NETWORK and SYSTEMS MEDICINE
Journal of Medical Systems Biology and Network Science

An Early Stage Researcher's Primer on Systems Medicine Terminology

Journal:	*Network and Systems Medicine*
Manuscript ID	SYSM-2020-0003.R1
Manuscript Type:	Comprehensive Review
Date Submitted by the Author:	n/a
Complete List of Authors:	Zanin, Massimiliano; Universidad Politécnica de Madrid, Centro de Tecnología Biomédica Aitya, Nadim Basilio, José Baumbach, Jan; Technical University of Munich, Experimental Bioinformatics Benis, Arriel Behera, Chandan Bucholc, Magda Castiglione, Filippo Chouvarda, Ioanna Comte, Blandine Dao, Tien-Tuan Ding, Xuemei Pujos-Guillot, Estelle Filipovic, Nenad Finn, David Glass, David Harel, Nissim Iesmantas, Tomas Ivanoska, Ilinka Joshi, Alok Zouaoui Boudjeltia, Karim Kaoui, Badr Kaur, Daman Maguire, Liam McClean, Paula McCombe, Niamh Miranda, João Luís de Moisescu, Mihnea Pappalardo, Francesco Polster, Annikka; Centre for Molecular Medicine Norway Prasad, Girijesh Rozman, Damjana; Univerza v Ljubljani Fakulteta za Kemijo in Kemijsko tehnologijo, Institute of Biochemistry, Faculty of Medicine Sacala, Ioan Sanchez-Bornot, Jose Schmid, Johannes Sharp, Trevor Solé, Jordi Spiwok, Vojtech

	Spyrou, George Stalidzans, Egils; University of Latvia, Computational Systems Biology group; Latvian Biomedical Research and Study Centre, Stres, Blaž; University of Ljubljana, Department of Animal Science Sustersic, Tijana Symeonidis, Ioannis Tieri, Paolo; C National Research Council of Italy, IAC Institute for Applied Computing Todd, Stephen Van Steen, Kristel; University of Liege, Medical Genomics; WELBIO (Walloon Excellence in Lifesciences and Biotechnology), University of Liege Veneva, Milena Wang, Da-Hui Wang, Haiying Wang, Hui Watterson, Steven Wong-Lin, KongFatt Yang, Su Zou, Xin Schmidt, Harald; Maastricht University, Department of Pharmacology & Personalised Medicine, Faculty of Health, Medicine & Life Science
Keywords:	Genetics, Graphs and Networks, machine learning, Medicine
Manuscript Keywords (Search Terms):	Systems medicine, multi-scale modelling, multi-scale data science
Abstract:	Background: Systems Medicine is a novel approach to medicine, i.e. an interdisciplinary field that considers the human body as a system, composed of multiple parts and of complex relationships at multiple levels, and further integrated into an environment. Exploring Systems Medicine implies understanding and combining concepts coming from diametral different fields, including medicine, biology, statistics, modelling and simulation, and data science. Such heterogeneity leads to semantic issues, which may slow down implementation and fruitful interaction between these highly diverse fields. Methods: In this review we collect and explain over one hundred terms related to Systems Medicine. These include both modelling and data science terms and basic systems medicine terms, along with some synthetic definitions, examples of applications, and lists of relevant references. Results: This glossary aims at being a first aid kit for the Systems Medicine researcher facing an unfamiliar term, where he/she can get a first understanding of them, and, more importantly, examples and references for keep digging into the topic.

An Early Stage Researcher's Primer on Systems Medicine Terminology

Massimiliano Zanin [01,*], Nadim A. A. Aitya [02], José Basilio [03], Jan Baumbach [04], Arriel Benis [05], Chandan K. Behera [06], Magda Bucholc [07], Filippo Castiglione [08], Ioanna Chouvarda [09], Blandine Comte [10], Tien-Tuan Dao [11], Xuemei Ding [12], Estelle Pujos-Guillot [13], Nenad Filipovic [14], David P. Finn [15], David H. Glass [16], Nissim Harel [17], Tomas Iesmantas [18], Ilinka Ivanoska [19], Alok Joshi [20], Karim Zouaoui Boudjeltia [21], Badr Kaoui [22], Daman Kaur [23], Liam P. Maguire [24], Paula L. McClean [25], Niamh McCombe [26], João Luís de Miranda [27], Mihnea Alexandru Moisescu [28], Francesco Pappalardo [29], Annikka Polster [30], Girijesh Prasad [31], Damjana Rozman [32], Ioan Sacala [33], Jose M. Sanchez-Bornot [34], Johannes A. Schmid [35], Trevor Sharp [36], Jordi Solé-Casals [37], Vojtěch Spiwok [38], George M. Spyrou [39], Egils Stalidzans [40], Blaž Stres [41], Tijana Sustersic [42], Ioannis Symeonidis [43], Paolo Tieri [44], Stephen Todd [45], Kristel Van Steen [46], Milena Veneva [47], Da-Hui Wang [48], Haiying Wang [49], Hui Wang [50], Steven Watterson [51], KongFatt Wong-Lin [52], Su Yang [53], Xin Zou [54], Harald H. H. W. Schmidt [55]

[01] Centro de Tecnología Biomédica, Universidad Politécnica de Madrid, 28223 Pozuelo de Alarcón, Spain. Tel. +34 91 067 92 50. Email: massimiliano.zanin@gmail.com

[02] Intelligent Systems Research Centre, School of Computing, Engineering and Intelligent Systems, Ulster University, UK. Tel. +44 (0)28 7167 5522. Email: atiya-n@ulster.ac.uk

[03] Center for Physiology and Pharmacology, Inst. of Vascular Biology and Thrombosis Research, Medical University of Vienna, Austria. Tel. +43-1-40160-31162. Email: jose.basilio@meduniwien.ac.at

[04] TUM School of Life Sciences Weihenstephan, Technical University of Munich,

85354 Freising, Germany. Tel. +49-8161-71-2136. Email: jan.baumbach@wzw.tum.de

[05] Faculty of Technology Management, HIT- Holon Institute of Technology, 5810201 Holon, Israel. Tel. +972 (0)3-5026892. Email: arrielb@hit.ac.il

[06] Intelligent Systems Research Centre, School of Computing, Engineering and Intelligent Systems, Ulster University, UK. Tel. +44 (0)28 7167 5522. Email: behera-c@ulster.ac.uk

[07] Intelligent Systems Research Centre, School of Computing, Engineering and Intelligent Systems, Ulster University, BT48 7JL, UK. Tel. +44 28 7167 5398. Email: m.bucholc@ulster.ac.uk

[08] CNR National Research Council, IAC Institute for Applied Computing, 00185 Rome, Italy. Tel. +39 06 4993 7352. Email: filippo.castiglione@cnr.it

[09] Lab of Computing, Medical Informatics, and Biomedical Imaging Technologies,
School of Medicine, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece. Tel. +302310999247. Email: ioannach@auth.gr

[10] Université Clermont Auvergne, INRAE, UNH, Plateforme d'Exploration du Métabolisme, MetaboHUB Clermont, F-63000 Clermont-Ferrand, France; Tel: +33 473624687 Email: blandine.comte@inrae.fr

[11] Biomechanics and Bioengineering Laboratory (UMR CNRS 7338), Université de Technologie de Compiègne, 60200 Compiègne, France; and Labex MS2T "Control of Technological Systems-of-Systems", CNRS and Université de Technologie de Compiègne, 60200, Compiègne, France. Tel. 03 44 23 43 34. Email: tien-tuan.dao@utc.fr

[12] Intelligent Systems Research Centre, School of Computing, Engineering and Intelligent Systems, Ulster University, UK. Tel. +44 28 7167 5155. Email: x.ding@ulster.ac.uk

[13] Université Clermont Auvergne, INRAE, UNH, Plateforme d'Exploration du Métabolisme, MetaboHUB Clermont, F-63000 Clermont-Ferrand, France; Tel: +33 473624141 ; Email: estelle.pujos-guillot@inrae.fr

[14] Faculty of Engineering, University of Kragujevac, 34000 Kragujevac, Serbia; and Bioengineering Research and Development Center (BioIRC), 3400 Kragujevac, Serbia; and Steinbeis Advanced Risk Technologies Institute doo Kragujevac, 34000 Kragujevac, Serbia. Tel. +381-64-844-96-73. Email: fica@kg.ac.rs

[15] Pharmacology and Therapeutics, School of Medicine, Galway Neuroscience Centre, National University of Ireland, Galway, Republic of Ireland. Tel. + 353 (0)91 495280. Email: david.finn@nuigalway.ie

[16] School of Computing, Ulster University, UK. Tel. +44 28 9036 8296. Email: dh.glass@ulster.ac.uk

[17] Faculty of Sciences, HIT- Holon Institute of Technology, 5810201 Holon, Israel. Tel. +972 (0)3-5026534. Email: nissimh@hit.ac.il

[18] Kaunas University of Technology, Department of Mathematics and Natural Sciences, Studentu g., LT-51368 Kaunas, Lithuania. Tel. +37062537096. Email: tomas.iesmantas@ktu.lt

[19] Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University, 1000 Skopje, Macedonia. Tel. 00 389 70 870 924. Email: ilinka.ivanoska@finki.ukim.mk

[20] Intelligent Systems Research Centre, School of Computing, Engineering and Intelligent Systems, Ulster University, UK. Tel. +44 28 7167 5127. Email: a.joshi@ulster.ac.uk

[21] Laboratory of Experimental Medicine (ULB 222), Medicine faculty, Université libre de Bruxelles, CHU de Charleroi, Belgium. Tel. 0032 71.924705. Email: Karim.Zouaoui.Boudjeltia@ulb.ac.be

[22] Biomechanics and Bioengineering Laboratory (UMR CNRS 7338), Université de Technologie de Compiègne, 60200 Compiègne, France; and Labex MS2T "Control of Technological Systems-of-Systems", CNRS and Université de Technologie de Compiègne, 60200, Compiègne, France. Tel. 03 44 23 43 97. Email: badr.kaoui@utc.fr

[23] Northern Ireland Centre for Stratified Medicine, Biomedical Sciences Research Institute, Ulster University, UK. Tel. +44 28 7167 5662. Email: kaur-d1@ulster.ac.uk

[24] Intelligent Systems Research Centre, School of Computing, Engineering and Intelligent Systems, Ulster University, UK. Tel. +44 28 7167 5508. Email: lp.maguire@ulster.ac.uk

[25] Northern Ireland Centre for Stratified Medicine, Biomedical Sciences Research Institute, Ulster University, UK. Tel. +44 28 7167 5675. Email: pl.mcclean@ulster.ac.uk

[26] Intelligent Systems Research Centre, School of Computing, Engineering and Intelligent Systems, Ulster University, UK. Tel. +44 (0)28 7167 5522. Email: mccombe-n@ulster.ac.uk

[27] Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Portalegre, P7300-110, Portalegre, Portugal; and CERENA - Centro de Recursos Naturais

e Ambiente, Instituto Superior Técnico, Universidade de Lisboa, 1049 001 Lisboa, Portugal. Tel. +351-245.300.200. Email: jlmiranda@ipportalegre.pt

[28] Faculty of Automatic Control and Computers, University Politehnica of Bucharest, Romania, RO-060042. Tel. 0040214029167. Email: mihnea.moisescu@upb.ro

[29] Department of Drug Sciences, University of Catania, 95125 Catania, Italy. Tel. +39 095/7384223. Email: francesco.pappalardo@unict.it

[30] Centre for Molecular Medicine Norway (NCMM), Forskningparken, Gaustadalléen 21, 0349 OSLO, Norway. Tel. 0046764241771. Email: a.v.polster@ncmm.uio.no

[31] Intelligent Systems Research Centre, School of Computing, Engineering and Intelligent Systems, Ulster University, UK. Tel. +44 28 7167 5645. Email: g.prasad@ulster.ac.uk

[32] Centre for Functional Genomics and Bio-Chips, Institute of Biochemistry, Faculty of Medicine, University of Ljubljana. Tel. +386-1-543-7591. Email: damjana.rozman@mf.uni-lj.si

[33] Faculty of Automatic Control and Computers, University Politehnica of Bucharest, Romania, RO-060042. Tel. 0040214029167. Email: ioan.sacala@acse.pub.ro

[34] Intelligent Systems Research Centre, School of Computing, Engineering and Intelligent Systems, Ulster University, UK. Tel. +44 (0)28 7167 5522. Email: jm.sanchez-bornot@ulster.ac.uk

[35] Center for Physiology and Pharmacology, Inst. of Vascular Biology and Thrombosis Research, Medical University of Vienna, Austria. Tel. +43-1-40160-31155. Email: johannes.schmid@meduniwien.ac.at

[36] Department of Pharmacology, University of Oxford, Oxford OX1 3QT, UK. Tel. +44-1865-271850. Email: trevor.sharp@pharm.ox.ac.uk

[37] Data and Signal Processing Research Group, University of Vic – Central University of Catalonia, 08500 Vic, Spain; and Department of Psychiatry, University of Cambridge, Cambridge CB2 3EB, UK; and College of Artificial Intelligence, Nankai University, Tianjin 300071, China Tel. +34938815519. Email: jordi.sole@uvic.cat

[38] Department of Biochemistry and Microbiology, University of Chemistry and Technology, 166 28 Prague, Czech Republic. Tel. +420 22044 3028. Email: Vojtech.Spiwok@vscht.cz

[39] The Cyprus School of Molecular Medicine, The Cyprus Institute of Neurology and Genetics, 2370 Nicosia, Cyprus. Tel. +357-22358600. Email: georges@cing.ac.cy

[40] Computational Systems Biology Group, Institute of Microbiology and Biotechnology, University of Latvia, LV1004 Riga, Latvia. Tel. +371 29575510. Email: egils.stalidzans@gmail.com

[41] University of Ljubljana, Biotechnical Faculty, Department of Animal Science, Jamnikarjeva 101, SI-1000 Ljubljana, Slovenia; and University of Ljubljana, Faculty of Civil and Geodetic Engineering, Jamova 2, SI-1000 Ljubljana, Slovenia; and Jozef Stefan Institute, Department of Automation, Biocybernetics and Robotics, Jamova 25, SI-1000 Ljubljana, Slovenia. Tel. +386 1 3203 869. Email: Blaz.Stres@bf.uni-lj.si

[42] Faculty of Engineering, University of Kragujevac, 34000 Kragujevac, Serbia; and Bioengineering Research and Development Center (BioIRC), 3400 Kragujevac, Serbia; and Steinbeis Advanced Risk Technologies Institute doo

Kragujevac, 34000 Kragujevac, Serbia. Tel. +381-63-80-30-336. Email: tijanas@kg.ac.rs

[43] Center for Research and Technology Hellas, Hellenic Institute of Transport, GR 570 01, Thermi, Thessaloniki. Tel. +30.2310.498.482. Email: ioannis.sym@certh.gr

[44] CNR National Research Council, IAC Institute for Applied Computing, 00185 Rome, Italy; and Data Science Program, Sapienza University of Rome, 00185 Rome, Italy. Tel. +39 06 49 93 73 49. Email: paolo.tieri@cnr.it

[45] Altnagelvin Area Hospital, Western Health and Social Care Trust, UK. Tel. +44 (028) 7134 5171. Email: Stephen.Todd@westerntrust.hscni.net

[46] BIO3- Systems Genetics, GIGA-R, University of Liege, 4000 Liege, Belgium; and BIO3 - Systems Medicine, Dept of Human Genetics, KU Leuven, 3000 Leuven, Belgium. Tel. +3243662692. Email: kristel.vansteen@uliege.be

[47] Independent researcher. Tel. +359 876 204 200. Email: milena.p.veneva@gmail.com

[48] State Key Laboratory of Cognitive Neuroscience and Learning, and School of Systems Science, Beijing Normal University, 10087 Beijing, China.Tel. 86-13466388445. Email: wangdh@bnu.edu.cn

[49] School of Computing, Ulster University, UK. Tel. +44 28 9036 8908. Email: hy.wang@ulster.ac.uk

[50] School of Computing, Ulster University, Northern Ireland, UK. Tel. +44 (0)28 90368981. Email: h.wang@ulster.ac.uk

[51] Northern Ireland Centre for Stratified Medicine, Ulster University, Derry, Co Londonderry, Northern Ireland, UK, BT47 6SB. Tel: +44 28 7167 5665. Email: s.watterson@ulster.ac.uk

[52] Intelligent Systems Research Centre, School of Computing, Engineering and Intelligent Systems, Ulster University, Magee Campus, Northern Ireland, UK. Tel. +44 28 7167 5320. Email: k.wong-lin@ulster.ac.uk

[53] Intelligent Systems Research Centre, School of Computing, Engineering and Intelligent Systems, Ulster University, UK. Tel. +44 (0)28 7167 5522. Email: scott_yang15@outlook.com

[54] Shanghai Centre for Systems Biomedicine, Key Laboratory of Systems Biomedicine (Ministry of Education), Shanghai Jiao Tong University, Shanghai, China. Tel. +86 021-34206059. Email: x.zou@sjtu.edu.cn

[55] Faculty of Health, Medicine & Life Science, Maastricht University, 6229 ER Maastricht, The Netherlands. Tel. +31 43-3881421. Email: hschmidt@ppmlab.net

* Corresponding author

Running title: Primer on Systems Medicine Terminology

Keywords: Systems medicine, multi-scale modelling, multi-scale data science

Authorship Confirmation Statement:

M. Z. designed and directed the project. MZ further wrote part of the manuscript, specifically: the abstract, the introduction, and terms “Complex networks”, “Complex systems”, “Correlation networks”, “CRISP-DM”, “Cross-validation”, “Data analysis software”, “Data mining”, “Decision Trees”, “Erdős–Rényi model”, “Feature selection”, “Frequentist statistics”, “Functional networks”, “Hidden Conditional Random Fields”, “Interactome”, “Multi-layer networks”, “Network Analysis Software”, “Network medicine”, “Null models”, “Phase transition”, “Random Forest”, “Random graphs”, “Scale-free networks”, “Support Vector Machine”, “Systems biology”, and “Time-evolving networks”.

