with similar clinical features but different locations can also be performed aiming to identify possible risk factors which contribute to the development of diseases. Each center can access to the database and extract information about a single set of patients or about the entire cohort of patients. Moreover, simulation on clinical data set can be performed on subset of clinical data. The core database is implemented using a MySQL instance whose data structure is partially reported in Figure 5.

Patients table, as well as Exams and Diagnosis, and therapies table contain information regarding geographic issues. Also, the system has been enriched to treat patients from extra European countries thus which any patient may move through countries by having access to their personal data, where for instance Italian ones can use their social security number. Information about residence, nationality and region are also inserted and they are used to map the patient. The mapping operations allow to locate a individual in a well-defined geographical area for research epidemiological. In the patients table, other relevant information are risk factors, cause of death and infection classification. Tables Exams and Diagnosis contain standard information which can be used to study clinical and virus diffusion cases through the several health centers participating to the project and that host patients. Also clinical treatments across different cases are also mapped

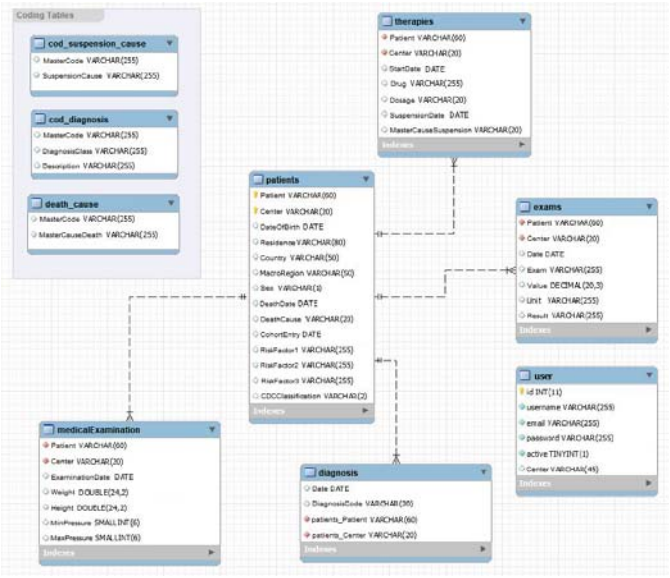

Fig. 5. (Extracted portion of) Entity Relation Database Instance.

and analyzed. In table Treatment, data about type and name of the drug, dosage, start and end date of treatment and possible reason for suspension are included.

B. System Access and Application

A web-based application has been implemented by using PHP 7, HTML5 and CSS3 by offering a simple user interface, that is also able to import data from laboratories (e.g., used for Covid-19 swab tests results), as well as administrative information. Users can be clinicians, data manager or administrator from single health structure. Super user

figures have been designed to fulfill large scale access to information of all Italian centers participating to the project. The web-based application allows to perform a patient survey by selecting specific features and parameters. Moreover, clinicians can access only to the list of patient of his medical center. The administrator, instead, can access to the entire cohort of patient of all medical centers. This is very important to investigate the behavior of an infection disease on a regional or national basis.The most common functionalities concern (i) the addition, modification and removal of a patient, (ii) the visualization of clinical and geographical patient information, (iii) the collection and distribution of data within a virtual map to a best and complete vision of a possible epidemic or pandemic. For each disease, the system generates a map to visualize the degree of infection and the number of infected patients for for each region. Another important functionality regards the statistical and machine learning analysis of data. It is available only for the administrator user and allows to perform statistical evaluation also with the possibility to engage with external statistical systems which provide environmental or clinical data.

III. Application for pandemic data

The system has been defined on top of several Infectious Diseases databases from some Italian clinical centers. Each center provide its own database where clinical treatments and management are reported. The system presents a common simple user interface, where each user belonging to the VPN may access by using login page (see Figure 6). An example of patient addition is shown in Figure 7.

Fig. 6. Login page.

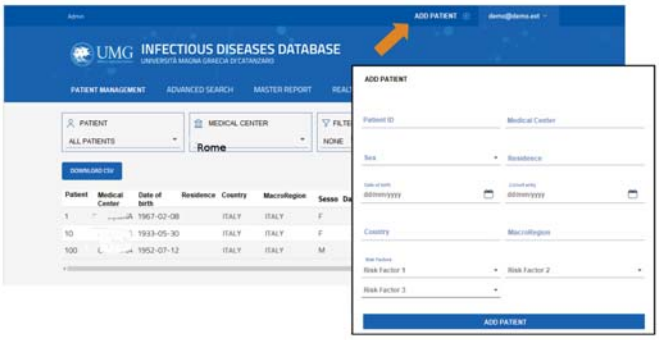

Fig. 7. Patient addition.

Clinicians may perform statistical analysis on subset of clinical data as well as on therapies results. For example, Figure 8 reports number of patient for each medical center, the number of patient in the cohort, the number of patient in follow-up and the patient distribution by sex.

A functionality based on geographical mapping (Italy at regional scale) has been performed, through geodecoding of data and instantaneous display of patients located in the various regional or continental areas. The collection and distribution of this data within a virtual map facilitates the complete vision of a possible epidemic or pandemic. For example, Figure 9 reports the number of infected patients with COVID-19 in

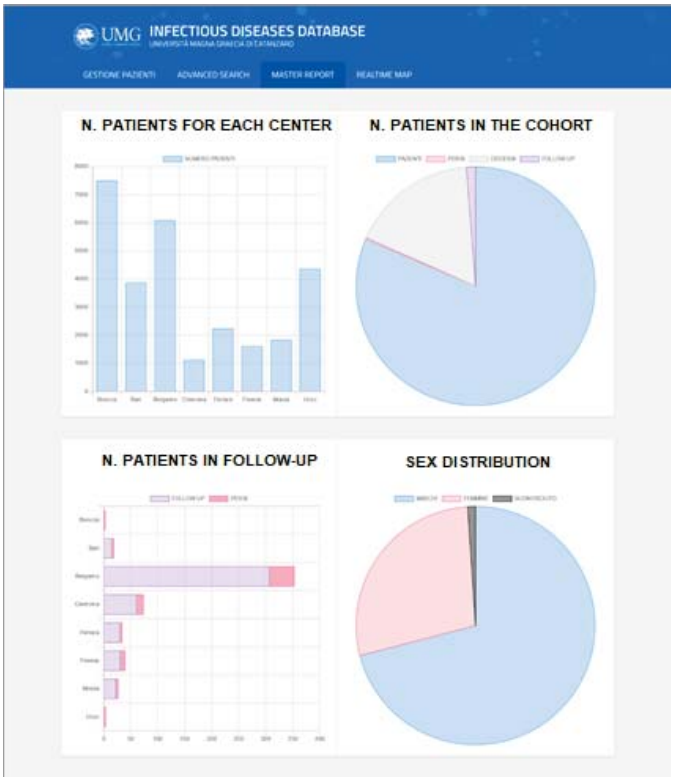

Fig. 8. Preliminary results.

Italy in a preliminary Covid-19 diffusion period (beginning of February 2020).

The framework allows to compare clinical data as well as drugs information to identify features for epidemiology information.

IV. Results

The system scales on 8 health infection hubs (health structures) through Italian territory, mapping latter 5 years with more than 100 thousands entries. The system is particularly useful to monitor data useful to control therapies and follow up. For instance it is possible to map patients which have been monitored with respect to their latter control visit in one of the center and which reacted to the therapy. We are also studying the possibility of monitoring patients with covid-19 which have symptoms with respect to those that does not present symptoms. The difference can be studied with respect to the possibility of considering different drug treatments used by patients across latter years as well as environmental phenomena.

V. Conclusions

Clinical and geographical integration allows to perform epidemiological and statistical studies about infection disease. This is useful to identify possible environmental risk factors associating environmental data with possible pathological cases. The proposed system improves the analysis of clinical

Fig. 9. COVID-19 manifestation in Italy between 20 and 21 February 2020.

data and supports the decision-making process by the medical staff.

ACKNOWLEDGE

The paper has been been partially supported by PON VQA project. Authors are grateful to all members of the MASTER italian project.

REFERENCES

[1] I. Abubakar, H. Stagg, T. Cohen and L. Rodrigues, "Infectious Disease Epidemiology", Oxford University Press, 2016.
[2] G. Canino, P. H. Guzzi, G. Tradigo, A. Zhang and P. Veltri, "On the analysis of diseases and their related geographical data", IEEE journal of biomedical and health informatics, vol. 21(1), pp. 228-237, 2015.
[3] G. Tradigo, P. Vizza, P. Veltri, P. Lambardi, F. M. Caligiuri, G. Caligiuri and P. H. Guzzi, "SISTABENE: an information system for the traceability of agricultural food production", 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 2304-2309, 2019.
[4] R. S. Kirby, E. Delmelle and J. M. Eberth, "Advances in spatial epidemiology and geographic information systems", Annals of epidemiology, vol. 27(1), pp. 1-9, 2017.
[5] P. Vizza, G. Tradigo, P. Veltri, P. Lambardi, C. Garofalo, F. M. Caligiuri, G. Caligiuri and P. H. Guzzi, "Tracking agricultural products for wellness care", 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 2064-2067, 2018.
[6] P. H. Guzzi, D. Mercatelli, C. Ceravolo and F. M. Giorgi, "Master Regulator Analysis of the SARS-CoV-2/Human Interactome", Journal of clinical medicine, vol. 4, pp. 982, 2020.
[7] WHO Coronavirus Disease (COVID-19) Dashboard. Available online at https://covid19.who.int/
[8] G. Tradigo, P. Vizza, P. Veltri and P. H. Guzzi, "An Information System to Track data and processes for food quality and bacterial pathologies prevention", SEBD, 2019.
[9] G. Tradigo, C. Pagliaro, G. Canino, P. H. Guzzi and P. Veltri, "GIS for the analysis and monitoring of environmental issues related to life quality", SEBD, pp. 320-326, 2015.
[10] G. Tradigo, C. Pagliaro, G. Canino, F. Casalinuovo, C. Graziani and P. Veltri, "A model for the Geographical Analysis and monitoring of agricultural areas example and tests in south Italian regions", HealthGIS, 2014.
[11] G. Tradigo, M. Cannataro, P. H. Guzzi, F. Casalinuovo, P. Veltri and C. Graziani, "Health risk assessment of zoonotic infections agents through plant products in areas with high livestock pressure", In Proceedings of the First ACM SIGSPATIAL International Workshop on Use of GIS in Public Health, pp. 72-74, 2012.
[12] A. Qayum, R. Arya, P. Kumar and A. M. Lynn, "Socio-economic, epidemiological and geographic features based on GIS-integrated mapping to identify malarial hotspots", Malaria Journal, pp. 14-192, 2015.
[13] F. Destefano, E. A. Eaker, S. K. Broste, D. L. Nordstrom, P. L. Peissig, R. A. Vierkant, K. A. Konitzer, R. L. Gruber and P. M. Layde, "Epidemiologic research in an integrated regional medical care system: The Marshfield Epidemiologic Study Area", Journal of Clinical Epidemiology, vol. 49(6), 1996.
[14] C. T. Guo, Q. B. Lu, S. J. Ding et al., "Epidemiological and clinical characteristics of severe fever with thrombocytopenia syndrome (SFTS) in China: an integrated data analysis", Epidemiology and Infection, vol. 144(6), pp. 1345-1354, 2016.
[15] C. B. Ocampo, N. J. Mina, M. I. Echavarria, M. Acuna, A. Caballero, A. Navarro, A. Aguirre, I. S. Criollo, F. Forero, O. Azuero and N. D. Alexander, "VECTOS: An Integrated System for Monitoring Risk Factors Associated With Urban Arbovirus Transmission", Global Health: Science and Practice, vol. 7(1), pp. 128-137, 2019.
[16] F. Riccardo, M. Ajelli, X. Andrianou, A. Bella et al., "Epidemiological characteristics of COVID-19 cases in Italy and estimates of the reproductive numbers one month into the epidemic", medRxiv, 2020.
[17] E. Puca, J. Arapovic, C. Popescu, I. Christova, et al., "Short epidemiological overview of the current situation on COVID-19 pandemic in Southeast European (SEE) countries", The Journal of Infection in Developing Countries, vol. 14(05), pp. 433-437, 2020.
[18] L. I. Zambrano, M. A. Sierra-Santos, H. M. Mejia-Santos, et al., "Spatial Distribution of COVID-19 in Honduras at the Early Phase of the Pandemic Using Geographic Information Systems (GIS)", 2020.
[19] Eugenio Vocaturo, Ester Zumpano, Pierangelo Veltri"Image preprocessing in computer vision systems for melanoma detection", IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018.