

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/331418148>

Multiscale Computing in Systems Medicine: a Brief Reflection

Conference Paper · December 2018

DOI: 10.1109/BIBM.2018.8621361

CITATIONS

3

READS

175

10 authors, including:



Huiru Zheng
Ulster University

334 PUBLICATIONS 3,431 CITATIONS

SEE PROFILE



Jyotsna Wassan
University of Delhi

20 PUBLICATIONS 93 CITATIONS

SEE PROFILE



Mihnea Alexandru Moiescu
Polytechnic University of Bucharest

74 PUBLICATIONS 471 CITATIONS

SEE PROFILE



Lacramioara Stoicu-Tivadar
Polytechnic University of Timisoara

106 PUBLICATIONS 344 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



CoDBIS (Computing, Database and Information Systems) Group: Optimization of DWs [View project](#)



development of models [View project](#)

Multiscale Computing in Systems Medicine: a Brief Reflection

Huiru Zheng
School of Computing
Ulster University
 Northern Ireland, United Kingdom
 h.zheng@ulster.ac.uk

Lacramioara Stoicu-Tivadar
Automatics and Applied Informatics
University Politehnica Timisoara
 Timisoara, Romania
 lacramioara.stoicu-tivadar@aut.upt.ro

Ioan Stefan Sacala
Automatics and Industrial Informatics
University Politehnica of Bucharest
 Bucharest, Romania
 ioan.sacala@acse.pub.ro

Jyotsna Talreja Wassen
School of Computing
Ulster University
 Northern Ireland, United Kingdom
 Wassen-JT@ulster.ac.uk

João Miranda
 Instituto Politécnico de Portalegre
 CERENA, Instituto Superior Técnico
 Portugal
 joaoluismiranda@tecnico.ulisboa.pt

Almir Badnjevic
F. of Engineering and Natural Sciences
International Burch University Sarajevo
 Bosnia and Herzegovina
 almir.badnjevic@ibu.edu.ba

Mihnea Alexandru Moisescu
Automatics and Industrial Informatics
University Politehnica of Bucharest
 Bucharest, Romania
 mihnea.moisescu@upb.ro

Mihaela Crisan-Vida
Automatics and Applied Informatics
University Politehnica Timisoara
 Timisoara, Romania
 lacramioara.stoicu-tivadar@aut.upt.ro

Ivan Chorbev
Faculty for Computer Science and
Engineering, University Ss Cyril and
Methodius in Skopje
 ivan.chorbev@finki.ukim.mk

Boro Jakimovski
Faculty for Computer Science and
Engineering, University Ss Cyril and
Methodius in Skopje
 boro.jakimovski@finki.ukim.mk

Abstract— Today’s modelling approaches in Systems Medicine are increasingly multiscale, containing two or more submodels, where each operates on different temporal and/or spatial scales. In addition, as these models become increasingly sophisticated, they tend to be run as multiscale computing applications using computational infrastructures such as clusters, supercomputers, grids or clouds. Constructing, validating and deploying such applications is far from trivial, and communities in different scientific disciplines have chosen very diverse approaches to address these challenges. Within this paper we reflect on the use of Multiscale Computing within the context of Systems Medicine. Multiscale Computing is widely applied within this area, and instead of summarizing the field as a whole we will highlight a set of challenges that we believe are of key relevance to the Systems Medicine community.

Keywords—*Multiscale Computing; Systems Medicine; Ambient Assisted Living (AAL); Cloud Computing; Data Science; High Performance Computing (HPC); Internet of Things (IoT).*

I. INTRODUCTION

Nowadays, very large volumes of both biological and clinical data are being generated, sometimes in complex datasets, other times in unstructured schemes. Most of the currently publicly available datasets are fragmentary, frequently underpowered. Hence, as not that much structured and annotated data is yet available, although more and more datasets of sufficient quality are being assembled, more effort should be made to coordinate the primary data generating entities (MD, physiologists, pharmacologists, microbiologists, ...), to produce the data in machine readable format, with all accompanying metadata correctly structured and attached onto the databases.

In the meantime, the resources for multiscale computing (storage, processors, bandwidth, etc.), the related infrastructures, and IT tools (communication protocols, cloud, open software, etc.) are becoming more and more available. Namely, within the Horizon Europe framework, research

fields are being revisited and innovation ecosystems are being outlined. For example, a number of national grids connected into EU wide consortia can offer computational capacity, such as ELIXIR [1] or others.

Therefore, increasingly, the current challenges in the healthcare research and services, combined with the recent enhancements in Multiscale Computing, are calling for a systematic and quantitative analysis, as well as innovative approaches for decision aiding/making. For that, Operational Research (OR) tools and methodologies are very useful too, and the related contributions are becoming important in several societal challenges.

Among others, the most recent initiatives in Multiscale Computing analyzed within the COST Action “OpenMultiMed” include:

- 3D-1D coupled simulations of blood flow in the Middle Cerebral Artery [2].
- Ambient Assisted Living (AAL) systems, which are based on the interoperability and the proper integration of multiple medical devices;
- Cloud based services and client applications for medical ecosystems;
- Data Science specific issues, namely, integrative analysis, dedicated algorithms, and generation of multiscale datasets;
- Internet of Things (IoT) and related ecosystems as a key data provider for Multiscale Modelling, for Multiscale Computing, and addressing educational applications too.
- Tools and methodologies of Operational Research applied to healthcare services (ORAHs).

This paper considers the following parts: in Section 2, we intend to highlight key multiscale computing challenges in the

context of healthcare and assisted living settings; Section 3 tackles Multiscale computing in the context of the emerging IoT domain. Finally Section 4 gives a conclusion.

II. MULTISCALE COMPUTING AND DATA SCIENCE

Multiscale modelling is useful for understanding biological systems which are inherently divided into different scales or functional units which interact with each other at different levels of hierarchy (Fig. 1). Multiscale systems assume that the growing amounts of highly diverse (multiscale) data relevant to human health and disease are the key to address current and future medical challenges. Transforming these data into effective and economical medical solutions requires appropriate means for multiscale data modelling, integration and analysis. Multiscale data (i.e. scales) are potentially used simultaneously to describe a system or a phenomenon. For example, biological systems are multi-scale: -a cellular network represents molecular interaction, cell population dynamics rely on intercellular protein signaling, which relies on fluctuations of intracellular protein concentrations, distributed between tissues via vascular system, receiving and exchanging signals also with human intestinal microbiome, and so on [3].

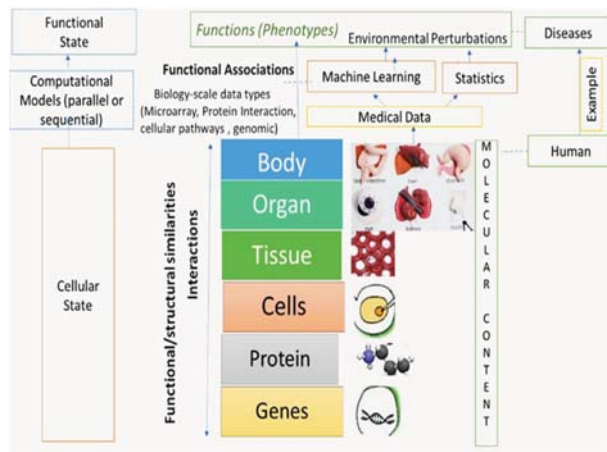


Fig. 1. Biological scales in relation to the modelling and experimental approaches of multiscale systems biology

With the rapid development of data acquisition techniques, scientists have generated a large amount of data at different biological scales in spatial and temporal dimensions, that largely do not overlap in most published datasets. Multi-scale systems biology is concerned with experiments and hypotheses that involve different scales of biological organization from intracellular molecular interactions to cellular behaviour and the behaviour of cell populations, for which different methodologies can be used [4]. Fig. 1 illustrates biological scales in relation to experimental approaches of systems biology. Computational modelling over different scales can reveal which functional units are - i) correlated with environmental functions (e.g. diseases in humans); ii) highly expressed components in the biological sample at a specific time and place; iii) useful for understanding the biological configuration and its impact on organisms. The computational tools have also been developed for integrating biological scale data to identify the network of connections between the biological components and to develop system models to simulate and predict the network

behaviour in response to human functions (e.g. diseases and medicine).

The communication between cellular and functional components is crucial and hence is useful to study their interaction supporting human systems biology [3,4].

A. Integrative analysis

Integrative analyses aids in synthesizing and transforming multi-scale data into valuable biological insights by using an ensemble of computational techniques or by using combinations of experimental data. The integrative approach allows combining different algorithms or biological datasets to increase the performance of computational analyses ranging from machine learning, statistical analysis to optimization and network-based approaches. This supports the comprehensive analysis of data towards a better understanding of biological systems [3].

1) Data Integration

The abundance and heterogeneity of biological data have made data integration approaches increasingly popular for systems biology. Integration could be useful in case of normalization, formatting, and dimensionality of biological datasets, the presence of noisy points in medical data, selection of informative biological components (e.g. genes) from datasets, and scalability with the number and size of datasets [5]. Gligorijevic' et al. [5] identified using methods of data integration such as MAGIC (for gene prediction), MRLP (for disease association prediction), PRINCE (for disease-gene prioritization), PreDR (for drug repurposing) etc., enabling analysis of diverse biological datasets. Krempel et al. [6] recently created Cancer Systems Biology Database (CancerSysDB), a resource for determining cancer-related medical information across multiple source studies as a useful Use Case. Also, integrating prior biological knowledge such as information from Gene Ontology, phylogeny into computational models to predict the phenotypes forms an interesting basis for biological medicinal studies (e.g. IMG tool) [7]. However, the structure and distributions of biological data obtained from different tasks such as sequencing-based or a microarray-based or medical imaging, makes it challenging to combine or integrate. The development of computational methods for data integration of large-scale biological data is an active area of research.

2) Algorithm Integration

Combining algorithms in a sequential and/or parallel manner potentially could provide performance improvements and new biological insights (Fig.2.). A principled framework combining different ML models into a unified model for studying a biological ecosystem dealing with the variety of tasks such as drug-disease prediction. The predominant ML methods useful in integration are :- i) kernel-based methods (KB), probabilistic models (PB), tree-based methods (TB) and regression-based models ((RB) [8]. KB methods such as support vector machines, principal component analysis, etc. are useful for data mining and analysis and work by mapping data to the desired feature space, e.g. for learning tasks, such as clustering, classification, regression, correlation, feature selection, etc. [8]. Probabilistic models such as neural networks gain concepts from probability to model the relationships between various feature points present in the data [8]. TB methods such as random forest depend highly on hierarchical way of partitioning the space in a recursive manner [8]. These work on decision rules for predicting the

outcome of interest. RB methods depend highly on finding the causal effect relationship between the various feature variables. There is no one standard from the above-discussed models which performs better in every biological use case; hence combining different models (i.e. ensemble of algorithms) may perform better than the traditional models. Wassan et al. combined TB methods with RB methods to attain better performance over human microbiome to associate them with diseases [9]. Gorunescu et al. [10], constructed an ensemble of TB algorithms, collaborating with each other to make an “intelligent” efficient global decision framework for breast cancer detection and various stages for liver fibrosis in chronic hepatitis C [10]. Cunningham et al. [11] reported that an ensemble of neural networks produces better predictions for medical decision support systems. Badnjevic et al [12] have developed automated system based on neural network and fuzzy logic to predict COPD and asthma disease and proved that these systems are very useful when available to rural areas [13]. Overall, ensemble learning could prove useful for meta-strategy for the medical field and serves an interesting domain

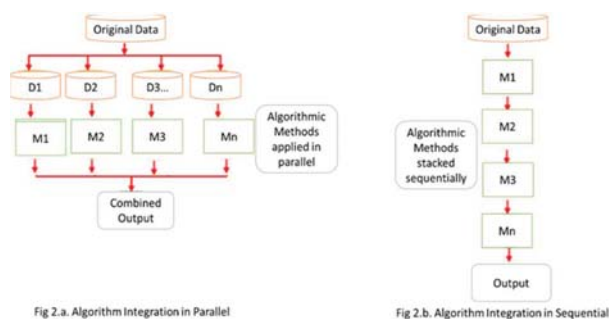


Fig. 2. Illustration of the Algorithm Integration

B. Multiscale algorithms

1) Multi-grid methods

Multi-grid methods have been adapted from mathematical models related to solving elliptic boundary-value problems using a hierarchy of discretizations. In general, developing a multiscale grid solution involves selecting an appropriate local level procedure/process and thereafter selecting methods for transferring information across, by developing appropriate equations. Such an approach is very useful in problems exhibiting multiple scales of data. Particularly, in medical applications the partial differential equations (PDEs) based analysis with segmentation and registration of medical images has proved successful [14]. Multi-grid approaches supports almost all stages of patient care. For example, Chen et al. [15], simulated tumor growth in 2D and 3D grids that demonstrated the capabilities of the approach in analysing the progression of tumors with complex morphologies [15]. Fogolori et al. [16] applied Poisson–Boltzmann equation [16] for grid-based modelling over the functional units of structural biology (e.g. proteins, D.N.A etc.). Poisson–Boltzmann equation was integrated into a multi grid-based technique for calculating finite differences to attain accurate and faster solutions for studying the protein structures [17]. The discrete approximations of geometrical quantities with multigrid optimization algorithm are useful for identification of cellular structures in biology.

2) Multi-agent systems

Multi-agent systems serve as an integrative collection of autonomous intelligent units that collaborate in dynamic ways to solve the problems of medical domains such as monitoring medical protocols for tumor growth, organ transplant, medicine.

3) Bayesian Learning

A Bayesian model determines joint multivariate probability distribution to determine conditional independence between variables. Such models are useful for stochastic processes and describing interactions between various functional units (such as genes, proteins etc.) and their role in determining human functions. The analysis potentially dictates restricted use of medicines or wound healing [18]. Multi-agent systems aid in solving distributed problems by coordinating their actions and collaborating to derive a solution for the medical problem. A medical problem is decomposable into subproblems to be handled with various agents such as communication manager (to monitor patients), database manager (to maintain patient health records) and data analyser to create patient models (to make predictions on patient’s data). For example, the PalliaSys project, based on multi-agent systems was aimed to improve the care for palliative patients [18]. Mfumu et al. [19], recently presented the use of multi-agents to monitor epidemics and chronic diseases in DR Congo [19]. The main aim of such systems is to improve collaboration and shorten the response time to manage health data between actors in society (patients, doctors etc.).

4) Semantic data analytics

The data vocabularies and biomedical ontologies provide useful domain knowledge for associating patterns in the biological data with phenotypes. As highlighted by He et al.[20], human health is majorly powered by following data analytics strategies:- i) natural language processing, (2) data sharing, integration, mining and (3) ontology engineering. The semantic structures facilitate the data integration tasks and develop integrative framework addressing for computing with the data such as linking human’s biological structural data to contextual and environmental factors of diseases [21]. For example, recently Zhange et al., created a semantic structure highlighting Ontology for Cancer Research Variables (OCRV) adapting resources from the National Cancer Institute (NCI) Thesaurus [21].

5) Network-based approach

The recent advances in techniques of systems biology supports the linkage of human diseases result with perturbations of cellular components especially gene/protein/molecular configurations [22]. Network-based methods are useful for prioritizing molecular basis of human diseases such as finding neighboring biomolecules that are responsible for similar diseases and hence to infer how likely it is that a structure (e.g. gene) is associated with functions (e.g. diseases). Molecular networks, including Protein-Protein Interactions, metabolic, regulatory, genetic and co-expression networks are useful for studying the functional interactions. It’s interesting to study whether similar biomolecules are responsible for phenotypically overlapping diseases [23]. Le et al.[24] demonstrated a heterogeneous network by integrating the PPI network and phenotypic disease–gene relationships[24]. Hau et al. used ranking phenomena based on network-based modelling for the prediction of novel diseases associated with microRNAs. Recently, Wang et al. [25] proposed the use of multiplex networks in metagenomics

which is easily extensible to determine microbial genes in humans to be associated with diseases and hence in systems gene regulatory networks, protein–DNA interactions, protein–protein interactions in translational processes or enzymatic reactions [26]. Nicholson et al. [27], suggested a hypothetical (Pachinko) model to study and visualize the interactions between genes and drug reactions in humans. The Bayesian Networks as proposed by Troyanskaya et al. [28], used expert knowledge for combining heterogeneous data with biological domain knowledge for predicting the gene functions. Freidman et al. [26] highlighted the study of gene expressions from microarray data using the Bayesian networks. The applicability of Bayesian model to study the omics data sets and its relation to toxicology and nutrition was well studied by Joyce et al. [29]. The experimental time series measurements serve as an important phenomena for modelling dynamic biological network systems. The feedback mechanism in such systems are replicable with Dynamic Bayesian Networks [30].

C. Generating the multiscale datasets

The vast majority of bioinformatic software packages that are currently used (and benchmarked) in the field of metagenomics, metatranscriptomics, metabolomics and proteomics, including draft genome assembly from microbiological systems are in general not written for distributed memory systems such as Open MPI. As such, large scale bioinformatic 'omics projects of complex environmental systems require access to single high performance computing nodes each with a combination of a large contiguous memory bank and a high cpu core number, in addition to large storage space. This brings us to two faceted problem that researcher in the field of “big data” face: the need for complete refurbishment of existing utilities toward OpenMPI on one side and availability of single high performance computing nodes each with a combination of a large contiguous memory bank and a high cpu core number that support the existing software, on the other. These effective scaling of complex analysis of primary data is in fact needed in order to generate the necessary data matrices of e.g. genes, transcripts, proteins, metabolites, compatible with the developed downstream approaches of multidimensional and multiscale data analyses, modelling in multiscale computing. Once large datasets from 'omics technologies are agglomerated from a number of patients or nationwide, on-site analyses using large genomic, proteomic, enzyme, pathway and metabolomic databases (that are themselves under constant updating) are going to be conducted on primary data, hence facing this double computational/software problem. The fast development of the field of systems medicine and biology (decreasing sequencing costs, increasing sequencing depth as well of the proteomics, metabolomics, lipidomics, including large numbers of samples, replicates, technical replicates, tools efficiently scaling with the number of samples, the use of docker technology, novel computational solutions, the use of systems biology or medicine scaling and modelling approaches) has introduced highly interesting opportunities not only to produce more new data, but to largely reanalyze the existing primary datasets from scratch, using meta-analyses in order to provide more firm ground over the particular field of interest. So far this approach has been practically impossible. It is to be expected that once the bottle necks of structuring primary data and computing reproducibility are solved many other labs that currently work on other aspects contributing to systems medicine are going to pick up the novel tools as well,

especially as wider use is also projected to bring forward standardization of data analysis approaches, making the outcomes more comparable between groups on the European and global scale, hence forging the more coordinated way towards novel discoveries of industrial and medical relevance.

III. MULTISCALE COMPUTING AND THE IoT

Internet of Things (IoT) is one of the most important research priorities related to the evolution of internet towards integrating real objects. A comprehensive description is provided in [31]: IoT “is a concept and a paradigm that considers pervasive presence in the environment of a variety of things/objects that are able to interact with each other and cooperate with other things/objects to create new applications/services and reach common goals”. Thus IoT is based on technologies and architectures, for large-scale, loosely coupled systems, with the capability to identify and address virtual representations of physical devices. IoT generic components include: IoT Standards and Ecosystems, Event Stream Processing, IoT Device Management, IoT Platforms, IoT Analytics, and IoT Security. [32]. The IoT Reference Model proposes the following sub-models: IoT Domain Model, IoT Information Model, IoT Functional Model, IoT Communication Model, and IoT Security Model [31]. In practice the following generic components are identified: Smart devices, Network, Data processing, Data storage, Data aggregation, data analytics, and process integration. A set of communication models for IoT is represented by: device to device communication, device to IoT platform communication, device to gateway and data aggregation. The relation between IoT and Multiscale Computing (MSC) and Multiscale Modeling and Simulation (MMS) has not been extensively documented in research literature. Some aspects have been partially identified. Two perspectives can be identified in order to document a relation: IoT as data provider for Multiscale Modeling and Multiscale Modeling as a way to model complex processes.

A. IoT as data provider for Multiscale Modeling

IoT can become a data provider for Multiscale Modelling by facilitating access by routing and recording data acquired from sensors attached to smart objects. In this context, the sensory information can be used to create data-driven multiscale simulations. Another aspect is related to the fact that large scale heterogeneous systems connected through IoT can provide data for computation and analysis at different resolution levels. Relevant enabling technologies are related to Big Data, Semantic representation and Multiscale context awareness. Big Data as an evolution in data analytics addressing challenges related to the unstructured data, data volume and processing time. Semantic representation addresses data exchange formats and is closely related ontologies and machine-interpretable representations of domain knowledge. Context awareness systems address complex environments in terms of location, identity and components and relations. Context sensing capability is provided by Wireless Sensor Networks technologies that allow for the collection of data from different types of sensors distributed in the physical environment. Cross-domain sensing and coupling of sensors is an important aspect that may affect the performance and reliability of multiscale computing applications. Multiscale context awareness research project INCOME [33] provides solutions for multiscale context management in IoT. IoT sensors can provide

data measured at different scales associated to the sub-systems of a complex system as well as aid in scale separation mapping. IoT connects heterogenous devices and allows for a centralised data storage and data aggregation. In relation to Systems Medicine the following aspects have been identified. Specialized medical devices as well as smart sensors can aid in clinical investigations. Data stored and aggregated through IoT platforms can be used in computational multiscale analysis and modelling. Disease progression and treatment response can be monitored with wearable or implanted smart devices. Disease prevention can be modelled and monitored.

IoT devices, linked together with the aid of an IoT platform can aid in scale bridge problems as well as in model validation and verification. As multiscale modelling extends from micro and macro scale towards the analysis of the effects on the human body and groups of people, IoT becomes an important model and means for data gathering in a non-invasive ubiquitous manner.

B. Multiscale Modeling of IoT Ecosystem

Multiscale Modeling as a way to model complex processes addresses the ecosystem of things created within IoT. The complex processes include Machine to Machine and Human to Machine Interaction. Relevant enabling technologies are related to Heterogenous objects, Heterogenous distributed systems (P2P, Wireless Sensor Networks, Cloud Computing), Complex Systems of Systems. IoT as a complex systems is not a simple set of subsystems and involves data and energy transformation, interaction, interoperability, feed-back and feed-forward structures, self-organization and self-management. The System of Systems concept is described in terms of "Meier's criteria": operational and managerial independence, distribution, emergent behavior as a result of component behavior and evolutionary development [34].

C. Case study: Multiscale computing and modelling in IoT applied in education in a technical faculty

This sub-section describes the planning, development and results of laboratory and project activities during the first to third year for the students of the Faculty of Automation and Computers in a technical University and continuity to Master and PhD programs activities as steps to prepare good professionals using IoT and integrating it in multiscale modelling with applications in digital healthcare.

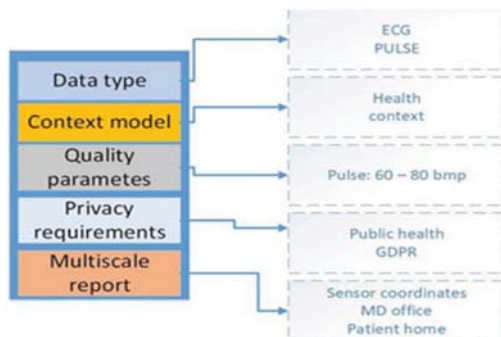


Fig. 3. Multilevel context for multiscale and IoT practicing

The workstyle and content are preparing technical students to develop complex healthcare information systems and manipulate concepts as multiscale modelling and computing in relation with new IoT environment. The activities are

performed during 1 semester with 14 weeks, 2 hours/week, teaching the students to work with sensors, and modelling a complex system. The educational prototyping systems are using Arduino kits, and specific components and sensors, collecting the information in the Cloud anytime and anywhere (Fig. 2.).

Gaining the basic abilities, the future engineers or informaticians have the possibility to follow the Master in Healthcare Informatics and the PhD program with thesis on topics related to digital healthcare. Fig. 3 presents the complexity of the process. The sensors and kits that are a basis in IoT for healthcare applications are of great help as real time systems being very quick in performing their operations and are easy to configurate in complex systems connected on Cloud or mobile applications presenting interoperability and continuity potential for medical processes.

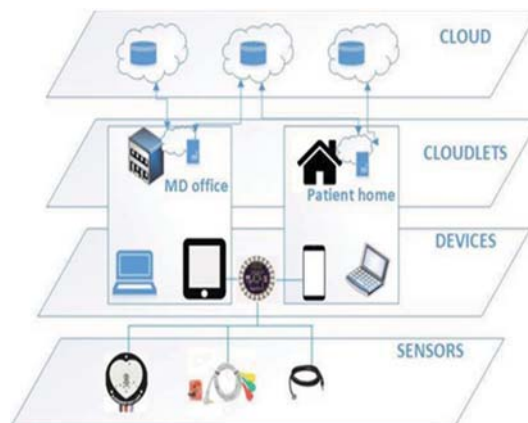


Fig. 4. Complex process of multiscale and IoT approach in education

During the lab activities students learn gradually how to work with Arduino UNO and sensors in different contexts. The sensors retrieve data regarding patient health, such as pulse, ECG, pulse oximetry, and environmental ones influencing health status, as temperature and humidity. These signals are sent wireless to the doctor's devices for monitoring in remote locations. The student's training to develop skills related to the Internet of Things domain starts with the 1st year of study, when they learn C# language and receive small scale projects. The components of the projects are Arduino UNO modules, biomedical sensors (ECG, pulse) or environmental sensors (temperature, humidity). They continue in the 2nd year with the Object-Oriented programming lecture where they learn to work using embedded systems with classes, objects, inheritance, and polymorphism. In the 3rd year of study they learn to connect these embedded systems with C# applications and save the data received from the sensors into databases in the cloud. During the Master for Informatic Systems Applied in Healthcare program they will develop a more complex system, with a Hardware module (Arduino and components), C# application (ASP.NET web application), databases, cloud (e.g. Azure, Firebase) or using WiFi to create IoT (Internet of Things) applications. Students with abilities for research in digital healthcare continue with a PhD program where currently they focus on Internet of Things, Virtual Reality and Deep Learning applied in healthcare. The activities with students are managed almost entirely (documentation and test for exams) using an on-line system. University Politecnica Timisoara Campus Virtual (CVUPT) is an academic online and communication platform for

academic staff and students. This platform is used extensively, students accessing the documentation associated with a discipline, receive different news from the professor, upload different projects or taking tests for exams.

During the master program project, the master students continue attaching a WIFI shield and save the data in different cloud platforms (e.g. Microsoft Azure IoT hub). After collecting the data, the information is accessed by a prototype of an information system for a general practitioner, a specialist in a hospital, a pharmacy, a lab and the insurance house. There is also connection with mobile phone applications for the physician and the patient. The applications context is changing yearly, relating healthcare monitoring, prevention, fitness and healthcare management.

During current PhD activities students with research topics related to digital healthcare work using different techniques for deep learning applied to medical data in ob-gyn department for prevention, and virtual reality and gestures for medical education.

For theoretical support in multiscale modelling and computing related to IoT to be developed and disseminate, we consider education as an important player and implementations and practice a fundamental activity for the future of the domain.

IV. DISCUSSION AND CONCLUSIONS

There is tremendous output of primary data at various levels, scales, tissues, timeframes, geolocations, subpopulations, in the form of classical group comparisons or data collected longitudinally. However, these datasets are not made available in computer readable format to large extent, preventing the use of available advanced computational tools and coupling these tools to additional nationwide metadata.

Due to the limitations presented in this manuscript, many of the computational approaches developed so far have not been tested on really comprehensive datasets, despite the potential that they contain.

Therefore some level of benchmarking, also using completely randomly assembled data, would also be feasible, in order to provide the extent of false positive predictions in comparison to true primary data. Also, as starting from the primary data is complex process of deconvolution of sequencer files, NMR files, GC/MS files, databases, metadata into standard data matrices, a more standardized approach would be more productive.

ACKNOWLEDGMENT

Authors would like to thank the support of COST Action OpenMultiMed (CA 15120). H Zheng and JW Wassan received support from the EU MetaPlat project (H2020-MSCA-RISE-2015). J.L.Miranda. thanks Instituto Politécnico de Portalegre, and Centro de Recursos Naturais e Ambiente at Instituto Superior Técnico (CERENA/IST) within the support of FCT-Fundação para a Ciência e a Tecnologia through the project UID/ECI/04028/2013.

REFERENCES

- [1] <https://www.elixir-europe.org/> (7.11.2018)
- [2] Derek Groen, Joris Borgdorff, Carles Bona-Casas, James Hetherington, Rupert W Nash, Stefan J Zasada, Ilya Saverchenko, Mariusz Mamonski, Krzysztof Kurowski, Miguel O Bernabeu, et al. Flexible composition and execution of high performance, high fidelity multiscale biomedical simulations. *Interface Focus*, 3(2):20120087, 2013.
- [3] M. Meier-Schellersheim, I. Fraser and F. Klauschen, "Multiscale modeling for biologists", *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, vol. 1, no. 1, pp. 4-14, 2009.
- [4] J. Walpole, J. Papin and S. Peirce, "Multiscale Computational Models of Complex Biological Systems", *Annual Review of Biomedical Engineering*, vol. 15, no. 1, pp. 137-154, 2013.
- [5] V. Gligorijević and N. Pržulj, "Methods for biological data integration: perspectives and challenges", 2018.
- [6] R. Krempel, P. Kulkarni, A. Yim, U. Lang, B. Habermann and P. Frommolt, "Integrative analysis and machine learning on cancer genomics data using the Cancer Systems Biology Database (CancerSysDB)", *BMC Bioinformatics*, vol. 19, no. 1, 2018.
- [7] V. Markowitz, "The integrated microbial genomes (IMG) system", *Nucleic Acids Research*, vol. 34, no. 90001, pp. D344-D348, 2006.
- [8] J.Wassan, H.Wang and H.Zheng, "Machine Learning in Bioinformatics", 2018.
- [9] J. Wassan, H. Wang, F. Browne and H. Zheng, "A Comprehensive Study on Predicting Functional Role of Metagenomes Using Machine Learning Methods", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pp. 1-1, 2018.
- [10] F. Gorunescu and S. Belciug, "Evolutionary strategy to develop learning-based decision systems. Application to breast cancer and liver fibrosis stadialization", *Journal of Biomedical Informatics*, vol. 49, pp. 112-118, 2014.
- [11] [P. Cunningham, J. Carney and S. Jacob, "Stability problems with artificial neural networks and the ensemble solution", *Artificial Intelligence in Medicine*, vol. 20, no. 3, pp. 217-225, 2000.
- [12] Almir Badnjevic, Lejla Gurbeta, Eddie Custovic. An Expert Diagnostic System to Automatically Identify Asthma and Chronic Obstructive Pulmonary Disease in Clinical Settings, *Scientific Reports*, Springer. 2018.
- [13] Lejla Gurbeta, Mirjana Maksimovic, Enisa Omanovic-Miklicanin, Almir Badnjevic, Ervin Sejdic, A telehealth system for automated diagnosis of asthma and chronic obstructive pulmonary disease. *Journal of the American Medical Informatics Association* 2018
- [14] S. Angenent, E. Pichon and A. Tannenbaum, "Mathematical methods in medical image processing", *Bulletin of the American Mathematical Society*, vol. 43, no. 03, pp. 365-397, 2006.
- [15] Y. Chen, S. Wise, V. Shenoy and J. Lowengrub, "A stable scheme for a nonlinear, multiphase tumor growth model with an elastic membrane", *International Journal for Numerical Methods in Biomedical Engineering*, vol. 30, no. 7, pp. 726-754, 2014.
- [16] F. Fogolari, A. Brigo and H. Molinari, "The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology", *Journal of Molecular Recognition*, vol. 15, no. 6, pp. 377-392, 2002.
- [17] J.O. Dada and P. Mendes, "Multi-scale modelling and simulation in systems biology". *Integrative Biology*, 3(2), pp.86-96, 2011.
- [18] D. Riano, A. Moreno, and A. Valls, "PalliaSys: Agent-Based Palliative Care". In *IEEE 4th Conf. on Intelligent Systems Design and Applications (ISDA'04)*, 2004
- [19] J.C.T. Mfumu, A. Mercier, and C. Ocelllo "Towards an Agent-based Model to Monitor Epidemics and Chronic Diseases in DR Congo". In *HEALTHINF 2018* co-located with the 11th International Joint Conference on Biomedical Engineering Systems and Technologies, 2018.
- [20] Z. He, C. Tao, J. Bian, M. Dumontier and W. Hogan, "Semantics-Powered Healthcare Engineering and Data Analytics", *Journal of Healthcare Engineering*, vol. 2017, pp. 1-3, 2017.
- [21] H. Zhang, Y. Guo, Q. Li, T. George, E. Shenkman, F. Modave and J. Bian, "An ontology-guided semantic data integration framework to support integrative data analysis of cancer survival", *BMC Medical Informatics and Decision Making*, vol. 18, no. 2, 2018.
- [22] X.Wang, N. Gulbahe and H. Yu, "Network-based methods for human disease gene prediction", *Briefings in Functional Genomics*, vol. 10, no. 5, pp. 280-293, 2011.
- [23] Y. Li and J. Patra, "Genome-wide inferring gene-phenotype relationship by walking on the heterogeneous network", *Bioinformatics*, vol. 26, no. 9, pp. 1219-1224, 2010.

- [24] D. Le, "Network-based ranking methods for prediction of novel disease associated microRNAs", *Computational Biology and Chemistry*, vol. 58, pp. 139-148, 2015.
- [25] H.Wang, H.Zheng, F.Browne, R.Roche, R.Dewhurst, F.Engel,M. Hemmje, and P.Walsh , "Analysis of rumen microbial community in cattle through the integration of metagenomic and network-based approaches". In *Bioinformatics and Biomedicine (BIBM)*, 2016 IEEE International Conference on (pp. 198-203). IEEE, 2016.
- [26] N. Friedman, M. Linial, I. Nachman and D. Pe'er, "Using Bayesian Networks to Analyze Expression Data", *Journal of Computational Biology*, vol. 7, no. 3-4, pp. 601-620, 2000.
- [27] J. Nicholson, "Global systems biology, personalized medicine and molecular epidemiology", *Molecular Systems Biology*, vol. 2, 2006.
- [28] O.Troyanskaya, K. Dolinski, A. Owen, R. Altman and D. Botstein, "A Bayesian framework for combining heterogeneous data sources for gene function prediction (in *Saccharomyces cerevisiae*)", *Proceedings of the National Academy of Sciences*, vol. 100, no. 14, pp. 8348-8353, 2003.
- [29] A. Joyce and B. Palsson, "The model organism as a system: integrating 'omics' data sets", *Nature Reviews Molecular Cell Biology*, vol. 7, no. 3, pp. 198-210, 2006.
- [30] S. Kim, S. Imoto and S. Miyano, "Dynamic Bayesian network and nonparametric regression for nonlinear modeling of gene networks from time series gene expression data", *Biosystems*, vol. 75, no. 1-3, pp. 57-65, 2004.
- [31] Ovidiu Vermesan, Peter Friess, Patrick Guillemin, Harald Sundmaeker, Markus Eisenhauer, Klaus Moessner, Franck Le Gall, and Philippe Cousin. *Internet of things strategic research and innovation agenda*. RIVER PUBLISHERS SERIES IN COMMUNICATIONS, page 7, 2013.
- [32] Dragos Repta, Mihnea Alexandru Moiescu, Ioan Stefan Sacala, Ioan Dumitrache, and Aurelian Mihai Stanescu. *Towards the development of semantically enabled flexible process monitoring systems*. *International Journal of Computer Integrated Manufacturing*, 30(1):96-108, 2017.
- [33] Jean-Paul Arcangeli, Amel Bouzeghoub, Valérie Camps, Marie-Françoise Canut, Sophie Chabridon, Denis Conan, Thierry Desprats, Romain Laborde, Emmanuel Lavinal, Sébastien Leriche, et al. *Income-multi-scale context management for the internet of things*. In *International Joint Conference on Ambient Intelligence*, pages 338-347. Springer, 2012.
- [34] Mark W Maier. *Architecting principles for systems-of-systems*. In *INCOSE International Symposium*, volume 6, pages 565-573. Wiley Online Library, 1996.