N. A. wrote part of the manuscript, specifically terms “Clinical decision support systems” and “Systems dynamics”.

J. Basilio wrote part of the manuscript, specifically terms “Gene Set Enrichment Analysis”, “Mediation analysis”, and “NetworkAnalyst”.

J. Baumbach wrote part of the manuscript, specifically the term “Systems medicine”.

A. B. wrote part of the manuscript, specifically terms “Digital Health”, “Digital Twin”, and “Medical informatics”.

C. K. B. wrote part of the manuscript, specifically terms “Quantitative systems pharmacology”, and “Systems dynamics”.

M. B. wrote part of the manuscript, specifically the term “Clinical decision support systems”.

F. C. wrote part of the manuscript, specifically the term “Agent-based modelling”.

I. C. wrote part of the manuscript, specifically terms “Complex systems”, “Exposome”, and “Physiome”.

B. C. wrote part of the manuscript, specifically terms “metabolomics”, “standards”, “ontologies”, “data fusion and data integration”, and “FAIR principles”.

T.-T. D. wrote part of the manuscript, specifically terms “Deep Learning”, “Machine Learning”, “Model Verification and Validation”, “Multiscale modelling”, and “Parameter Sensitivity Analysis and Uncertainty Quantification”.

X. D. wrote part of the manuscript, specifically the term “Clinical decision support systems”.

E. P.-G. wrote part of the manuscript, specifically terms “metabolomics”, “standards”, “ontologies”, “data fusion and data integration”, and “FAIR principles”.

N. F. wrote part of the manuscript, specifically terms “Artificial neural networks”, “Biomaterials”, “Cellular automata”, “Dissipative particle dynamics”, “Finite Element Method”, “Finite Volume Method”, “*In silico* modelling”, “Lattice Boltzmann method”, “Multiphysics systems”, “Multiscale modelling”, “Precision medicine”, “Simulated annealing”, “Smoothed-particle hydrodynamics”, “Solid-fluid interaction”, and “Surrogate model”.

D. P. F. wrote part of the manuscript, specifically terms “Clinical decision support systems” and “Quantitative systems pharmacology”.

D. H. G. wrote part of the manuscript, specifically the term “Clinical decision support systems”.

N. H. wrote part of the manuscript, specifically terms “Digital Health”, “Digital Twin”, and “Medical informatics”.

T. I. wrote part of the manuscript, specifically the term “Bayesian statistics”.

I. I. wrote part of the manuscript, specifically the term “Graph embedding”.

A. J. wrote part of the manuscript, specifically terms “Quantitative systems pharmacology” and “Systems dynamics”.

K. Z. B. wrote part of the manuscript, specifically the term “Deep Learning”.

B. K. wrote part of the manuscript, specifically terms “Deep Learning”, “Lattice Boltzmann method”, “Machine Learning”, “Model Verification and Validation”, “Multiscale modelling”, and “Parameter Sensitivity Analysis and Uncertainty Quantification”.

D. K. wrote part of the manuscript, specifically the term “Clinical decision support systems”.

L. P. M. wrote part of the manuscript, specifically the term “Clinical decision support systems”.

P. L. M. wrote part of the manuscript, specifically the term “Clinical decision support systems”.

N. M. wrote part of the manuscript, specifically terms “Clinical decision support systems” and “Quantitative systems pharmacology”.

J. L. M. wrote part of the manuscript, specifically terms “Decision Support Systems”, “Model robustness”, “Probabilistic Risk Analysis”, and “Systems Engineering”.

M. A. M. wrote part of the manuscript, specifically terms “Context Awareness Systems”, “Internet of Things”, and “System of Systems”.

F. P. wrote part of the manuscript, specifically the term “Agent-based modelling”.

A. P. wrote part of the manuscript, specifically terms “Bayesian networks” and “Clustering”.

G. P. wrote part of the manuscript, specifically the term “Clinical decision support systems”.

D. R. wrote part of the manuscript, specifically terms “Biological networks”, “Metabolomics” and “Object oriented modelling”.

I. S. wrote part of the manuscript, specifically the term “Context Awareness Systems”.

J. M. S.-B. wrote part of the manuscript, specifically terms “Clinical decision support systems”, “Granger causality” and “Small-world network”.

J. A. S. wrote part of the manuscript, specifically terms “Gene Set Enrichment Analysis”, “Mediation analysis”, and “NetworkAnalyst”.

T. S. wrote part of the manuscript, specifically the term “Quantitative systems pharmacology”.

B. S. wrote part of the manuscript, specifically the introduction and terms “metaboAnalyst”, “microbiomeAnalyst”, and “Variation partitioning”.

J. S.-C. wrote part of the manuscript, specifically terms “Morphometric similarity networks”, “Permutation test”, and “Structural covariance networks”.

V. S. wrote part of the manuscript, specifically the term “Multiscale Biomolecular Simulations”.

G. M. S. wrote part of the manuscript, specifically terms “Computational Drug Repurposing” and “Systems bioinformatics”.

E. S. wrote part of the manuscript, specifically terms “Constraints”, “Parameter estimation”, “Parameter identifiability”, and “Time scale separation”.

T. S. wrote part of the manuscript, specifically terms “Artificial neural networks”, “Biomaterials”, “Cellular automata”, “Dissipative particle dynamics”, “Finite Element Method”, “Finite Volume Method”, “*In silico* modelling”, “Lattice Boltzmann method”, “Multiphysics systems”, “Multiscale modelling”, “Precision medicine”, “Simulated annealing”, “Smoothed-particle hydrodynamics”, “Solid-fluid interaction”, and “Surrogate model”.

I. S. wrote part of the manuscript, specifically terms “Biomechanics”, “Biofluid mechanics”, and “Bioheat transfer”.

P. T. wrote part of the manuscript, specifically the term “Biological networks”.

S. T. wrote part of the manuscript, specifically the term “Clinical decision support systems”.

K. V. S. wrote part of the manuscript, specifically terms “Integrative analysis” and “Statistical networks”.

M. V. wrote part of the manuscript, specifically terms “Bayesian filtering”, “Bayesian smoothing”, and “Nvidia Clara”.

D.-H. W. wrote part of the manuscript, specifically terms “Quantitative systems pharmacology” and “Systems dynamics”.

H. W. wrote part of the manuscript, specifically the term “Clinical decision support systems”.

H. W. wrote part of the manuscript, specifically the term “Clinical decision support systems”.

S. W. wrote part of the manuscript, specifically the term “Quantitative systems pharmacology”.

K. W.-L. wrote part of the manuscript, specifically terms “Clinical decision support systems”, “Quantitative systems pharmacology”, and “Systems dynamics”.

S. Y. wrote part of the manuscript, specifically the term “Clinical decision support systems”.

X. Z. wrote part of the manuscript, specifically the term “Quantitative systems pharmacology”.

H. H. H. W. S. wrote part of the manuscript, specifically the introduction and the term “Systems medicine”.

All authors have reviewed and approved of the manuscript prior to submission.

The manuscript has been submitted solely to Systems Medicine, and is not published, in press, or submitted elsewhere.

Author Disclosure Statements:

No competing financial interests exist.

Abstract:

Background: Systems Medicine is a novel approach to medicine, i.e. an interdisciplinary field that considers the human body as a system, composed of multiple parts and of complex relationships at multiple levels, and further integrated into an environment. Exploring Systems Medicine implies understanding and combining concepts coming from diametral different fields, including medicine, biology, statistics, modelling and simulation, and data science. Such heterogeneity leads to semantic issues, which may slow down implementation and fruitful interaction between these highly diverse fields.

Methods: In this review we collect and explain over one hundred terms related to Systems Medicine. These include both modelling and data science terms and basic systems medicine terms, along with some synthetic definitions, examples of applications, and lists of relevant references.

Results: This glossary aims at being a first aid kit for the Systems Medicine researcher facing an unfamiliar term, where he/she can get a first understanding of them, and, more importantly, examples and references for keep digging into the topic.

Introduction

While death has always been the end of every man's life, mankind has been trying to delay that as much as possible. It is thus not surprising that one of the most ancient forms of science, if not the first, has been medicine, starting with documents going back to ancient Egypt and Greece [1]. In the last century, technical advances (from vaccines to genome sequencing) have supposed a revolution in medicine, and have allowed a substantial reduction in mortality rates. Yet, this trend is now experiencing diminishing returns: new drugs are nowadays developed less frequently and at a higher cost, they are beneficial to smaller subsets of the population, and consequently have less impact in life expectancy. In parallel, mankind has recently witnessed an IT revolution, in which data are gathered and processed at unprecedented rates, given birth to applications that would have appeared as science fiction as recently as twenty years ago. Following the theory of Kondratiev waves [2], postulating the existence of waves of forty to sixty years with high sectoral growth, could it be that the next wave will have medicine at its focus, and specifically through the merging of both revolutions?

Such merging is actually taking the form of the so-called Systems Medicine, an interdisciplinary field of study that looks at the human body as a system, composed of interacting parts, and further integrated into an environment [3, 4]. It considers that these complex relationships exist on multiple levels, and that they have to be understood in light of a patient's genomics, behaviour and environment. The analysis of a disease then starts with real data, coming from a large number of patients (thus to ensure that the natural variability is taken into account) and covering all aspect of them, from genetics to the environment. Machine learning and mathematical models are then developed, aimed at finding the most efficient way of disrupting the disease in a specific patient.

Even after this oversimplified description, it is clear that systems medicine requires skills and knowledge not considered in standard medical curricula, or alternatively the collaboration between researchers of different backgrounds. The

revolutionary idea behind systems medicine is thus responsible for its main drawback: the need for understanding and combining concepts coming from diametral different fields, including statistics, modelling and simulation, and data science [5]. The researcher wanting to enter this world will face an additional problem: while a large number of books and papers can be found on, e.g., data mining concepts, these are usually not written with a medical practitioner in mind. Not just the required background, but even the basic terminology can become a major barrier.

This review addresses the semantic issues this implies, which may slow down implementation and fruitful interaction between these highly diverse fields, by providing the first version of the Systems Medicine Dictionary[1]. Specifically, the practitioner coming from medicine will in it find a large number of modelling and data science terms, along with some synthetic (albeit comprehensive) definitions and a list of relevant references. Similarly, a researcher with a background in modelling and data will here find an explanation of the basic systems medicine terms. It is worth noting that these definitions are not exhaustive, as both their selection and the corresponding content has been guided by the personal view of the authors. Additionally, some terms here described represent fields of research on their own, whose characterisation can hardly be contained in a monographic book. This work thus represent the first aid kit for the systems medicine researchers facing an unfamiliar term. They will here get a first understanding of it; and, more importantly, examples and references for keep digging into the topic.

Science in general, and medicine in particular, can benefit from approaches different from what was done before, as these can have multiplicative effects on knowledge and understanding in general; this may lead to new insights and ideas for new hypotheses, and eventually to breakthroughs unattainable via the old and tested ways of thinking and acting. In turn, this requires crossing discipline boundaries and provide new angles to old information. We expect this glossary

[1] We plan this glossary to be updated in the future; we will therefore welcome any suggestion coming from readers.

to be especially useful to the younger readership, e.g. PhD students and early-career researchers, as they are at a better position to break away from old conventionalisms while significantly boosting their careers.

Concepts from Systems Medicine, Modelling and Data Science

All terms are here included in alphabetical order, and are further listed in Table 1. Table 2 also reports a list of the acronyms that appear in the text, and the corresponding meaning. Finally, underlines words, e.g. Agent-based modelling, refers to terms that are here defined.

Agent-based modelling	Artificial Neural Networks	Bayesian filtering
Bayesian networks	Bayesian smoothing	Bayesian statistics
Biofluid mechanics	Bioheat transfer	Biological networks
Biomaterials	Biomechanics	Cellular automata
Clinical decision support systems	Clustering	Complex networks
Complex systems	Computational Drug Repurposing	Constraints
Context awareness systems	Correlation networks	CRISP-DM
Cross-validation	Data analysis software	Data fusion and data integration
Data mining	Decision Tree	Decision Support Systems
Deep Learning	Digital Health	Digital Twin
Dissipative particle dynamics	Erdős–Rényi model	Exposome
FAIR principles	Feature selection	Finite Element Method

Finite Volume Method	Frequentist statistics	Functional networks
Gene Set Enrichment Analysis	Granger causality	Graph embedding
Hidden Conditional Random Fields	Imputation	In silico modelling
Integrative analysis	Interactome	Internet of Things
Lattice Boltzmann method	Machine Learning	Mediation analysis
Medical Informatics	metaboAnalyst	Metabolomics
Model robustness	Model Verification and Validation	Morphometric similarity networks
Multiphysics systems	Multi-layer networks	Multiscale Biomolecular Simulations
Multiscale modelling	Network Analysis Software	networkAnalyst
Network medicine	Null models	Nvidia Clara
Object oriented modelling	Ontologies	Parameter estimation
Parameter identifiability	Parameter Sensitivity Analysis and Uncertainty Quantification	Permutation test
Phase transition	Physiome	Precision medicine
Probabilistic Risk Analysis	Quantitative systems pharmacology	Random Forest
Random graphs	Scale-free networks	Simulated annealing
Small-world network	Smoothed-particle hydrodynamics	Solid-fluid interaction
Statistical bioinformatics	Statistical Networks	Support Vector Machine
Surrogate model	Systems biology	Systems bioinformatics
Systems dynamics	Systems Engineering	Systems medicine
System of Systems	Standards	Structural covariance networks

Time-evolving networks	Time scale separation	Variation partitioning
Virtual physiological human		

Table 1. List of the terms here described.

2SSP	Two-Stage Stochastic Programming
AAL	Ambient Assisted Living
ABM	Agent-based modelling
AI	Artificial Intelligence
ANN	Artificial neural networks
BI	Business Intelligence
BIC	Bayes Information Criteria
BPPV	Benign paroxysmal positional vertigo
CA	Cellular automata
CDSS	Clinical decision support system
CFD	Computational Fluid Dynamics
DDA	Drug-disease association
DDI	Drug–drug interaction
DPD	Dissipative particle dynamics
DSS	Decision Support System
DT	Decision Tree
EEG	Electro-encephalography
FBA	Flux balance analysis
FEA	Finite element analysis
FEM	Finite element method
fMRI	Functional magnetic resonance imaging
FVM	Finite Volume Method
GCN	Gene co-expression network
GRN	Gene regulatory network
GSEA	Gene Set Enrichment Analysis
HCRF	Hidden Conditional Random Fields

HMS	Healthcare Monitoring System
HSH	Health Smart Homes
ICT	Information and communication technologies
IoMT	Internet of Medical Things
IoT	Internet of Things
IT	Information Technology
LB	Lattice Boltzmann
LDL	Low density lipoprotein
MEG	Magneto-encephalography
MFA	Metabolic flux analysis
MICE	Multiple Imputation by Chained Equations
MMS	Multiscale Modelling and Simulation
MSC	Multiscale Computing
NLP	Natural Language Processing
PaaS	Platform as a Service
PCA	Principal-component analysis
PIN	protein interaction network
PK/PD	Pharmacokinetic/pharmacodynamic
PPI	Protein-protein interaction
PRA	Probabilistic risk analysis
QM/MM	Quantum mechanical and molecular mechanical
QSP	Quantitative systems pharmacology
RF	Random Forest
RFE	Recursive Feature Elimination
RSM	Response surface models
SA	Simulated annealing
SDK	Software Development Kit
SPH	Smoothed-particle hydrodynamics
TF	Transcription factor
t-SNE	t-distributed stochastic neighbour embedding
UPR	Unfolded protein response

Table 2. List and explanation of the acronyms used throughout the review.

Agent-based modelling. Agent-based modelling (ABM) (also known as Individual based modelling, Multi-agent Systems and Multi-agent autonomous Systems) is a modelling/simulation paradigm especially suited for studying complex systems, i.e. systems composed of a large number of heterogeneous interacting entities, each having many degrees of freedom. A very open definition of this mathematical discrete modelling paradigm is to represent a physical or biological system on the basis of entities (called agents) with defined properties and behavioural rules, and then to simulate them in a computer to reproduce the real phenomena and to perform what-if analysis [6]. Agents have thus to be understood as autonomous entities, each one with an internal state representing its knowledge about the environment, and rules (or algorithms) to interact with other agents. This broad definition can then encompass from simple particles to autonomous software with learning capabilities. To illustrate, these can be from "helper" agents for web retrieval [7, 8], robotic agents to explore inhospitable environments [9], up to lymphocytes in an immune system reaction simulation [10, 11, 12]. Roughly speaking, an entity is an "agent" if it is distinguishable from its environment by some kind of spatial, temporal, or functional attribute: an agent must be identifiable. Additionally, agents can be identified on the basis of four basic properties: autonomy, i.e. the behaviour of each agent is not guided by rules defined at a higher tier; social ability, that is, their capacity of interacting with other agents; reactivity, in that they react to perceived changes in the environment; and pro-activeness, i.e. the ability to take the initiative. Moreover, it is also conceptually important to define what the agent "environment" in an ABM is. This can be implicitly embedded in the behavioural rules or be explicitly represented as a different "modelled object" with a well-defined set of characteristics which influence the agent's behaviour.

An ABM simulation may start from simple agents, locally interacting with simple rules of behaviour, responding to perceived environmental cues and trying to achieve a local goal. Yet, the simplicity of the composing elements does not derive in the simplicity of the overall dynamics. From this simple configuration, a

synergy may emerge, which leads to a higher-level whole with much more intricate behaviour than the component agents (holism, meaning all, entire, total).

If the first examples of agent-based models were developed in the late 1940s, only computers could really show their modelling power. These include the Von Neumann machine, a theoretical machine capable of reproduction [13], i.e. of producing an identical copy of itself by following a set of instructions. This idea was then improved by Stanislaw Ulam [14], by suggesting machines to be built on paper, as collections of cells on a grid. This idea inspired von Neumann to create the first of the models later termed cellular automata (CA). Building on top of these, John Conway constructed the well-known “Game of Life”, a simple set of rules that allow evolving a virtual world in the form of a two-dimensional checkerboard, and which has become a paradigmatic example of the emergence of order in nature. How do systems self-organize themselves and spontaneously achieve a higher-ordered state? These and other questions have been deeply addressed in the first workshop on Artificial Life (ALife) held in the late 1980s in Santa Fe. This workshop shaped the ALife field of research [15], in which ABM models are the main form of modelling and simulation.

ABM proved very successful in theoretical biology. In this specific research domain, ABM is emerging as the best modelling paradigm able to accommodate the need to represent more than one level of space-time description thus fitting the multi-scale specification. Beyond the aforementioned works on the immune system, examples include cancer modelling [16, 17], or epidemics predictions [18, 19]. For further discussions and examples, the reader may refer to [20].

Artificial Neural Networks. Artificial neural networks (ANN) are inspired by the neural networks that exist in mammal brains [21]. They represent a programming paradigm that helps a computer to process complex information in order to learn from the observational data. The network itself consists of connected units or nodes called artificial neurons (based on neurons in a biological brain) that are organised in layers. The first layer is called the input layer and is connected to

the input signals. The input layer is followed by one or more hidden layers, all the way to the output layer connected to the output signals. Analogous to the synapses in a biological brain, signals are transmitted from one neuron to another. The output of one artificial neuron is computed when a non-linear function is applied on the sum of its inputs. Usually, the weights and biases are added to adjust the learning process. Weights increase or decrease the strength of the signal at a connection, and biases represent the threshold to delay the triggering of the activation function. Mathematically, this can be represented as:

$$Output = f\left(\sum weight * input + bias\right).$$

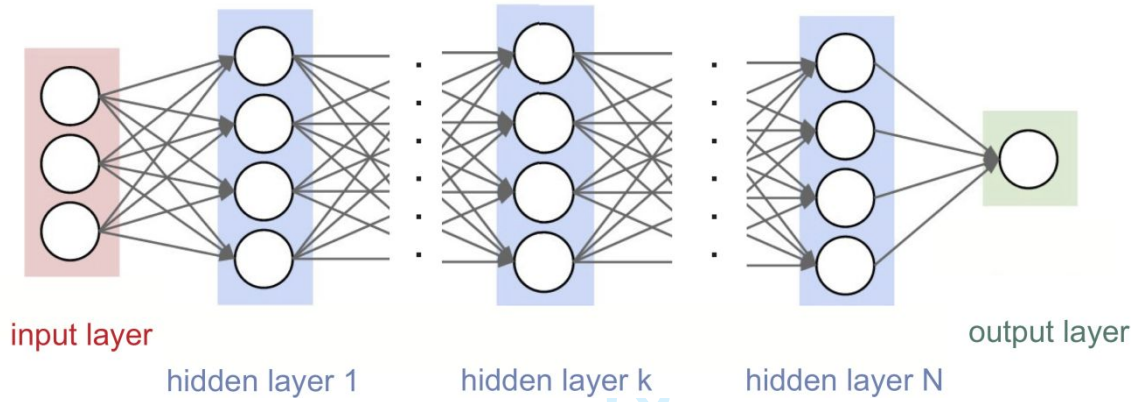

Figure 1. Graphical representation of Artificial neural network (ANN).

In order for ANN to learn from the provided data, they need to have a huge amount of information used as a training set. During the training period, the ANN's output is compared to the human-provided description of what should be observed (called *target*). If they are the same, weights are validated, and in case of incorrect classification, its learning will be adjusted [22]. In the end, an unknown signal (not used in the training set) will be used as the input, and we expect the network to correctly predict the output (this process is called *generalisation*). As an example, in the process of classification of images as images with a dog or cat, the training set would be thousands of images already classified as dog or cat image. After the training, the ANN should be able to classify future images based on the trained model.

Although ANNs were originally aimed at solving specific biology problems, over time their application extended to a wide spectrum of tasks, including systems medicine through genomics, drug repurposing, or personalized medicine. Not surprisingly, many reviews are available. For instance, Awwalu et al. investigated the adequacy of using ANN, among other artificial intelligence algorithms, in solving personalized medicine and precision medicine problems [23]. Ching et al. have developed ANN framework called Cox-nnet to predict patient prognosis from high throughput transcriptomics data [24]. Bica et al. have introduced a novel neural network architecture for exploring and integrating modalities in omics datasets, especially in cases where a limited number of training examples was available [25]. Also, some examples of application of deep neural networks could be found in using neural networks to learn an embedding that substantially denoises expression data, making replicates of the same compound more similar [26]. Donner et al. used ANNs to identify drugs with shared therapeutic and biological targets, even for compounds with structural dissimilarity, revealing functional relationships between compounds and making a step forward towards the drug repurposing based on expression data [26].

Bayesian filtering. A class of methods that allows estimating the current state, i.e. the value of the observed variable(s), based on noisy measurements of the current and previous states. For instance, the spread of infectious diseases could be modelled with the help of Bayesian filters, where the time-varying variables are e.g. estimations of the number of susceptible, infected, healed, and dead individuals taken in the current and some previous time moments [27]. For more information, see [28].

Bayesian networks. Bayesian networks (also known as Bayes networks, belief networks, Bayes/Bayesian models and probabilistic directed acyclic graphical models) are a type of directed graphical model (i.e. a graph expressing the

conditional dependencies between variables) that combine graph theory and probability theory (see also Bayesian statistics). They present a formalism designed to address problems involving uncertainty and complexity. The Bayesian network approach can be seen as both a statistical as well as an AI-like knowledge-representation formalism. It is a useful tool for describing mechanisms involving stochasticity, cohort heterogeneity and knowledge gaps, which are common features of medical problems, and has been utilised for diagnosis, treatment selection, and prognosis [29] as well as for analysing probabilistic cause-effect relationships (i.e. estimating the likelihood of a set of factors to be contributing to an observation, for example the relationship between symptoms and potential underlying mechanisms). Bayesian networks are constructed as directed acyclic graphs, where nodes represent unique variables that have a finite set of mutually exclusive states, whereas edges represent conditional dependence and the absence of edges conditional independence [30]. For each variable A with parents $B_1,B_2,...,B_n$, there is a conditional probability table P given as $P(A|B_1,B_2,...,B_n)$ [30]. Importantly, Bayesian networks satisfy the local Markov property, meaning that nodes are conditionally independent of its non-descendants given its respective parents. This characteristic permits a simplification of joint distributions within the model, allowing for efficient computation. In the most simple approach a Bayesian network is specified using expert knowledge, in the case of complex interactions the network structure and parameters need to be learned from data.

Inference and learning in Bayesian networks. Given probability tables of the variables in a Bayesian network and conditional independencies, joint probability distributions can be calculated and utilised to infer information within the network and for structural learning. This approach can be used for different probabilistic inference methods, e.g. for estimating the distribution of subsets of unobserved variables given observed variables (so-called evidence variables). Furthermore, Bayesian networks can be utilised to express causal relationships and combine domain knowledge with data, and, importantly, can thus be used for probabilistic parameter estimation.

Examples of the use of Bayesian networks in medicine include the diagnosis and prediction of disease trajectory [31, 32, 33], healthcare planning [34, 35], and molecular data analysis [36]. While this is a popular and successful option for modelling in the medical domain, they should be used with caution in complex problems with multiple feedback loos and closed-loop conditions.

Most relevant limitations. Bayesian networks commonly rely on prior knowledge/belief for construction and inference, thus the quality and usefulness of a respective network is directly dependent on the usefulness and reliability of this prior knowledge. In the case of expert-constructed networks it may furthermore be challenging to translate this knowledge into probability distributions. Bayesian networks are constructed as acyclic graphs and thus do not support the implementation of feedback-loops [37], although this may be addressed using dynamic Bayesian networks [38]. Bayesian networks have limited ability to deal with continuous variables, a limitation most commonly addressed by discretizing these variables, which in turn has tradeoffs [39]. Lastly, Bayesian learning and inference can become very computationally expensive, to the point that a network becomes impossible to compute and the search space needs to be reduced using different heuristics (for example, see [40, 41]).

Bayesian smoothing. A class of methods for reconstructing previous state(s), having noisy measurements of the current and the previous states. Brain imaging is an example of an area that can take advantage of the Bayesian filters and smoothers relying on sensor measurements of different values. For examples, see [28].

Bayesian statistics. Bayesian statistics is a Bayesian interpretation of probability in which probability expresses a degree of belief in an event, as opposed to a fixed value based upon frequency - see frequentist statistics.

The basic framework of Bayesian analysis is quite straightforward. Prior distributions are associated with parameters of interest to represent our initial beliefs about them, e.g. based on objective evidence, subjective judgment, or a combination of both. Evidence provided by further data is summarized by a likelihood function, and the normalized product of prior and the likelihood forms a posterior distribution. This posterior distribution contains all the currently available information about the model parameters. Note that this is different from the standard frequentist approach, and that both methods do not always give the same answers; and this is fuelling an ongoing debate between proponents of both approaches [42, 43, 44]. At the same time, the use of a Bayesian approach yields results that go beyond what obtainable through a frequentist perspective [45, 46, 47]. In what follows, the most important points of Bayesian and frequentists disagreements and differences are discussed: prior distributions, sequential analysis and confidence intervals.

The (subjective) choice of prior distribution. The specification of prior distribution is a matter of ongoing concern for those contemplating the use of Bayesian methods in medical research [48]. It is not without a reason that frequentists object to this concept. Any conclusions drawn from the posterior distribution will be impacted by this choice. If the prior distribution is informative, i.e. already carries strong evidence for certain values of unknown parameters, then new data might have no significant impact at all (which is not a bad thing if our prior distribution reflects the truth). Many authors devoted their thoughts to the formalization of the prior distribution selection. [49, 50, 51, 52] have all made suggestions regarding the elicitation and quantification of prior opinions of clinicians. However, it is still a very difficult task. Even minor mistakes in the prior elicitation can propagate to significant errors in the posterior inferences. The subjectivity in the elicitation of expert opinions is the main critique of Bayesian approach. Actually, in very complex problems such elicitation might even be

impossible to many parameters. However, uninformative priors, the kind that also have a claim to objectivity, are the Bayesian response [53]. In fact, there is a strong movement toward objective uninformative priors in Bayesian community.

This struggle to develop the objective Bayesian framework produced quite many different approaches on how to devise objective prior distribution. The most famous of these is the Jeffreys-rule prior [54]. Reference priors [55, 56] are a refinement of the Jeffreys-rule priors for higher dimensional problems and have proven to be remarkably successful from both Bayesian and non-Bayesian perspectives. Maximum entropy priors [57] are another well-known type of noninformative prior, although they often also reflect certain informative features of the system being analysed. Invariance priors, as mentioned above, matching priors [58] and admissible priors [59] are other approaches being extensively studied today. Methods on how to select a prior distribution from this vast universe of possible distributions are discussed in [60]. Caution is advised when considering a noninformative distribution. Sensitivity analysis should always be performed, because in small sample cases, noninformative prior distribution can still influence the posterior results [61]. On the other hand, arbitrariness is not so unfamiliar to frequentists practices as well.

Sequential analysis. The Bayesian approach includes a generally accepted stopping rule principle: once the data have been observed, the reasons for stopping the experiment should have no effect on the evidence reported about unknown model parameters. Frequentists practice, on the other hand, is different. If there are to be interim analysis during the clinical trial, with the option for stopping the trial early should the data look convincing, frequentists feel that it is mandatory to adjust allowed error probability (down) to account for the multiple analysis [42].

Stopping rules are especially important in clinical trials, and Bayesians pick up on this theme as early as 1992, with four seminal papers on colorectal cancer clinical trials [62, 63, 64, 65, 66]. Currently, Bayesian stopping rules are being used in all phases of trials - see [46] for a complete review. In fact, the increasing

use of Bayesian statistical methods in clinical research is supported by their capacity to adapt to information that is gathered during a trial, potentially allowing for smaller, but yet more informative trials, and for patients to receive better treatment [67].

Confidence intervals. The concept of confidence intervals is purely frequentists. However, the way it is (wrongly) interpreted is Bayesian. Confidence interval represents the precision of a parameter estimate as the size of an interval of values that necessarily include estimate itself. A true understanding of the concept would look like this: if new data were to be repeatedly sampled, the same analysis carried out and a series of 95% confidence intervals calculated, 19 out of 20 of such intervals would, in the long run, include the true value of the quantity being estimated [68]. However, many researchers (mistakenly and fundamentally incorrect) interpret this interval as a 0.95 probability that the true parameter is in the interval. If one would be truly Bayesian from the beginning of the analysis, Bayesian credible intervals [69] would be considered as exactly the probability that the unknown parameter is contained in it. In fact, in certain prior distribution cases, Bayesian credible intervals are exactly the confidence intervals, only the interpretation is different.

The interplay of Bayesian and Frequentist analysis. Currently, there is a trend of using notions from one type of approach to support analysis of another approach. Of many topics, several should be mentioned in this brief note: empirical Bayesian analysis, where prior distribution is estimated from the data [70]; approximate model selection methods, like BIC (Bayes Information Criteria [71]), similar to the usage of Akaike Information criteria; robust Bayesian analysis [72] which recognize the impossibility of complete subjective specification of the model and prior distribution, etc. From the frequentist theory viewpoint, the most convincing argument in favour of the Bayesian approach is that it intersects widely with the three notions of classical optimality, namely, minimaxity, admissibility and equivariance [73].

Biofluid mechanics. Biofluid mechanics is the application of principles of fluid mechanics on the dynamics of motion of biofluids inside and around of living organisms and cells [74]. The main applications of biofluid dynamics are the study of the circulatory system with the blood-flow inside vessels of various sizes, the study of the respiratory system with the air-flow inside the lungs, but also the lubrication of synovial joints [75]. The study of biofluid dynamics has allowed many therapeutic applications as artificial heart valves [76], stents and in the future artificial lungs [77]. Biofluid dynamics can be studied with simulations and experiments. Computational Fluid Dynamics (CFD) simulations can be used to better understand the flow phenomena of the biofluids inside the complex geometry of vessels. Biofluid dynamics can also be studied with *in vivo* experiments, with the use of non-invasive medical imaging methods as doppler ultrasound and magnetic resonance imaging, invasive methods as angiography but also with more straightforward methods as the pressure cuff used to measure blood pressure [78].

Bioheat transfer. Bioheat transfer concerns the rate of heat transfer between a biological system and its environment. Main difference concerning heat transfer of biological systems to non-biological ones is the blood perfusion through the extended network of vasculature in biological systems that directly affects the local temperature of the living tissue [79]. Main research subjects of bioheat transfer are the thermal interaction between the vasculature and tissue, tissue thermal parameter estimation [80], human thermal comfort, thermoregulation, safety of heat transfer to living tissue due to microwave, ultrasound or laser exposure due to environmental exposure or for therapeutic applications [81]. Because biochemical processes are governed by local temperature, bioheat transfer also plays a major role in the rate of these processes.

Biological networks. The concept of complex networks represents a powerful tool for the representation and the analysis of complex systems, and especially to describe their internal interaction structure. Recently, the so-called network biology approach [82] has been fruitfully applied in many different biological areas, from gene regulation, to protein-protein interactions, to neural signals [83], to finally hit clinical applications: network medicine is today at the forefront of modern quantitative approaches in medical sciences [84]. Here, with no claim of exhaustiveness, we list the main types of biological networks.

Protein-protein interaction networks. Protein-protein interactions (PPIs) are physical contacts, stable or transitory, between two or more proteins created by electrostatic forces between the so-called protein surfaces, i.e. the "exposed" regions of the three-dimensional structures of folded proteins. These contacts are at the base of most biological functions, as for instance of signal transduction, cell metabolism, membrane transport, or muscle contraction. It is thus clear that the analysis of how proteins interact between them is essential to understand cellular processes in healthy and in pathological conditions. Sets of proteins and their interactions are generally referred to as protein interaction networks (PINs), mathematically represented by undirected graphs. The specific analyses performed on PINs depends on the overall goal of the study; to illustrate, one may try to identify the most prominent element for a given function (e.g. gene target prioritization) [85], or the set of lethal proteins in a cell [86]. Methods for the detection of protein interaction encompass experimental (e.g. yeast-two-hybrids, mass spectrometry) or *in silico* (ortholog-based) approaches [87, 88].

Gene Regulatory Networks. Gene regulatory networks (GRNs) are networks of causative and regulative interactions (biochemical processes such as reactions, transformations, interactions, activations, inhibitions: the links) between transcription factors and downstream genes (the nodes), represented with directed graphs and inferred by gene expression data.

Methods to extrapolate GRNs are based on information-theoretic criteria, co-expression metrics, or regression approaches, among others. For example, the mutual information (MI) approach is often used, i.e. a dimensionless metric that states how much the knowledge of a random variable tells about another one. A value of MI of zero indicates that the two variables are completely independent; on the other hand, MI > 0 implies that they are connected, as knowing one of them is equivalent to (partially) knowing the other. Thus, if MI > 0 for the expression of two genes, we can infer that one of them is (partially, at least) driving the other [89].

While created in an indirect way, inferred GRNs aim at representing real physical, directed, and quantitatively determined interaction events, both between genes and, and between them and their products. The final aim is the discovery of key functional relationships between RNA expression and chemotherapeutic susceptibility [90]. Recently, data from single-cell gene expression have become mature and have been approached using partial information decomposition to detect putative functional associations and to formulate systematic hypotheses [91, 92].

Validation of GRNs has traditionally been performed in two ways. On the one hand, one can resort to “gold standards”, i.e. sets of interactions that have been validated; on the other hand, one can observe the biological system under study *in vitro*, by inducing a perturbation and by observing whether the real and predicted effects coincide [93, 94].

Gene Co-Expression Networks. Gene co-expression networks (GCNs) are basically RNA transcript–RNA transcript association networks: nodes of the network correspond to genes, which are pairwise connected when an appreciable transcript co-expression association between them exists. Networks are then calculated by estimating some kind of similarity score from expression data and by applying a significance threshold; the result is usually a undirected graph. In reconstructing GCNs, normalization methods, co-expression correlation (e.g. Pearson’s or Spearman’s correlation measures), significance and relevance

estimation are calculated. Graphical Gaussian Models (e.g. “concentration graph” or “covariance selection” models) are also used, along with edge removal based on gene triplets analysis (e.g., the ARACNE tool), regression methods and Bayesian networks [95].

Signalling Networks. Signalling pathways are cascades of molecular/chemical interactions and modifications to carry signals from cell membrane receptors to the nucleus to arrange proper biological responses to stimuli, on human or microbial levels. The process of reconstructing signalling networks has typically been based on gene knockout techniques, which are effective in describing cascades in a linear or branched manner. Nevertheless, recent screens suggest a switch from such cascades to networks with complex interdependencies and feedbacks [96], which require methods able to infer aspects and features of signalling processes from high-throughput -omic data in a faster and systemic way. In general, such inference problem can be reduced to the definition of suitable optimal connected subgraphs of a network originally defined by the available data; examples include the Steiner tree approaches (based on the shortest total lengths of paths of interacting proteins), linear programming, and maximum-likelihood (e.g. tagging proteins as activators or repressors to explain the maximum number of observed gene knockout). Alternatives include the use of probabilistic network, e.g. network flow optimization (Bayesian weighting schemes for underlying protein-protein interaction networks coupled with other -omics data), network propagation (gene prioritization function that scores the strength-of-association of proteins with a given disease), or information flow analysis (based on the identification of proteins dominant in the communication of biological information across the network) [97, 98].

Metabolic Networks. Metabolic network reconstruction is generally referred to as the annotation process of genes and metabolites for the determination of the metabolic network’s elements, relationships, structure and dynamics [83]. It can be identified on human, microbial and their joint co-metabolic levels. It is usually possible to infer the enzymatic function of individual proteins, or to reconstruct larger (or even whole) metabolic networks. Techniques such as metabolic flux

analysis (MFA) and its improvements (for example, isotopically nonstationary metabolic flux analysis), and flux balance analysis (FBA) have become largely utilized for the predictions of concurrent fluxes of multiple reactions. Recently, computational approaches coupling metabolic flux analysis with mass spectrometry have been also implemented. Single enzyme function prediction can be carried out by resorting to machine learning, especially when the enzyme does not show significant similarity to existing proteins; or to "annotation transfer" approaches, based on the use of reference databases or orthologs to tag specific DNA sequences. Comparative pathway prediction methods use established functional annotations to check for the existence of new reactions, while explorative pathway prediction techniques (not using existing annotations), can be graph-theoretic (e.g., by weighting paths of metabolite connectivity) or constraint-based (e.g., elementary mode analysis), or both [99, 100].

Transcription factor networks. When talking about disease and transformation from health to disease, we cannot avoid the transcription factor (TF) networks that were enabled by technological advances, such as genome-wide large-scale analyses, genome editing, single-cell analyses, live-cell imaging, etc. Enhancer locations and target genes are keys to TF network models [101]. The original definition of enhancers is that they represent functional DNA sequences that can activate (enhance) the rate of transcription from a heterologous promoter, independent of their location and orientation [102]. Determining the function of enhancers and whether TFs bind to them was accelerated by the CRISPR/Cas9 and other genome editing technologies, as well as by the data collected within the large-scale efforts, such as the Human Epigenome, ENCODE, etc. If we combine the experimental evidence of TFs binding to specific promoter or enhancer DNA elements, at specific genomic loci, we can construct TF network models and maps, to predict biological behaviour *in silico* and further guide experimental research. In principle, the TF network models are simple, consisting of sub-networks with nodes (genes and proteins) and edges that link the TFs to their functional targets. More complex models can nevertheless be used, for instance integrating Boolean and Bayesian approaches – see [101] for a review.

Transcription factors work predominantly in a tissue specific manner to define the cell phenotypes. For a maximal output, different TFs usually cooperate and synergise, to modulate changes in gene expression [103]. A TF network map is a graph where we can see which TFs directly regulate a gene by binding to one of its promoter or enhancer elements. A TF network map includes the basic biochemical knowledge, similarly as the metabolic network map. It links the TFs with target genes, taking into account the proper physiological or patophysiological conditions and signals (endogenous and external), as well as the context of the time (development, aging, circadian, etc.). Several approaches have been developed to model and/or graphically represent the TF networks, such as the PetriNets [104] and the ARACNE algorithm that has been recently upgraded to suit also the single-cell gene expression data [105]. The NetProphet 2.0 [106] is another algorithm for TF network mapping that can as accurately as possible identify TF targets. Another representation of TF networks are the maps that are built directly from transcriptome data by applying the enrichment procedures. These maps show if the expression of individual TFs is related. For example, the KEGG pathways [107] and TRANSFAC database were used for functional enrichment studies [108]. Gene sets containing over five elements were constructed and tested for enrichment using the *PGSEA* package and the TFs were merged based on their ID irrespective of their binding sites. In this manner the TF enrichment analyses confirmed an increased unfolded protein response (UPR) and metabolic decline after depleting one of the genes from cholesterol synthesis in the liver [109].

Biomaterials. Biomaterial is a synthetic material that is used to replace part of a living system or to function in intimate contact with living tissue [110, 111]. Although there are different definitions of a biomaterial, the Clemson University Advisory Board for Biomaterials has officially defined a biomaterial as “a systemically and pharmacologically inert substance designed for implantation within or incorporation with living systems”. One must differ biomaterial from

biological material (i.e. bone matrix or tooth enamel), which is produced by a biological system. Other materials that should be differentiated are artificial materials that are simply in contact with the skin (i.e. hearing aids and wearable artificial limbs), which are not biomaterials since the skin acts as a barrier with the external world. The main applications of biomaterials include assistance in healing, to improve function and correct abnormalities or replacement of a body part that has lost function due to disease or trauma. Advances in many fields, including surgery, have permitted materials to be used in many cases and wider scope [112, 113].

Biomechanics. Biomechanics is the application of classical mechanics to the study of biological systems. Laws of physics for statics, kinematics, dynamics, continuum mechanics and tribology are applied for the study of biological systems from a single cell to whole human bodies [114]. Biomechanics studies are employing both experiments and numerical simulations. Experiments in biomechanics are performed *in vitro* and *in vivo*. Common experiments include measurements of kinematics and dynamics of human motion (gait analysis) [115], [116], soft tissue deformation and impact studies (tension-compression tests, impact tests, three-point bending tests) [117], electromyography for neuromuscular control [118], but also experiments at microscopic level with dynamic loading of cells with microscopic cantilevers setups [119]. Simulation of biomechanics systems has allowed the testing of conditions that would be dangerous to test with human participants or biological tissue, with applications ranging from vehicle safety with simulated crash tests using active human body models, study of biological systems with complex geometries that is not possible to measure their deformation response with experiments, as brain deformation during head impacts and faster and easier to perform parametric studies. However, it is important when using a simulation model to consider the range of parameters for which the model is valid.

Cellular automata. Cellular automata (CA) are defined as abstract and discrete (spatially and temporally) computational systems that showed its application as general models of complexity and as more specific representations of non-linear dynamics in a variety of scientific fields. CA are composed of a finite (countable) set of homogeneous and simple units, called *atoms* or *cells*. These cells have an internal status that can take a finite set of values, and that is updated at each time step through functions or dynamical transition rules – generally as a function of the states of cells in the local neighbourhood. It should be mentioned that CA are abstract, meaning they can be specified in purely mathematical terms and physical structures can implement them. Since CA are computational systems - they can compute functions and solve algorithmic problems, therefore displaying complex emergent behaviour. Because of that, they are attracting a growing number of researchers from the cognitive and natural sciences interested in pattern formation and complexity in abstract setting [120]. CA have also been applied to some medical problems, as for instance image segmentation [121, 122] or infection modelling [123, 124, 125].

Clinical decision support systems. Clinical decision making involves clinicians making decisions about patient diagnosis and treatment [126]. Clinical decision making has traditionally largely been determined by human expertise. As of now, clinicians still make the final decisions upon weighing across evidence, for example from clinical data records.

Various statistical and mathematical methods [127], and knowledge-based approaches using dictionary-defined knowledge (e.g. with “if-then” rules) [128] have now been used to aid clinical decision making, resulting in more quantitative, standardized, accurate and objective decisions. This has led to the development of medical or clinical decision support systems (CDSSs), often in

the form of computer software or health technology, aiding human experts with interpretation, diagnosis and treatment [129].

The rise of artificial intelligence, particularly machine learning, has led to another form of CDSSs that is "non-knowledge-based". Some of these approaches, e.g. deep learning algorithms, have been claimed to outperform human experts in diagnosis of specific illness [130]. However, interpretability or explainability of the results of such approaches hinder their use in practice [131]. It should be noted that CDSSs still remain not as highly adopted by users, perhaps partially due to general lack of engagement from clinicians, physicians or health specialists [132].

Clustering. In data mining, any problem involving the division of data into groups (clusters), such that each one of them contains similar records (according to some similarity measures), and that dissimilar records are organised into different clusters. It is also called *unsupervised learning*, as no a priori information about the structure of the groups is used. An alternative definition of clustering is proposed in Ref. [133]: "*partition a given data set in groups, called clusters, so that the points belonging to a cluster are more similar to each other than the rest of the items belonging to other clusters.*"

While consensus on a unique classification of clustering algorithms has not been achieved, it is customary to divide such algorithms according to their underlying hypothesis [134]:

- *Hierarchical-based*. Hierarchical clustering combines instances of the data set to form successive clusters, resulting in a tree form called dendrogram. Clusters are equal to individual instances in the lowest level of the tree, and upper levels of the tree are aggregations of the nodes below. Agglomerative and divisive clustering can be distinguished, depending on whether each observation starts in its own cluster, or in the complete set.

- *Partitions-based*. As opposed to the previous group, partitions-based methods start from the complete data set and divide it into different disjoint subsets. Given a desired number of clusters, the process is based on assigning instances to different clusters and iteratively improving the division, until an acceptable solution is reached. Note that partitions-based methods are different from divisive hierarchical methods because, firstly, they require predefining the number of clusters; and secondly, because of their iterative nature. The well-known K-means algorithm [135], possibly the most commonly used clustering algorithm [136, 137], belongs to this class.
- *Density-based*. If the previously described algorithms assess the similarity of instances through a distance measure, density-based algorithms rely on density measures; clusters are thus formed by groups of instances that form a high-density region within the feature space. This presents the advantage of a lower sensitivity to noise and outliers. Among the most used algorithms belonging to this family, the DBSCAN [138] is worth mentioning.
- *Probability-based*. Probability-based clustering combines characteristics of both partitions-based and density-based approaches. The most important of these clustering approaches are mixture models [139], which are probabilistic models used to model heterogeneity and represent the presence of subpopulations (latent subgroups) in an overall population. The probabilistic component makes them a useful approach for complex (especially multimodal) data and can be used to obtain statistical inferences about the property of latent subgroups without any a priory information about these subgroups. In practice this is achieved using Expectation-Maximization algorithms [140]. Important advantages are the flexibility with regards to choosing subgroup distributions and the possibility of obtaining “soft” stratification.

Complex networks. Born at the intersection of physics, mathematics and statistics, the theory of complex networks has proven to be a powerful tool for the analysis of complex systems. Networks are mathematical objects composed of nodes, pairwise connected by links [141, 142, 143]. Their flexibility, and indeed their success, resides in the fact that the identity of those elements is not defined a priori; for instance, networks can be used to represent from people and their social connections [144], market stocks and their correlations or co-ownership [145], to genes and their co-regulation [146]. In all cases, networks allow reducing such complex systems into simple structures of interactions, which can easily be studied by means of mathematical (algebraic) tools, while removing all unnecessary details.

The simplest way of reconstructing networks, and indeed the first one from a historical perspective, is to directly map each element composing a system to a node, and map explicit relationships between elements as links. Consider the example of a gene co-regulation network: nodes would represent genes, with pairs of them being connected when it is known (e.g. from direct biological experiments) that one of the two genes is regulating the second. Once the full network is reconstructed, its structure can be studied through a broad set of existing topological metrics [147], designed to numerically quantifying specific structural features; and by using these metrics as input to data mining models [148].

In spite of the interesting results that could be obtained through this simple understanding of networks, it was soon apparent that many real-world systems needed more detailed descriptions. Specifically, it is worth noting that a simple network reconstruction implies three hidden assumptions: that links are constant through time; that nodes are connected by just one type of relationship; and that relationships are explicit. Breaking these three hypotheses gave birth respectively to time-evolving, multi-layer and functional networks.

Complex systems. Systems composed of a large number of elements, interacting in a non-linear way between them. As opposed to more simple systems, these interactions are essential to understand the behaviour of the complete system, and in some cases, they can even be more relevant than the individual elements [149, 150, 151]. Due to this, the study of complex systems goes beyond the reductionism paradigm, where understanding is based on splitting to smaller subsystems that are simpler to understand. In other words, while the reductionistic approach works bottom-up, the systems view required to understand complex systems is a top-down one. Complex systems displays two important properties. On one hand, a nonlinear behaviour, and thus tools originating in nonlinear analysis have been used in this domain – to illustrate, the analysis of time series describing the dynamics of complex systems often resort to the use of metrics of complexity [152], fractal dimension [153], sample entropy [154] and other types of entropies [155] to quantify the irregularity, or detrended fluctuation analysis to quantify long-range correlations [156]. On the other hand, emergence refers to the behaviours that may unexpectedly emerge, leading to order or disorder, and that cannot be explained by the dynamics of the system's units. Adaptation is considered as one of the qualities of complex systems, and this is a property that can be observed in the biomedical domain [157].

Computational Drug Repurposing. Drug repurposing or repositioning is the detection of novel indications for existing drugs, in order to treat new diseases [158]. A major advantage of the drug repurposing strategy is that it involves approved compounds that have passed the toxicological safety screening process and have a known pharmacokinetic profile: repositioned drugs can hence enter directly to clinical Phase II, making the clinical phase process much faster than that newly developed drugs, and thus more cost-effective. Computational drug repurposing approaches aim to optimise and accelerate the drug repurposing procedures providing also means for candidate drug prioritization. Computational drug repurposing methods include the following:

Structure-based virtual screening (molecular docking), Ligand-based methods (Pharmacophore model, Quantitative structure-activity relationship and Reverse docking methods) [159], Transcriptomic-based methods [160], GWAS-based methods [161], Literature-based discovery methods [162], and Network-based, Multi-source data integration and Machine-Learning approaches [163].

Constraints. In mathematics, constrains are conditions that must be fulfilled by some parameters (or solutions) of a model, in order to make the latter realistic. In the case of mathematical modelling of complex biological systems, different constraints can be implemented for parameters like value range of variables, limitations of sum of parameters, transition speed and other type of information. To illustrate, the angle of joints in the human arm cannot take any value, but must comply with some physical limitations [164]. There are 1) general constraints that are true for any system (mass conservation, energy balance), 2) organism level constraints - consistent limitations for all experimental and environmental conditions for a particular organism (range of viable metabolite concentrations, homeostatic constraint) and 3) experiment level constraints - environmental condition dependent constraints for particular organism (biomass composition, cellular resources) [165].

Context awareness systems. Context awareness systems address complex environments in terms of location, identity, components and relations. Context refers to information that describes an entity (person, location, object) [166]. The study of such complex environments has been made possible by the availability of Wireless Sensor Networks technologies, which allow heterogeneous sensors, distributed in a physical environment, to share their measurements. Still, these technologies do not protect from problems like cross-domain sensing and coupling of sensors; in order to preserve performance and reliability, the data fusion has to be performed with caution [167]. Context awareness systems have an important role in the design of Healthcare Monitoring Systems (HMS), Health

Smart Homes (HSH) and Ambient Assisted Living (AAL), which facilitate the acquisition of both ambient and medical data from sensors. Such systems also may include reasoning capabilities consisting of data processing and analysis as well as knowledge extraction [168].

Correlation networks. Functional complex networks created by considering the correlation between the dynamics of pairs of nodes.

CRISP-DM. CRISP-DM stands for Cross-Industry Standard Process for Data Mining, an industrial group that proposed a methodology for organising the data analysis process in six standard steps [169, 170]. Since that, the term CRISP-DM has been used to indicate both the group itself and the methodology. The six steps are:

- *Business (or Problem) understanding*: initial understanding of the objectives and requirements of the analysis to be performed; these are expressed as a data mining problem, and should include a preliminary roadmap or execution plan.
- *Data understanding*: in this second phase, data are collected and a first analysis is executed, in order to familiarise with them; identify quality problems; discover initial insights, and formulate initial hypotheses; and identify relevant data subsets.
- *Data preparation*: data received by the researchers are seldom ready to be processed; on the contrary, they usually require an initial preparation. This covers all of the activities required to construct the final data set, from selecting those data that are really relevant, to data cleaning and pre-processing. This is one of the most important steps of the whole process, as the success of the final analysis strongly depends on it; and is responsible for most of the time and resources consumed in a data

analysis project, as data preparation is usually performed iteratively and without a fixed recipe. See [171, 172, 173] for a review of techniques and the motivations for data preparation.

- *Modelling*: phase in which data mining algorithms are applied and parameters are calibrated to optimal values. Some algorithms covered in this review are Artificial Neural Networks, Decision Trees, Random Forests and Support Vector Machines. While each one of these models have specific requirements on the format of input data, and are built on top of hypotheses on the patterns to be detected, in practice multiple algorithms are suitable in any given problem. In these situations, multiple models are optimised and compared; the models reaching a higher performance are passed to the next phase for a final evaluation.
- *Evaluation*: model evaluation cannot be understood only from a data mining perspective, e.g. in terms of the achieved classification score; a business perspective should also be taken into account. Only when all relevant questions have been addressed, can one then move to the deployment of the extracted knowledge.
- *Deployment*: when all of the information about the business problems has been gathered, the information and knowledge then has to be organised and presented.

Cross-validation. In data analysis, cross-validation (also known as *rotation estimation* and *out-of-sample testing*) refers to any technique used to validate a data mining model, i.e. to quantify how it will generalise to an independent data set, re-using a single data set. The initial data set is divided into multiple subsets, which are used to train or validate the model; this guarantees that the same data are never used in both tasks [174].

Data analysis software. With the widespread adoption of data-based solutions in many real-world scenarios, it is not surprising to find a large number of analytic

solutions, spanning from cloud pipelines to commercial and freeware software, and both stemming from research activities or having a commercial nature. The most important are here listed, classified according to their underlying structure in cloud, non-cloud and hybrid tools.

Non-cloud (or local) solutions. Commercial and freeware software tools for data analysis, which are designed to work on a local (or at least, non-cloud) environment. In this category, one can find:

- KNIME [175] (www.knime.com);
- SPSS Modeller [176] (www.ibm.com/products/spss-modeler);
- RapidMiner [177] (rapidminer.com);
- Alteryx (www.alteryx.com).

These software platforms usually have a broad focus, allowing to process any (or most) kind of data; and they allow to construct models by connecting *modules* in a graphical interface.

Cloud-based solutions. Also known as Platform as a Service (PaaS), are solutions based on full cloud environments, and on the creation of web-based pipelines in which data are fed, processed, and returned to the user in a completely automatic way. The most notable solutions include:

- Google's ML Engine (cloud.google.com/ml-engine);
- Amazon's SageMaker (aws.amazon.com/sagemaker);
- Microsoft's Azure (studio.azureml.net).

This approach presents two advantages: a complete scalability, and a simplified user experience. At the same time, they usually provide a limited spectrum of possible analysis - for instance, Google ML Engine completely relies on Tensor Flow algorithms [178].

Hybrid solutions. These solutions position themselves in between the two families previously described. While they are designed for cloud deployment, they can easily be installed in a local infrastructure; and they shift the focus towards an intuitive representation of the results and simplified user experience. Among others, these include:

- Sisense (www.sisense.com);
- Looker (looker.com);
- Zoho Analytics (www.zoho.com/analytics);
- Tableau (www.tableau.com).

They usually allow to summarise data on high-level dashboards, with specific applications including business analytics [179] or website usage tracking. They nevertheless do not provide the analytical flexibility required by systems medicine applications.

Data fusion and data integration. Data fusion is the process of integrating multiple data sources to produce more consistent, accurate, or useful information than that provided by a single data source, whereas data integration refers to heterogeneous data obtained from different methods or sources, that are merged to produce meaningful and valuable information. In the field of system/personalized medicine, progress has been made regarding data integration, with large sets of comprehensive tools and methods (e.g. Bayesian or network-based methods), especially for multi-omics processing [180].

Data mining. General term describing the process of discovering patterns in data sets through the use of statistical and mathematical algorithms. Its definition

overlaps with that of machine learning; and the term is also used to denote the modelling step of the CRISP-DM process.

Decision Tree. In data mining, Decision Trees (DT) denote classification algorithms that rely on comprehensive tree structures, and that classify records by sorting them based on attribute values. Each node in a decision tree represents an attribute in an instance to be classified, while each branch represents a value that the attribute can take - see Fig. 2 for a simple graphical representation. Decision trees can be generalised to target continuous values, in which case they are usually referred as *regression trees*.

Let us denote by *D* the set of training instances that reach a node. The general procedure to build the tree is:

- If all the instances of *D* belong to the same class, then the node is a *leaf node*.
- Otherwise, use an attribute to split the set *D* into smaller subsets. These subset will then feed subsequent nodes, by applying this procedure recursively until a stop condition is met.

The main differences between the many implementations of DTs available in the literature reside in the criteria used to decide the splitting point. Among others, Gini index is used in CART [181], SLIQ [182], SPRINT [183]; information gain is used in ID3 [184] and in the well-known C.45 [185].

The main advantage of DTs is their simplicity, both in the software implementation and in the interpretation of results; and their capacity of handling both numerical and categorical variables, thus implying little data preparation. This has fostered their use in medical applications, as reviewed, for instance, in [186, 187]. They nevertheless suffer from a less-than-perfect performance. The concept of DT further underpins the Random Forest classification algorithm.

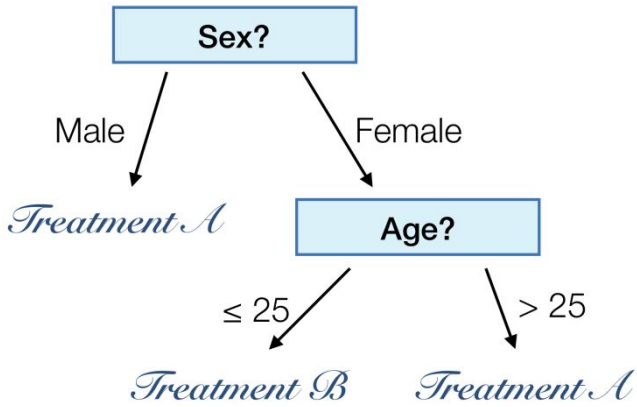

Figure 2. Example of a simple Decision Tree model, trained to choose between two treatments as a function of the age and sex of the patient.

Decision Support Systems. Decision Support Systems (DSSs) are information systems, i.e. systems designed to collect, process and make available information, focused on supporting different types of decisions [188]. DSSs typically deal with business and management challenges; can be completely customized by including multiple user interfaces and flexible architectures; and implement Optimization/Mathematical Programming tools for solution strategy and report. DSS are able to provide a complete view of the activities and flows within large and complex real production systems, integrating the supply of raw materials, the production phases, the products distribution, and the recovery within the sustainable and closed-loop supply chains. DSS in the form of standardized, enterprise-wide information systems were widely implemented in multiple sectors, including industry supply chains (e.g., pharmaceutical, manufacturing, agri-food [189]) and healthcare services (e.g., Clinical decision support systems [126, 127, 128, 129, 130]).

Deep Learning. Artificial neural networks, which form the basis of deep learning, were developed in the 1940s as a model for the human brain [190]. While this model has attracted the interest of researchers in previous periods, it made a

significant leap in learning and classification with the development of deep learning systems based on the layered learning structure of the human brain. One of the main reasons for this is that computational infrastructure needed to satisfactorily operate these complex structures that contain hundreds of layers and thousands of neurons have only appeared in the last decade.

Deep learning systems are mainly defined by the fact that each important feature of the phenomenon to be learned is automatically recognized by the algorithm and each group of features is learned by a separate artificial neural layer [191]. For example, in an image recognition system developed for human face recognition, different facets of the face, such as lines, eyes and mouths, and the general lines of the face are learned by different layers. Deep learning-based methods have greatly improved performance in Computer Vision and Natural Language Processing (NLP), and are integrated into many of the technologies currently used.

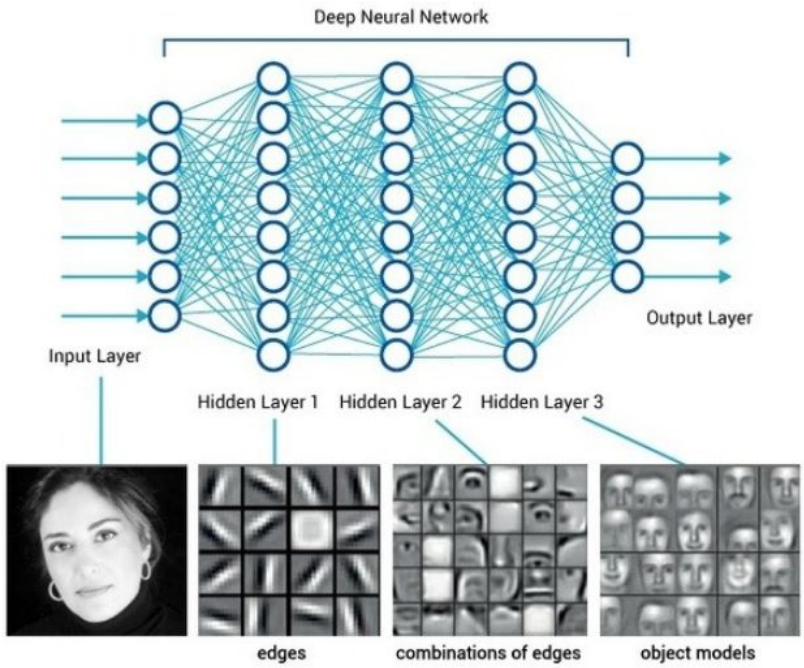

Figure 3. Deep Learning system developed for human face recognition. Source: https://www.quora.com/What-do-you-think-of-Deep-Learning-2

Digital Health. The term Digital Health (or d-Health) is used for denoting the massive and ubiquitous use of information and communication technologies (ICT) in health, healthcare, and medicine fields [192]. Digital Health covers the range of technologies used in health and medicine from genome sequencing of the microbes in the human organs, such as the gut and the skin, through genome sequencing, to the use of smartphone for supporting online telemonitoring (exposome level). The main goals of digital health are to improve healthcare customer follow-up and engagement, in parallel of resources and cost optimization from the health organizations and providers. As a part of the fourth digital revolution, "Digital Health" is using Internet of Things (IoT) and Business Intelligence (BI) for delivering personalized healthcare and medicine services. However, Digital Health is taking healthcare from a paternalistic medicine wherein physicians are defining and deciding how to treat the patient to being patient-centred. Patient-centred in the Digital Health context means that the electronic tools, hardware and software, are enhancing the healthcare customers experience and engagement by providing them with the decision support tools for getting better health outcomes and by considering their way of life and constraints [193, 194]. Nevertheless, Digital Health reduces direct human-human interactions and thus may induce a dehumanization of healthcare. Within Digital Health, a sub-subject has to be highlighted: the development of methods allowing improving healthcare customers', practitioners' and other caregivers' (like patient's family members) experience, engagement and interactions, by considering the digital environment as another kind of point-of-care similarly to clinics, pharmacies, and hospitals. One limitation of a dynamic and fast development of Digital Health lies in local regulations which have the objective of keeping health-related data and information confidential and safe, and allowing their use in ways ensuring data availability and integrity only for relevant individuals (patients and their related one when relevant, professional, and specific organizations). Digital Health is a full component of the Systems Medicine paradigm by allowing a dynamic view of individuals from the nano-level (e.g. gene expression as a response to an environmental change) to the mega-

level (e.g. population interactions/reactions -discussions- on social networks as a response to an epidemic announcement).

Digital Twin. The concept of Digital Twin is a bridge between the physical world, which can consist of a living system (i.e. an animal or a vegetal, an individual or a population) or a cyber-physical system (e.g. a biological process, a drug production line, a health monitoring service). A Digital Twin is a virtual or more accurately a computational representation of a real-world object [195]. This kind of "duplicate" is allowing designing, implementing, and testing models in a virtual environment before or instead of performing these operations in a real-world context. From a Systems Medicine perspective, the digital twin is allowing building models of living systems (from the cell components level to the world population level for building and evaluating from biological to epidemiological models) by using socio-demographics, biological, clinical, communicational data collected by healthcare customers and caregivers (see Medical Informatics) and/or generated by Internet of Things objects (see Digital Health) [196, 197].

Dissipative particle dynamics. Dissipative particle dynamics (DPD) is a stochastic simulation technique used to study dynamical and rheological properties of fluids, both simple and complex. It involves a set of particles, representing clustered molecules or fluid regions, moving in a continuous space and at discrete time steps. This meso-scale approach disregards all atomistic details that are not considered relevant to the processes addressed. Internal degrees of freedom of particles are replaced by simplified pairwise dissipative and random forces, in order to conserve momentum locally and ensure a correct hydrodynamic behaviour.

This technique facilitates the simulation of the statics and dynamics of complex fluids and soft matter systems. The main drawback is high computing power, but this has improved due to the high performance computing, which is now combined with this technique [198].

Among others, DPD can be used for modelling the transport of low density lipoproteins (LDLs) through arterial wall and analysing plaque formation, where the force of attraction of oxidase LDL molecules to the wall is modelled in the DPD solution as spring force with experimentally determined coefficient [199]; for creating semicircular canal models with simplified geometry, showing the behaviour of the fluid inside the canal, cupula deformation and movement of otoconia particles in order to analyse benign paroxysmal positional vertigo (BPPV) [200]; or for modelling self-healing materials used for corrosion analysis and protection [201].

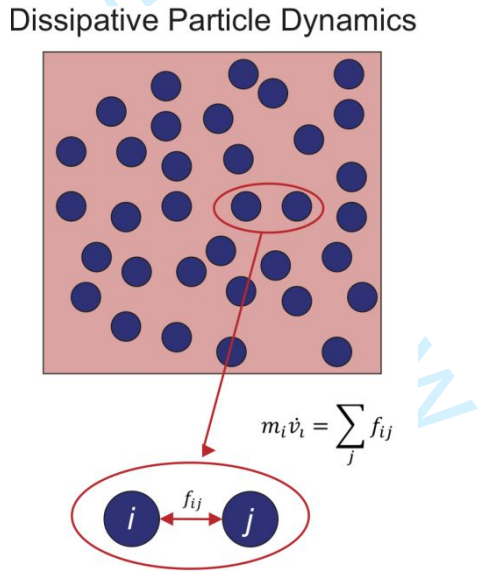

Figure 4. Schematic representation of a dissipative particle dynamics (DPD) model.

Erdős–Rényi model. The Erdős–Rényi model is a model to construct random graphs in which all edges, or links, have the same probability of existing, i.e. they are independent. The model is usually denoted as $G(n, p)$, n being the number of nodes and p the probability for any link to be present. Therefore, the model starts

with n nodes, and each possible edge is included with probability p independent from every other edge.

The simplicity of this random network model makes it an ideal candidate for act as null model in the normalization of network properties, although special care is required when the underlying real network is connected by construction, or has any other fixed characteristic [202].

This simplicity also made possible the calculation of the expected characteristics of the graph, as a function of n and p, in an analytical way. Note that all these results are of a statistical nature, and hence that the error probability tends to zero; yet, counterexamples can always be found. Among others, the most well-known ones include [203]:

- If $np < 1$, then the graph will almost surely have no connected components of size larger than $O(\log n)$.
- If $np = 1$, then the graph will almost surely have a largest component of size $\approx n^{2/3}$.
- If $p < \frac{(1-\varepsilon)\ln n}{n}$, then the graph will be disconnected, i.e. it will contain isolated nodes.
- Conversely, if $p > \frac{(1-\varepsilon)\ln n}{n}$, then the graph will likely be connected.

Exposome. Exposome is the systems approach for disease study that takes into account the interaction of internal biological mechanisms with the environment, in other words, the interplay of genetic, epigenetic and environmental factors. The concept was first introduced by Wild in 2005, and encompasses for exogenous and endogenous components [204]. A series of technological advances can be regarded as enabling technologies in this highly ambitious paradigm, including sensor networks monitor the air quality and make available the data, big data research, progress in microbiome analysis and metabolomics.

The study of endocrine disruptors and their role in pregnancy is one of the examples of this approach [205, 206]. Other work relates to cancer, and chronic diseases at large, involving pollutants, metabolism, inflammation, and diet. There are large initiatives worldwide aiming to create synergies and build knowledge on this new field of research, as for instance: https://www.projecthelix.eu/, https://humanexposomeproject.com/, http://metasub.org/.

FAIR principles. In an open-science approach, making scientific research, data and dissemination accessible, four principles for scientific data management and stewardship, were defined as Findability, Accessibility, Interoperability, and Reusability (FAIR), by the Force11 working group (https://www.force11.org/, [207]). The principles do apply not only to data but also to algorithms, tools, and workflows. These objectives are now becoming expectations from funding agencies and publishers, concerning the use of contemporary data resources, tools, vocabularies and infrastructures to assist research discovery and reuse by third-parties.

Feature selection. In data analysis, the process of feature selection consists in applying algorithms designed to select a subset of features, from the original data set, for subsequent analysis. All other features are ideally irrelevant for the problem at hand, and are thus disregarded.

Feature selection yields two main benefits. On one hand, even when the studied data set is not of large size, it can help in data understanding, reducing training times and improving prediction performance. On the other hand, feature selection is essential when the features outnumber the instances. To illustrate, domains such as gene and protein expression, chemistry or text classification are characterised by the limited availability of instances to train models – e.g. few

patients and control subjects, few complete textual records, etc. Refs. [208, 209] extensively review methods for feature selection.

Feature selection methods are usually classified in three different families:

- *Filters* select subsets of variables, according to some rules, as a pre-processing step; in other words, this selection is not made taking into account the subsequent classification. One of the most relevant examples is the Recursive Feature Elimination (RFE), based on iteratively constructing a classification model and removing features with low weights (i.e. of low relevance) – note that the classification model here used is independent from any subsequent classification. When features are added, instead of being eliminated, the result is a forward strategy.
- *Wrappers* assess subsets of features according to their usefulness to the subsequent classification problem. When the number of variables is reduced, this is done by evaluating all possible variable combinations; on the other hand, when this is not computationally feasible, a search heuristic is implemented. Note that here the machine-learning algorithm is taken as a black box, i.e. it is only used to evaluate the features' predictive power. Wrappers can be computationally expensive and have a risk of overfitting in the model [210], in which case coarse search strategies may be applied.
- *Embedded techniques* are similar to wrappers, but integrate the search of the best subset of features within the classification model [211]. The classification is then formalised as an optimization of a two-part objective function, with a goodness-of-fit term and a penalty for a large number of variables. Embedded methods that incorporate variable selection as part of the training process may be more efficient in several aspects, as they make better use of the available data and are more computationally efficient. On the negative side, they are specific to a single learning algorithm, and are thus not generalisable.

Finite Element Method. Finite element method (FEM) is a numerical method that is used for solving problems in different fields of engineering and mathematical physics. They can be widely categorized into structural analysis, heat transfer, fluid flow, mass transport, and electromagnetic potential. The finite element method formulation of the problem requires solving a system of algebraic equations. Analytical solutions of these problems generally require the solution to boundary value problems for partial differential equations. The domain of interest is divided into a finite number of simpler parts called elements and the method calculates values of the unknowns at discrete number of points over the mentioned domain. The simple equations at each point of the model are then assembled into a larger system of equations that describe the entire problem. Analysis that is associated with solving a problem using FEM is called finite element analysis (FEA) [212] [213].

Examples of the application of FEM in medicine include the analysis of bone – hip implant interactions, to obtain the information about shear stress distribution [214]; the development of several inner and middle ear models, especially cochlea models and their analysis [215]; the computational model of arteries [216, 217, 218]; the detection and localization of ischemic cardiac diseases [219]; or the examination of electrospinning jet trajectories [220].

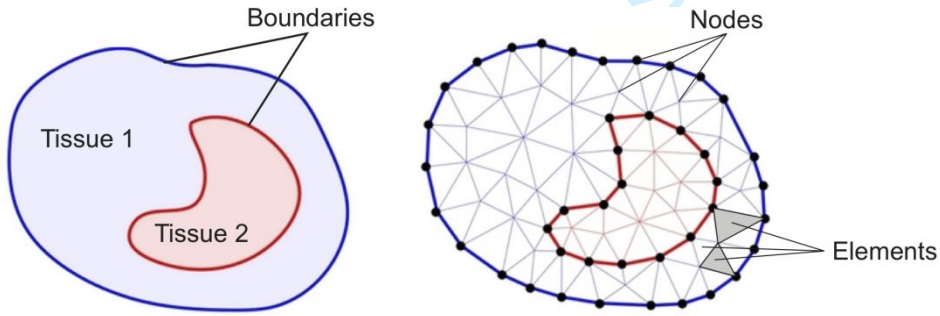

Figure 5. Schematic representation of a finite element method (FEM) model.

Finite Volume Method. Finite Volume Method (FVM) is a method that uses an approach to represent and solve partial differential equations in the form of algebraic equations. The term “finite volume” marks a small volume that surrounds each point (called node) in a mesh. By dividing the domain of interest in the form of mesh (structured or unstructured mesh), this method leads to robust schemes. Different conservation laws are used - elliptic, parabolic, hyperbolic etc. Finite volume method is often chosen when flux is of interest, since local conservativity of the numerical fluxes (conserved from one discretization cell to its neighbour) is a characteristic of this method. This is especially present in the field of fluid mechanics, semi-conductor device simulation, heat and mass transfer etc. By local conservativity it is meant that an integral formulation of the fluxes over the boundary of the control volume is obtained. A local balance is written on each discretization cell, which is called “control volume”. The fluxes on the boundary are discretized with respect to the discrete unknowns [221]. FVM can, for instance, be used in pharmacokinetic models [222].

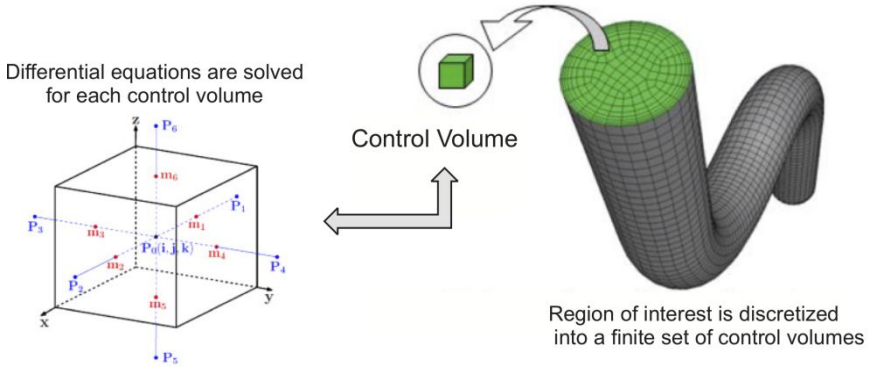

Figure 6. Schematic representation of a finite volume method (FVM) model.

Frequentist statistics. Frequentist statistics is an interpretation of statistics that considers the probability of a random event as being the long-run (in the sense of Neyman, Pearson and Wald tradition) proportion of occasions on which it occurs, conditional on some specified hypothesis [68]. For a different interpretation, see Bayesian statistics.

Functional networks. In all original studies focusing on complex networks, one inherent hypothesis was the fact that the structure of the network was easily observable: for instance, neural connections in the *C. elegans* can be obtained by physically looking at the organism. Yet, many real-world systems do not comply with this requirement: their structure is not observable, and we can only measure some aspects of the dynamics of the constituting elements. If one makes the hypothesis that the dynamics of each element is partly the result (or "the function") of the dynamics of its peers, then the structure of interactions can, in principle, be inferred from the individual dynamics: the result is called a functional network. The introduction of this latter representation has resulted in an important step forward in network science, allowing a broader focus including both structural and dynamical (functional) relations, and shifting the focus from the underlying physical structures to the flow of information developing on top of them [223, 224]. While a detailed description of the functional network theory is beyond the scope of this review, it is worth reporting a sketch of the standard way of reconstructing them. Let us suppose that a set of time series is available, each one describing the dynamics of one element (node) of the system; to illustrate, in neuroscience these typically correspond to measurements of electric (EEG) or magnetic (MEG) fields generated by the brain, or the consumption of oxygen by neurons (fMRI). The synchronicity between the dynamics of pairs of nodes is then estimated, using metrics like linear correlations or causalities. Finally, the resulting functional networks can be analysed alone, i.e. as standard networks [148]; or the relationships between the physical substrate and the functional connectivities can be explored.

Gene Set Enrichment Analysis (GSEA). Method to identify sets of functionally related genes that are enriched or depleted when comparing two biological states [225]. It does not require that individual genes are statistically scored as significantly altered, as it ranks all genes and compares this rank list with

predefined sets of genes, usually designated as molecular signatures. Since it does not require any definition of a threshold for up- or downregulation, it can identify even weaker changes of gene expression, which are significant for a gene set, but not for a single gene. The gene sets or molecular signatures used for the comparison with the rank list, are accessible through a public repository, and are based on known biological functions, pathways or cell types [226, 227]. Computation of the gene set enrichment can be performed with open software or a web platform of the Broad Institute (http://software.broadinstitute.org/gsea/index.jsp) [226]; on other web sites such as *Enrichr* (http://amp.pharm.mssm.edu/Enrichr/), or with packages of the Bioconductor R environment (https://www.bioconductor.org/). Other tools can also be used within the GSEA software:

- *Leading Edge Analysis*: examine the genes that are in the leading-edge subsets of the enriched gene sets. A gene present in many leading-edge subsets is likely to be of interest.
- *Enrichment Map Visualization*: Cytoscape plugin for functional enrichment visualization (http://www.baderlab.org/Software/EnrichmentMap)
- *Chip2Chip*: Converts the genes in a gene set from HUGO gene symbols to the probe identifiers for a selected target chip.
- *GSEAPreranked*: Runs the gene set enrichment analysis against a ranked list of genes, which you supply (e.g. mRNAseq).
- *CollapseDataset*: Creates a new dataset by collapsing each probe set into a single vector for the gene, which is identified by its HUGO gene symbol.

GSEA can also be improved by integrating external information, e.g. pathway or ontology information; some of the previously described software packages, including *Enrichr* and the *Bioconductor R* environment, include functions to perform this analysis.

Granger causality. Granger causality is a statistical method allowing to infer cause-effect relationship between events, or corresponding variables, through

exploitation of the concepts of explained variance and prediction. According to Granger [228], a signal X “Granger causes” Y if current and future values of Y can be better predicted using current and past observed values of X. Although formally known as Granger causality, this statistical method can be seen as a practical application of the earlier research in causality [229]. Since its formulation in the late 1960, Granger causality has been widely used in economics. As a result, Prof. C. W. Granger received the Nobel Prize in Economics in 2003.

The Granger causality has extensively been used in neuroscience, and specifically for the reconstruction of functional networks representing brain dynamics [230, 231] and of physiological networks in general [232]. More in general, this metric allows describing the causal relationship between pair of time series; it has thus been used to assess aspects from cardio-respiratory instability events [233], to the relationship between health care expenditure and its output [234].

Graph embedding. Graph embedding (also known as network embedding) is a representation of a graph in a vector space, where relevant graph features are preserved. Their advantage resides in the fact that vectors are easier to handle than full graphs in several domains of machine learning [148]. A lot of graph embeddings methods have been proposed for graph analysis in the following areas: nodes classification, edges (link) prediction, clustering and visualization. Graph embedding methods are categorized into three broad categories: (1) matrix factorization based, (2) random walk based, and (3) neural networks (or deep learning) based [235].

There are several challenges that need to be considered for using graph embeddings. The biggest challenge in learning a graph embedding is the choice of metrics, node and edges properties and features to be preserved in the vector representation. The learnt embeddings should represent the rich graph information including topological structure and auxiliary information. Moreover,

the graph has to be constructed in a way to represent nodes relations as well as to maintain the node proximity matrix in embedded space [236]. Next, different application domains have different prerequisites for a using a suitable graph embedding algorithm. Therefore, the embedding dimensionality decision based on graph size should meet application requirements. Unfortunately, it has been argued that in several real-world complex network applications, graph embeddings cannot represent the network's most important features [237].

In the biomedical domain, graph embeddings methods can be used to represent graphs for protein-protein interactions (PPI) [238], brain regions connections [239], infectious diseases modelling [240], chemical reactions between metabolism enzymes [241] or regulatory genes interactions [242]. [243] gives an overview and comparison of the use of graph embeddings methods in three important biomedical link prediction tasks: drug-disease association (DDA) prediction, drug–drug interaction (DDI) prediction, protein–protein interaction prediction; and two node classification tasks: medical term semantic type classification; and protein function prediction. [244] identifies relevant gene functions for a biological context using network representation learning with neural networks based graph embeddings method. In a neuroscience context, a random walk based graph embedding method is used for embedded vector representations of connectomes to map higher-order relations between brain structure and function [245].

Hidden Conditional Random Fields. Hidden Conditional Random Fields (HCRFs) are discriminative latent variable models, used for the classification of sequences of events; in other words, these models are useful to process inputs that are graphs of local observations [246]. Given one sequence, the HCRF tries to assign a single label to it, by introducing a set of latent variables corresponding to each element of the sequence, and by conditioning the label to those variables. Beyond providing rules to discriminate one label from all the others, HCRFs also

yield the structure that is shared among labels. This classification model has been proved to be efficient, provided enough instances are available to validate the hidden structure. While still not widespread in the medical domain, some applications of HCRFs include the analysis of brain dynamics [247] or the recognition of protein folding structures [248]. The main limitation of HCRFs is that no rules are presently known to define the optimal number of hidden states for a given problem; the solution, i.e. a trial-and-error process with cross-validation, can be computationally expensive.

Imputation. In statistics and data analysis, imputation refers to the set of techniques and algorithms used to handle missing data in the raw data set. These can be divided in three categories:

- Listwise deletion, i.e. the strategy of deleting any instance containing missing data. This approach, while extremely simple and easy to implement, an only be used when data are missing at random (as otherwise the deletion would introduce a bias), and when a large number of instances is initially available.
- Single imputation. Missing values are substituted by new values, according to some rules, and a new data set is therefore created. Techniques include hot-decking (when instances with missing values are substituted by other instances, chosen at random) and mean or median substitution (the missing value is filled with the mean or median of that feature).
- Multiple imputation. Missing values are replaced by values generated according to a statistical rule, e.g. Multiple Imputation by Chained Equations (MICE) [249] or Latent Class Analysis [250]. Multiple imputed data sets are generated and are analysed in parallel, for then extracting a single consolidated result.

Imputation is never perfect nor without impact. The choice of optimal missing value treatment depends on multiple factors, including the nature of data and their correlations, the amount and randomness of missing values.

***In silico* modelling.** *In silico* modelling involves the development of computer models to simulate a pharmacological or physiologic process [251, 252, 253, 254]. It is an extension of controlled *in vitro* experimentation. While mathematical electrophysiological models exist for decades (e.g. in electrophysiology of the heart), the increase in computing power available for research purposes with lower price has enabled larger scale models, for example including the cell nodes for a whole heart and incorporating personalised organ geometry based on medical imaging. Specialised platforms allow for executing the simulations and solving the numerical problems, nowadays typically in high-performance computing infrastructures. *In silico* modelling combines both the advantages of *in vivo* and *in vitro* experimentation, with the main advantage of not being subjected to the ethical considerations and lack of control that is the case with *in vivo* experiments. *In silico* models theoretically allow unlimited array of parameters to be included, contrary to the *in vitro* experiments that exist in isolation. This means that the results would be more realistic and applicable to the organism. Pharmacokinetic experimentation is often connected to the *in silico* modelling. In addition, complex *in silico* models have been applied to pathophysiological problems to provide information which cannot be obtained practically or ethically by traditional clinical research methods. These models have enabled to obtain valuable information in many fields - pure physiology, congenital heart surgery, obstetric anaesthesia airway management, mechanical ventilation and cardiopulmonary bypass/ventricular support devices. In spite of many advantages, the interested researcher should also be aware of one main drawback of *in silico* modelling, i.e. that not all strategies have been validated *in vivo* [255].

Integrative analysis. “Integration” may have different connotations, depending on the context [256]. In its most general sense, it refers to combining things, such as

two viewpoints, or multiple systems, or multiple data sets. For life science data and in particular functional genomics, Lu et al. [257] defined data integration as the "process of statistically combining data from different sources to provide a unified view and make large-scale statistical inference". For multi-omics data integration, clearly this definition is too limited, in that it only refers to statistics as a means and underappreciates the opportunities that lie in creatively combining analytic methodologies (for instance, statistics and machine learning). A more challenging definition for data integration in complex disease analysis involves the process of combining data within a generic framework that encompasses organizing principles for the interaction of different types of systems. This definition does not explicitly refer to statistical, bioinformatics or computational tools but to any approach that fits within a transdisciplinary viewpoint. It includes data fusion as well as more fancy and more elaborate forms of combining evidence from different data sets or sources [258]. Furthermore, it agrees with the definition of Oxley and Thorsen [259] as the process of connecting systems (which may have fusion in them) into a larger system. Apart from data integrative analysis, integrative analysis sometimes also refers to the integration of analytic tools or methods, to combine different analytic viewpoints to the same data.

Interactome. Map representing the whole set of molecular interactions in a particular cell. While usually interactome specifically refers to physical interactions, it can also be used to describe sets of indirect interactions among genes. As molecular interactions can occur between any pairs of molecules composing the cells (including proteins, nucleic acids, lipids, carbohydrates, and so forth), a great number of interactome maps can be defined; nevertheless, the most common and well-known include:

- The protein–protein interaction (PPI) network (PIN);

- The protein–DNA interactome, also called a gene-regulatory network, a network formed by transcription factors, chromatin regulatory proteins, and their target genes;
- Metabolic networks, representing metabolites and how they are converted into each other by enzymes.

For the corresponding mathematical representations of such maps, see biological networks.

Internet of Things. Internet of Things (IoT) is related to the evolution of the internet towards integrating real, everyday life devices called things.

A comprehensive description is provided in [260]: 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 aims at achieving a virtual representation of a set of physical devices through the deployment of technologies and architectures involving large-scale, loosely coupled systems.

Generally speaking, basic IoT systems components include: IoT Standards and Ecosystems, Event Stream Processing, IoT Device Management, IoT Platforms, IoT Analytics, and IoT Security [261]. An important aspect is the IoT Reference Model, the model that defines all architectural aspects of the system, and which is composed of the following sub-models: IoT Domain Model, IoT Information Model, IoT Functional Model, IoT Communication Model, and IoT Security Model [260]. Moving from a theoretical to a physical representation of IoT, this is usually composed of: Smart devices, Network, Data processing, Data storage, Data aggregation, data analytics, and process integration.

Communication between IoT elements can be addressed through multiple paradigms: 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 Modelling and Simulation (MMS) can be related to the following components: IoT as data provider for Multiscale Modelling and Multiscale Modelling as a way to experiment and validate complex processes with the aid of IoT.

Many synergies have been found between IoT systems and Multiscale Modelling. First of all, IoT can facilitate data provision to the modelling phase, by handling access, routing and recording of data acquired from sensors attached to smart objects. Secondly, IoT devices naturally measure the physical space at different resolution and conceptual levels, thus providing a multiscale view of the space. In addition, IoT can simplify the understanding of the raw data through technologies related to Big Data, semantic representations, ontologies and machine-interpretable representations of domain knowledge, and context awareness.

Multiscale IoT Systems for Experimental Multiscale Models can be used to acquire data at multiple scales corresponding to the scales selected in the Multiscale Model. Such IoT systems design use multiscale principles. 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 [262].

An important development of Internet of Things with applications in medicine is referred as Internet of Medical Things. The Internet of Medical Things (IoMT) can be described as an internet based environment connecting medical devices and services. Applications of IoT technologies in medicine are increasingly common

[263, 264, 265]. In cancer treatment studies blood pressure monitoring bracelets and tracking apps have been used to gather relevant information. Continuous Glucose Monitor (CGM) can be connected in an IoT environment to transmit data to mobile devices thus facilitating the analysis of blood glucose levels. A Bluetooth-enabled coagulation system has been used in connection to IoT environment in order to help patients become aware of potential blood clots and transmit results to healthcare providers. A wearable smart asthma monitor can detect symptoms related to asthma attacks and connected to an IoT environment can track and detect the inhaler.

Lattice Boltzmann method. Lattice Boltzmann (LB) method is a discrete numerical method used mainly for simulations of fluid flow [266, 267, 268, 269, 270]. The main advantage of this method is that it is not necessary to solve differential equations, which makes the implementation relatively simple and it is possible to parallelize the software. In LB method, fluid is observed as a set of fictional particles. These particles can move along the predefined directions, and the dynamics of their motion is modelled through their mutual collisions and further propagation in the observed domain. A special distribution function is defined, and this function depends on the state of neighbouring particles and has an identical form for all the particles, i.e. for all the nodes in the lattice mesh. Macroscopic quantities, such as density, pressure, velocity, are calculated using the components of the distribution function [271, 272].

Examples of the use of the Lattice Boltzmann method in medicine include the modelling of the motion of endolymph through the semicircular canals of the inner ear [273, 274]; and the analysis of the numerical and experimental transport of low-density lipoproteins (LDLs) through arterial walls [275]. Open-source software implementing LB methods are also available, see for instance https://www.openlb.net and https://palabos.unige.ch.

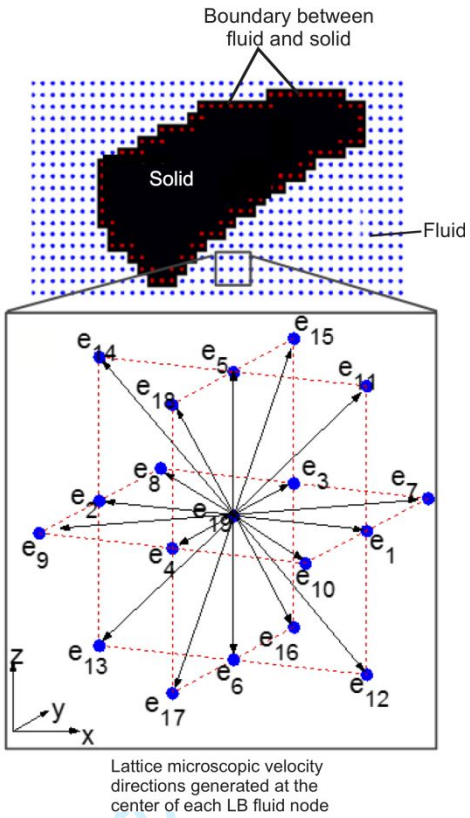

Figure 7. Graphical representation of the Lattice Boltzmann (LB) method.

Machine Learning. Machine learning is the science of using computers to discover new information from observations [276, 277]. There are several families of machine learning methods: supervised learning, unsupervised learning and semi-supervised learning. The choice of the strategy depends on the nature of the used data. A large and complex database is commonly required to develop a machine learning model. In system medicine field, bio-marker extraction or human genome classification is typical example of machine learning model. For further details, see also data mining, CRISP-DM, deep learning.

Mediation analysis. If two variables (an independent x and a dependent y) show a statistically significant correlation, it does not necessarily mean a direct causative link, as the correlation might be caused by a third variable (the mediator), which is often non-observable – and which is influenced by the independent variable and by itself influencing the dependent variable. A

mediation analysis can elucidate such interactions and dependencies and helps to differentiate between direct and indirect effects [278, 279]. This type of analysis can be performed with specific packages of the Bioconductor R environment or with add-ins of commercial software such as SPSS. It is important to note that a mediation effect can be full or partial – and that it can be moderated by additional parameters. Additionally, it has to be stated though that mediation analysis cannot be used to detect or analyse multiple interdependencies.

Medical Informatics. Medical informatics (also known as Health Informatics or Biomedical Informatics) is a science at the crossroad of information science, computer science, social sciences, and health and medical sciences. This research area deals with all the components of information systems (data acquisition, information and knowledge resources, devices and networks, regulation and ethics, and more) used for supporting and improving healthcare management (e.g. clinical knowledge management), delivery (e.g. patient-related data follow-up over time) and research (e.g. developing standards encoding diagnostic for epidemiological purposes) [280, 281, 282, 283]. Medical Informatics is an umbrella and the core for different sub-specialities such as clinical informatics, nursing informatics, public health informatics, consumer health informatics, and veterinary informatics. As a multidisciplinary field, the Medical Informatics playground consists of developing and investigating theories, models, methods, processes and systems, used for generating, storing, retrieving, using and sharing health and medical data, information, knowledge, and decision support. From an application perspective, medical informatics is actively and dynamically investigating and supporting health and medical reasoning by experimenting models and simulations across a wide spectrum: from molecules to populations, from a biological system point-of-view to a global population and One Health perspective. Moreover, end-users are a crucial component of the overall system in Medical Informatics. For efficiency reasons, researchers in the field of Medical Informatics have to continuously monitor the changes in different spheres such as the social, economic, ethical and

educational, and update their models in accordance to these changes. In recent years there has been an important and growing trend of applying algorithms and know-how from the fields of Business intelligence and automation in Medical Informatics, e.g. data and text mining, analysis, and information and knowledge management – see clinical decision support systems. From the integrative perspective of systems medicine, Medical Informatics investigates and delivers end-to-end frameworks supporting complex medical decisions, driven by evidence-based medicine for continuously improving health and disease management at the individual and populations levels [284]. One of the most critical parts of research done in Medical Informatics considers ethical and legal regulations and constraints in the technological side of medical field [285]. As new means of measuring, communicating and managing patients emerge, there is a need to continuously monitor and update the requirements for ensuring security, i.e. keeping confidentiality, integrity, and availability of health and medical data sensitive data.

metaboAnalyst. Part of the same family of websites including networkAnalyst and microbiomeAnalyst, this web site provides a visual analytics platform for meta-analysis of metabolomics data (www.metaboanalyst.ca) [286].

Metabolomics. Metabolomics is the scientific study of a set of metabolites present within an organism, cell, or tissue. It was also defined as a global measurement of small molecules (metabolites), which are produced or modified in an organism. Metabolites can also result from a stimuli (nutritional intervention, drugs, genetic perturbations, etc.), are present in a system (blood, urine, saliva, etc.) and accessible to analysis [287, 288]. Metabolomics is one of the functional level tools being employed to investigate the complex interactions between metabolites but also their regulatory roles through their interactions with genes, transcripts and

proteins. It is actually considered as a powerful phenotyping tool to better understand the biological mechanisms involved in the pathophysiological processes and identify biomarkers of metabolic deviations [289]. Indeed, it provides, at a molecular-level, multivariate information of multi-compartmental biological systems that reflect changes in biological processes [290].

microbiomeAnalyst. Part of the same family of websites including networkAnalyst and metaboAnalyst, this web site provides a visual analytics platform for meta-analysis of microbiome data (www.microbiomeanalyst.ca) [291].

Model robustness. Model robustness is a widely used concept in modelling under uncertainty, namely with Robust Optimization approaches. For that, the objective function of a Stochastic Linear/Quadratic Programming is modified by introducing penalization parameters related with non-desired attributes (e.g., high variability on solutions, non-satisfaction of products demands, over-designing of production capacities, non-utilization of expensive equipment), or probabilistic restrictions are modified by enlarging/narrowing "soft" bounds (e.g., "worst case" analysis) [292].

For instance, the Two-Stage Stochastic Programming (2SSP) [293] approach for the capacity expansion of a pharmaceutical supply chain allows both the promotion of solution robustness (by penalizing the deviations on the solutions, e.g., minimizing the solutions variance) and the model robustness (e.g., minimizing the expectances for the non-desired attributes). Namely: *i)* at the first stage, the capital and investment decisions must be taken (that is, the project variables are calculated "here-and-now"); *ii)* in the second stage, the uncertainty is introduced through a set of scenarios and the related probabilities (in this "recourse phase", it occurs the probabilistic calculation of the control variables).

Then, model robustness is obtained when the optimal solution does not present high values for the probabilistic measures of the attributes to avoid (namely: for the expectance of excess/unused production capacities that would imply larger investment costs; and for the expectance of unsatisfied products demands that would impact negatively the patient's health). Model robustness is also strongly connected with other concepts of interest, such as Model Verification and Validation, Parameter Sensitivity Analysis and Uncertainty Quantification, Probabilistic Risk Analysis. Several drawbacks can occur on model robustness developments, e.g., due to resource consuming, standard accuracy, or uncertainty see [294, 295] for details.

Model Verification and Validation. Model verification is a process to verify if a given model has been directly coded or mathematically represented; on the other hand, model validation aims at verifying if the implemented model is the right one for the biological system of interest. Model verification is a straightforward task, thanks to many direct techniques to check and debug computer programs. Model validation, on the other hand, is more complex, and is commonly performed using theoretical outcomes or experimental measurements. It is important to note that model validation of biological systems is extremely complex and difficult due to the lack of *in vivo* data and measurement protocols [296, 297].

Morphometric similarity networks. Morphometric similarity networks are graph-based representations of the structure of the brain [298]. The study of structural differences in the brain by topological analysis based on graph theory has the disadvantage of generating a connectivity matrix at the group level and, therefore, the connectivity parameters are calculated at the group level. Recently, a new technique has been developed that allows to generate a connectivity matrix at subject level based on the interregional similarity of multiple morphometric parameters measured by multimodal MRI [298]. Typical morphometric

measurements taken from multimodal image data for each brain region are: fractional anisotropy (FA), mean diffusivity (MD), magnetization transfer (MT), grey matter volume (GM), surface area (SA), cortical thickness (CT), intrinsic (Gaussian) curvature (IC), mean curvature (MC), curved index (CI) and folding index (FI). For each subject, these values will form a vector of morphometric measurements for each region. Then, the morphometric similarity matrix (MSM) of the subject will be obtained by calculating the Pearson's correlation between the vectors of the morphometric characteristics of each pair of regions. Finally, the morphometric similarity network (MSN) will be obtained by thresholding this MSM. Therefore, we end up with one network (MSN) per subject, which will allow us to calculate the (structural) connectivity parameters at the subject level. Recently some papers have been published that demonstrate the validity of this technique [299, 300].

Multiphysics systems. Multiphysics systems are systems consisting of more than one component, each governed by its own principle(s) for evolution or equilibrium (conservation or constitutive laws) [301]. Two possibilities for classification are related to the coupling:

- bulk couplings, i.e. through relations that are active in the overlapping domains of the individual components;
- couplings happening on idealized interfaces of lower dimension, e.g. through boundary conditions that transmit fluxes, pressures, or displacements.

Some examples of bulk-coupled multiphysics systems include radiation with hydrodynamics in astrophysics, electricity and magnetism with hydrodynamics in plasma physics (magnetohydrodynamics), and chemical reaction with transport in combustion or subsurface flows (reactive transport). Since forward models are simulated successfully, inverse problems, sensitivity analysis, uncertainty

quantification, model-constrained optimization, and reduced-order modelling are gaining more attention. The physical model is, in these advances, augmented by variables other than the primitive quantities in which the governing equations are defined. These variables may be sensitivity gradients, probability density functions, Lagrange multipliers, or coefficients of system-adaptive bases. Equations that govern the evolution of these auxiliary-dependent variables are often derived and solved together with other physical variables [302]. For an example of applications of multi-physics systems to medicine, see [220].

Multi-layer networks. Complex networks whose interactions are defined on more than one layer. In the standard complex network approach, links between nodes are usually of a single type, the only difference between them being a (generally, real) number, quantifying the weight of the connection. Nevertheless, considering all links as homogeneous can be an important constraint, as connections in real-world systems may be of different types. A biological example can help clarify this. One of the most interesting success in recent neuroscience has been the creation of a full map of the *C. elegans*' neural network, consisting of 281 neurons and around two thousand connections [303]. Yet, connections are not homogeneous: neurons can communicate through chemical and electrical (ionic) links, with completely different dynamics and time scales. Therefore, a correct representation should include two independent layers of connections. This resulted in the creation of the multi-layer network concept, i.e. graphs whose connections are organized in separate layers [304]. Multi-layer networks explicitly incorporate such heterogeneity, such that each link type (relationship, activity, category) is represented by a different layer, with the same node having different neighbours in each layer.

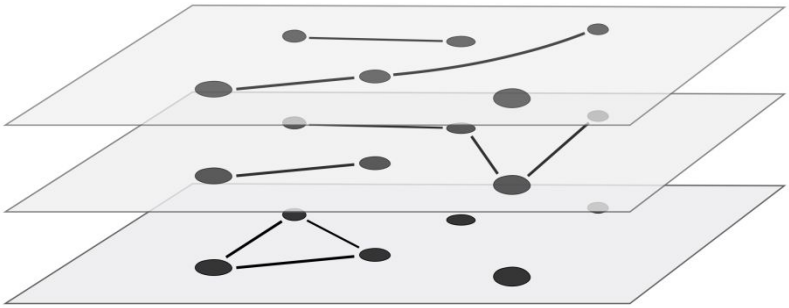

Figure 8. Example of a graphical representation of a multi-layer network composed of three layers.

Multiscale Biomolecular Simulations. Biomolecular simulations are computer simulations of molecular dynamics of biological systems, such as proteins, nucleic acids, saccharides, membranes and their complexes. Multiscale biomolecular simulations are simulations of molecular dynamics of biological systems at different levels of granularity, differing in spatial resolution and other aspects.

First attempts to simulate molecular systems started in 1950s. The first biomolecular simulation was published in 1977 by J. Andrew McCammon, Bruce R. Gelin and Martin Karplus (2013 chemistry Nobel Prize winner) [305]. The authors simulated several picoseconds of bovine pancreatic trypsin inhibitor in vacuum. An important milestone of biomolecular simulations was the development and refinement of biomolecular force fields (formulas and their parameters for calculation of potential energy from atomic coordinates) and simulation software. Packages CHARMM, AMBER, Gromos, Gromacs, NAMD, ACEMD and BOSS have been tuned for high performance on a wide range of machines and operation systems.

There are several types of granularity in multiscale biomolecular simulations. The main reason for interest in multiscale versions of biomolecular simulations is in the fact that these simulations are extremely computationally expensive. Each atom in a typical solvated biomolecular system interacts (covalently or non-covalently) with another approximately 5.000 atoms. These interactions must be

evaluated in every simulation step. The integration step of most biomolecular simulations is in a femtosecond scale. It is therefore necessary to carry out millions of steps (and evaluate interactions of millions of atomic pairs in each step) to simulate nanosecond time scales.

The first type of granularity is in modelling of interaction between atoms. There are two major models that make it possible to calculate energy and forces in a molecular system - quantum mechanics and molecular mechanics. Quantum mechanics models the system by solving Schrödinger equation for electrons. On the other hand, molecular mechanics represents atoms as particles connected by simple mechanical "springs" and interacting via interatomic potentials with simple mathematical descriptions. Electrons are not explicitly modelled. Quantum mechanics calculations are significantly more complex and, therefore, more computationally expensive. The advantage of quantum mechanics is that it does not require *ad hoc* sets of parameters for each class of molecules. Furthermore, most molecular mechanics models do not take into account the reactivity of the molecular systems. Molecular mechanics (with few exceptions) keeps the chemical structure fixed during the whole simulation, i.e. it disallows breakage and formation of covalent bonds in chemical reactions. For this reason quantum mechanics is used to study the mechanism of chemical reactions.

Enormous computational costs of quantum mechanics led to a mixed (multiscale) model of quantum mechanical and molecular mechanical (QM/MM) calculations. For example an enzymatic reaction can be studied on a model of enzyme with the substrates and active-site residues modelled by quantum mechanics and the rest of the system modelled by molecular mechanics.

This second type of granularity addresses the number of particles in the molecular system. These models differ in the number of atoms represented by a single particle. In a standard fine-grained ("all-atom model") model there is one particle representing one atom. All quantum mechanical models are all-atom models. Simplified versions called "united-atom models" represent certain groups of atoms, such as CH, CH_2 and CH_3, as a single particle. Such particle represents

the bulk properties of the whole group. This reduces the overall number of particles in the system and accelerates the simulation without significant loss of resolution.

Further coarse-graining in so-called "coarse-grained models" replaces multiple atoms, typically four non-hydrogen atoms, by a single particle. Coarse-grained simulations make it possible to study several orders of magnitude longer time-scales than all-atom simulations. The prize paid for this is loss of resolution. Coarse-grained simulations have been extremely successful in simulations of membranes, interfaces and related systems. They are less frequently used in studies requiring precise atomic resolution, such as in drug discovery. Models mixing all-atom and coarse-grained simulations (similarly to mixed QM/MM models) have been developed to address this problem.

There are examples of studies with further coarse-graining. For example, elastic network models of proteins represent individual amino acids as particles connected by harmonic springs. This representation of a protein resembles models used in civil engineering to test mechanical stability of constructions. They are used in biomolecular simulations, but more frequently, they are studied by static approaches such as normal mode analysis. Surprisingly, bulk mechanical properties of biomolecules can relatively accurately predicted using such simplified models.

The major aim of biomolecular simulations is to predict certain property of the biomolecular system. The third type of granularity is in depiction of such molecular properties. Biomolecular simulations produce trajectories - thousands of snapshots of thousands of atoms. These pieces of big data can be analysed to extract relevant low-dimensional properties of the systems. Such properties can be than used to build thermodynamic and kinetical models of the simulated system.

The last granularity is the computational granularity. As already mentioned biomolecular simulations are computationally expensive. Most software used in

biomolecular simulations has been developed to run in parallel on multiple cores of a CPU (multithreading) and multiple CPUs and node connected by Message Passing Interface. Recently Fast Multipole Method [306] is being introduced into biomolecular simulations in order to enable multiple levels of parallelism. Alternative hardware such as graphical processing units and special purpose hardware have been successfully used. The multiscale nature can be further extended by application of special multiple ensemble or multiple time scale methods.

Multiscale modelling. Multiscale modelling is a numerical approach to study the biological systems of interest at multiple time and length scales, i.e. in which multiple models at different scales of time and/or space are used simultaneously to describe one complex system [307]. To illustrate, a multi-cellular organism can be modelled at different levels, e.g. DNA, cells, fibres, and tissues; with each model getting input from the lower-level one [308].

Those models are commonly developed using a combination of several numerical methods. Finite element method could be used to model system behaviour at organ and tissue scales. Agent-based simulation could be used to model single cell or cell population behaviours. Molecular dynamics could be used to describe the movements of atoms and molecules. To make the link between scales, homogenization theory could be used. This theory allows constitutive behaviours at the macroscopic level to be described using the information from interactions between macroscopic and microscopic levels. There are two main multiscale modelling strategies. The first one is the hierarchical simulation in which the system behaviour is separately described and simulated for each scale and then the interaction is performed. The second one is the concurrent simulation in which all system behaviours and their interaction are simultaneously described and simulated. There is no time delay by using the

second strategy but the strategy is complex for model development and implementation.

The importance of multiscale modelling lies, on one hand, in the fact that available macroscale models are usually not accurate enough, and on the other hand, in the fact that microscale models are not efficient enough and/or offer too much information. By integrating both approaches, the idea is to find a compromise between accuracy and efficiency [309].

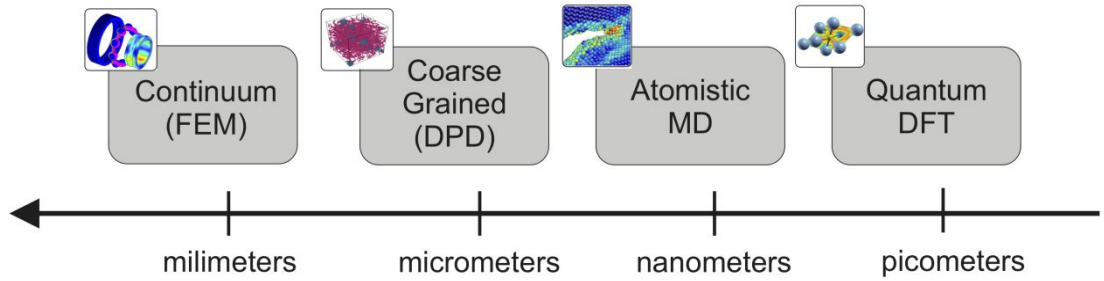

Figure 9. Graphical representation of the typical scales in a multiscale modelling.

Network Analysis Software.

- NetworkX [310]. Python library used for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks. This allows the creation of networks with different algorithms, evaluation of a large set of standard metrics, and finally display the results in an easily understood way. Freeware. Available at networkx.github.io.
- Cytoscape [311, 312]. Software specialized on the representation of networks, with some additional tools for the integration of biological data. It also provides some basic network analysis capabilities. Freeware. Available at www.cytoscape.org.
- Gephi [313]. Interactive visualisation and exploration platform. Freeware. Available at gephi.github.io.

- Pajek [314]. Software for representing complex networks, with some basic analysis capabilities. Freeware. Available at mrvar.fdv.uni-lj.si/pajek/.
- VisANT [315]. Software for the visual study of metabolic networks and pathways. Freeware. Available at visant.bu.edu.
- IBM ® i2 Analyst's Notebook. Software for the integration of social data and network analysis. Commercial. Information at www-03.ibm.com/software/products/en/analysts-notebook.
- SAS ® Social Network Analysis. Software for the analysis of social networks. Commercial. Information at support.sas.com/software/products/sna/index.html.

networkAnalyst. Part of the same family of websites including metaboAnalyst and microbiomeAnalyst, this web site provides a visual analytics platform for meta-analysis of differentially expressed genes or proteins (www.networkanalyst.ca) [316, 317]. It allows input of raw RNA-sequencing data, single or multiple gene expression tables or pre-calculated lists of differentially regulated genes with expression values. The input is then compared with known interaction networks covering not only various protein-protein interactomes, but also relations between genes and miRNAs; transcription factors, drugs or chemicals. By default, a first order network is computed, which can also be switched to a second order network to increase the number of interactors, or the zero-order network to decrease the number of nodes. If the complexity is too high, it can be reduced with filters on betweenness or degree. Another option is to calculate a minimum network, which comprises the least number of nodes that are required to link the input genes. The network can be downloaded in a Cytoscape-compatible SIF-format, but the standard routine is to visualize it within the web platform in an adjustable manner including up- or downregulation of expression levels and different layouts, which can be saved in SVG-format. Moreover, and most importantly, the network can then be statistically compared with different databases such as KEGG, Reactome, gene ontologies or transcription factor motifs to obtain functional

enrichment values. A module explorer can be applied to extract subnetworks with statistically elevated links and these can be further analysed for functional gene enrichments.

In case that the differential expression is computed on the *NetworkAnalyst* platform, gene clustering can be performed comprising heatmaps, principal-component analysis (PCA) or t-distributed stochastic neighbour embedding (t-SNE). Moreover, Gene Set Enrichment Analysis can be done and Venn- or Chord diagrams can be created for multiple comparisons.

Network medicine. General term to design applications of complex networks theory to medicine, and hence to the identification, prevention and treatment of diseases [84, 318]. It is buttressed by the idea that elements constituting our bodies at all scales (e.g. from genes, to cells and organs) do not exist in an independent fashion, but are rather connected by a dense set of interdependencies. Understanding one disease thus goes beyond the simple analysis of one element. For further examples, see biological networks.

Null models. In complex networks theory, a null model consists of a set of networks with some characteristics equal to the graph under study, while being random in all other aspects [319]. The simplest case is therefore a set of completely random networks, i.e. Erdős–Rényi graphs, which share the same number of nodes and links, but are otherwise completely random.

The main advantage provided by null models is that they allow breaking the coupling existing between different topological properties, and thus allow comparing networks with heterogeneous characteristics. To illustrate, the value

of a given topological metric can be normalized with what expected in the null model, thus helping to assess whether the observed value is special or, on the contrary, is the result of the other restrictions imposed in the model. The simplest solution involves the calculation of a Z-Score, which indicates how many standard deviations the observed metric is from the (null model's) expected value [202].

Nvidia Clara. Nvidia Clara is a computational platform that gathers CUDA accelerated tools for medical imaging and genomics. The Software Development Kit (SDK) provides libraries for computing, visualization and AI. The SDK allows the users to deploy their applications in any GPU platform they have access to. Within this platform, Nvidia Clara Medical Imaging provides tools for data annotation, training of AI models, and deployment in the case of medical imaging applications (e.g. computerized tomography (CT), magnetic resonance images (MRI), ultrasound, X-ray, and mammography). Adapting one of the included in the SDK pre-trained AI models with transfer learning accelerates the AI modelling as less time and training data are used. On the other hand, the Nvidia Clara Genomics platform gathers CUDA accelerated tools for genomics sequencing and analysis. Biomedical examples of the use of Nvidia Clara include the segmentation of images of brain tumours [320], and gene sequencing [321].

Object oriented modelling. For effective diagnosis and treatment of diseases we need to understand the dynamics of metabolism, including the metabolism of drugs. Here, the large scale computational models that describe dynamics from the metabolic, gene regulatory and signal transduction perspectives are of crucial value [322]. Different modelling approaches are in use, including the object oriented modelling. This technique is originally derived from machinery. Dymola (Dynamic Modeling Laboratory) has been developed by Dassault Systems, a branch of the Dassault group that produces also airplanes. Dymola sets the basics of object oriented modelling of the biological systems even if its initial

intention has been for use within automotive, aerospace and robotics process. In Dymola we can describe the entire multi-component systems and in this manner represent the real world as good as possible.

The basics of object oriented modelling is represented by a library of objects. An object is an element corresponding to components of mechanical, electrical, vehicle dynamics, etc., and also biological systems. In building the model, the objects from the library are moved by drag-and-drop and interactions between the model components are described by graphical connections that model the physical coupling of the components. The unique feature of object oriented modelling is that the models are intuitively organized to mimic the real physical or biological systems. In systems medicine we can imagine that large macromolecules (genes, mRNAs, proteins including enzymes and transcription factors, etc.) are objects. The signalling pathways represent links or information that is transferred through connections between these objects.

Nowadays, Modelica is used as the most popular programming language for object-orienting modelling. The benefit of Modelica is that the users can create their own libraries. *BioChem* has been designed as a library for metabolic pathways [323] that describes enzymatic reactions in different biochemical pathways. *SysBio* library [324] was initially used to construct the *SteatoNet* model with multi-layered regulation, including the transformation of genes to proteins and the transcriptional regulation [325]. Additionally, *SteatoNet* describes multiple tissues i.e. the liver and adipose tissue and their connections through the blood.

The beauty of object oriented modelling is that the number of parameters that need to be incorporated into the model is small. We can thus avoid problems with parameter estimation or model overfitting. This is possible due to observation of the normalised steady-state of the system's response, allowing modelling in the absence of parameters that describe the dynamics of the observed system. Another benefit of this type of modelling is the ability to incorporate specific data

towards i.e. personalisation. In this manner, the *LiverSex* has been produced as the first model describing the distinct liver metabolism of females and males [326].

Ontologies. Ontologies (also known as controlled vocabularies and semantic representation) can be defined as formal representations of knowledge in a certain domain, in an understandable way for people and computers [327]. They are made of defined classes of entities, structured in hierarchy where concepts are connected with standardized relationships [328]. In biomedical research, a great variety of ontologies have been developed to describe domain knowledge, for example, the Gene Ontology (GO) or the Disease ontology. BioPortal is a repository of biomedical ontologies, many of which can be openly reused. In addition, the Open biomedical Ontologies (OBO) is an established platform developed for interoperability and shared principles between ontologies [329]. The question of ontology relevance in the context of systems medicine has been particularly discussed. In fact, because of its intrinsic paradigm change, such ontologies must switch from a biological structure to a biological function architecture [330]. Beyond the existing ontologies, the US National Research Council proposed a new taxonomy for biology and medicine taking into account the multiple aspects of basic science and clinical characteristics to define disease endotype [331]. The development of phenotype-driven ontologies is also of great interest for the field [332]. However, with the explosion of heterogeneous clinical data and scientific information, harmonization between scientific communities as well as their participation to computational resources are essential for the future of ontologies in translational research and precision medicine [333].

Parameter estimation. Mathematical models in systems biology and systems medicine have a structure that characterizes interactions between elements of the system. Next level of detail are the parameters of interactions to quantify the intensity of interaction. Some of model parameters can be measured or found in

the literature while information about others is missing. Parameter estimation [334] can be used to estimate the unknown parameters by fitting of the model to the available experimental data. Usually it is solved as a numerical optimization problem where the difference between measured data and model calculations have to be minimized searching the best combination of unknown parameter values. Parameter estimation can have several results:

- The model behaviour fits the experimental data. It is not expected that model behaviour would match each and every measurement as they contain measurement errors and mathematical models are always simplifications of reality. Even in case of success, parameter identifiability should be checked (see Parameter identifiability).
- The model behaviour does not fit well to the experimental data. There can be several reasons: model definition and range limitation of estimated parameters have to be checked. Another problem can be the selection of inappropriate optimization method that leads to local minimum or stagnates [335].
- The model cannot reproduce the expected type of behaviour. This may be an indication that the structure of the model does not correspond to the system of interest; and that, without suitable changes in the model structure, a satisfactory behaviour as well as an identification of parameters cannot be reached.

Parameter identifiability. In case of successful parameter estimation, model parameters cannot be always trusted [334]. It can happen that a value of a particular parameter is not important for particular experimental set-up and any value can produce acceptable fit of model with experimental data. Another parameter unidentifiability reason can be structural unidentifiability [336] where the structure of model in combination with experimental results does not allow identification of particular parameters. For instance, if just summary flux of two

parallel metabolic pathway branches is measured, parameters defining each particular flux cannot be identified.

Parameter Sensitivity Analysis and Uncertainty Quantification. Parameter sensitivity analysis and uncertainty quantification are two important best practices when developing and simulating biological systems of interest. Parameter sensitivity analysis allows to determine which parameters are sensitive to the input variations with the used constitutive laws [337, 338]. This analysis is commonly time-consuming due to the repetitive nature of the procedure. Moreover, the determination of a plausible perturbation value range is also a difficult issue. A relative percentage (e.g. $\pm 10\%$) is usually used. Uncertainty quantification aims to model the uncertainties related to the system input values or variables and their propagation on the model outcomes through the used constitutive laws. A lot of data is commonly needed for uncertainty quantification. Data assumption could be performed with limited data samples but the accuracy level is questionable. Precise and imprecise probabilities could be used to model uncertainties. Monte Carlo is a classic example of uncertainty propagation method [339].

Permutation test. When we have to test between-group differences, for one or more values per subject, we can use a (non-parametric) permutation test to infer whether the difference between the two values is statistically significant or not. To do so, we need to generate random groups by shuffling the labels of the groups. The metric differences between the two resulting random groups are then used to create a reference distribution for each metric in order to reject or retain the null hypothesis that there are no differences between the groups. To ensure that the reference distribution is appropriate we need to generate thousands of random groups. With 1.000 random groups the smallest possible *p*-value is 10^{-3}, while with 100.000 random groups the smallest possible *p*-value decreases up to

10^{-5}. A practical way is to start with a not too large number of random groups, for instance 1.000, and increase this number to a larger one if the *p*-value is small enough to be interesting. Because this calculation can be computationally demanding, sometimes parallel computing is needed. One way to avoid it is to use other techniques based on tail approximation, which obtain accurate *p*-value with a drastically reduced number of permutations [340]. A typical case in which we will need to use the permutation test is when we are willing to test between-group differences in structural covariance analysis. In this case, we have the connectivity matrix at the group level and therefore the global connectivity measures are also at the group level. Testing differences between group level measures will require a permutation test.

Phase transition. The original meaning of the term *phase transition* is to be found in statistical physics, and especially in thermodynamics. When one defines the *phase* of matter as a state in which it has uniformly physical properties, a phase transition occurs when that matter undergoes a transformation between two states. To illustrate, water and ice are two phases (respectively liquid and solid), and the transition between both of them (i.e. the freezing process) is a phase transition. The term is nevertheless also used in a more general sense, to indicate any transition between two homogeneous and easy identifiable conditions of a system. For instance, when deleting nodes from a complex networks to simulate an attack to the system, the initial connected status and the final disconnected one are two phases, with a transition in between them [341].

Suppose one analyses the evolution of some metric describing the system as a function of an external parameter; in the previous example, the former can be the connectedness of the network, which is studied as a function of the number of removed links. Two types of transitions can then occur:

- First-order phase transitions, which exhibit a discontinuity in the first derivative of the metric (solid red line of Fig. 10). This implies that the system has an abrupt reaction to the change in the external parameter.
- Second-order phase transitions are continuous in the first derivative, but usually exhibit discontinuity in a second derivative (dashed blue line of Fig. 10). The response of the system is therefore smoother than in the previous case.

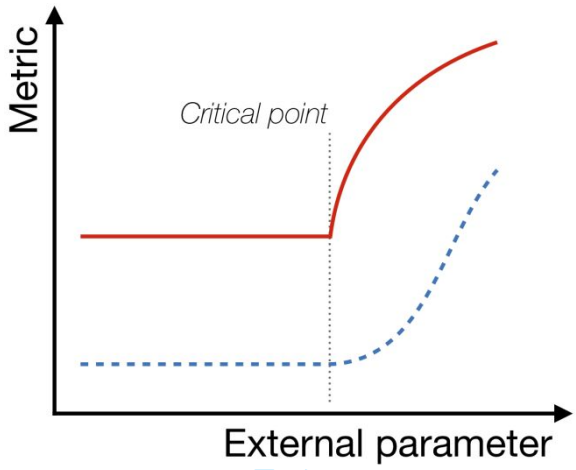

Figure 10. