POLITECNICO DI MILANO

PhD in BIOINGEGNERIA / BIOENGINEERING - 39th cycle

PNRR 118 PNRR Research Field: DEVELOPMENT OF ARTIFICIAL INTELLIGENCE, STATISTICS, DATA SCIENCE AND SOFTWARE ENGINEERING TOOLS FOR THE MANAGEMENT OF LARGE BIOMEDICAL DATA SETS: AN INTERDISCIPLINARY APPROACH

Monthly net income of PhDscholarship (max 36 months)
€ 1400.0
In case of a change of the welfare rates during the three-year period, the amount could be modified.

Context of the research activity	
Motivation and objectives of the research in this field	The availability of large amounts of data is one of the key conditions for studying the evolution of the state of well-being and the transition to pathology in different classes of subjects. In the field of biomedical research, in general, the purpose of the analysis of large amounts of data is the description of the health condition of the subjects, highlighting, with appropriate advanced tools, the onset of transition signs to pathological states.

The main challenge in this context is the inherent interdisciplinarity of this research, which requires an in-depth knowledge of the collected data, the phenomena they are associated to and the effects these latter have on pathological states together with the ability to apply techniques from multiple other fields. These include (i) statistical modelling, AI machine and deep learning techniques (ii) data management for the integration of heterogeneous sources and for the control of the (good) quality of data (iii) software engineering to allow for the design of reusable, maintainable and scalable software pipelines, devoted to properly orchestrate the execution of multiple data analysis tasks on potentially heterogeneous resources.

This interdisciplinary and interdepartmental PhD project addresses this challenge by developing a framework to |

POLITECNICO DI MILANO

	addresses this challenge by developing a framework to support the analysis of biomedical signals with methods from statistical modelling, machine learning and data science techniques. The framework will include development of libraries of tools to perform specific data management and analysis tasks. It will be built to exploit the computational and storage capacity of heterogeneous resources, from cloud to HPC clusters and, where needed, also quantum computers. Composing different tools to achieve complex analyses and checking the correctness of the biomedical findings will be key aspects of the project as well. The proposers are already involved in the activity of the DataCloud Interdepartmental Laboratory. The advanced computing facilities available in the laboratory will be made available to the research, together with the competences of all researchers involved in the Lab.
Methods and techniques that will be developed and used to carry out the research	Through this PhD Thesis we plan to find a structured and interdisciplinary way of working together that can provide an innovative contribution to Health Big data. As mentioned above, the PhD student will develop knowledge and skills in the treatment and quality assessment of large biological databases. As a contribution to the research he/she can utilize as a starting point, on a pre-existing database of fetal heart rate data with 20,000 recordings taken from measurements on women at different periods of pregnancy. The student will analyse the data to identify events that can act as predictors of fetal issues. This important exercise will allow the PhD student to understand the nature of data to be treated and relevant data integration and data analysis problems to be addressed. The student will extend the database, integrating other data sources and managing periodic updates. Moreover, he/she will develop a catalogue of analytics based on statistical methods, machine learning, deep learning and feature-based modelling. Altogether, this will constitute an information extraction system that will be evaluated and optimized in terms of reproducibility, classification accuracy and prediction utility.

	The second year of activity will be dedicated to the proposal of feature-based models for clustering of subjects. The estimated subgroups will be subsequently identified as clusters with different level of the pathology we focus on (including healthy subjects) and hopefully explained in terms of key features. This approach should also address a dynamical framework for the clustering problem, since subjects change their health state (healthy/ill) through time. Moreover, during the second and mostly the third year of the project, the generalizability of the proposed data integration approach and catalogue will be studied and evaluated on other large biological databases. The thesis will study how to adapt clustering and classification methods, which can follow the dynamic evolution of the database and the signals it contains. Moreover, an approach to verify and guarantee the quality of Bio data analyses will be studied. Pipelines of management, analysis, classification and quality of data will be proposed.
Educational objectives	The PHD course is organized in order to achieve important educational objectives. The PhD student is encouraged to take courses proposed by both the PhD School and the PhD in Bioengineering in order to communicate and extend their multidisciplinary knowledge of bioengineering. The supervisor and his research team continuously support the development of research. Seminars and proposed courses encourage an interdisciplinary approach. The laboratory activity completes the research path. The final objective is for the candidate to develop a strong theoretical knowledge of the subjects covered by his doctorate.
Job opportunities	Job opportunities include research in academic and private institutions in Italy and abroad and in industry. Research bodies such as IRCCS also receive PhDs. Spin-offs and start-ups are equally interesting and often even generated by the results of doctoral research. Employment in this field offers various interesting

	opportunities.
Composition of the research group	2 Full Professors 1 Associated Professors 4 Assistant Professors 2 PhD Students
Name of the research directors	PROF. MARIA GABRIELLA SIGNORINI

Contacts
Maria Gabriella Signorini mariagabriella.signorini@polimi.it +393355938411 https://www.deib.polimi.it/ita/personale/dettagli/367006

Additional support - Financial aid per PhD student per year (gross amount)	
Housing - Foreign Students	--
Housing - Out-of-town residents (more than 80Km out of Milano)	--

Scholarship Increase for a period abroad	
Amount monthly	700.0 €
By number of months	6

National Operational Program for Research and Innovation	
Company where the candidate will attend the stage (name and brief description)	PHILIPS Philips Center, Amstelplein 2, 1096 BC Amsterdam, The Netherlands (Address of principal executive offices)
By number of months at the company	6
Institution or company where the candidate will spend the period abroad (name and brief description)	Philips Research Eindhoven High Tech Campus 34 5656 AE Eindhoven The Netherlands
By number of months abroad	6

Additional information: educational activity, teaching assistantship, computer availability, desk availability, any other information

Educational activity: The student will be encouraged to attend to courses at POLIMI or abroad in International Schools.

Teaching assistantship: There are various forms of financial aid for activities of support to theteaching practice. The PhD student is encouraged to take part in these activities, within thelimits allowed by the regulations.

Computer and desk availability: the student will be allowed to access facilities of the DEIB.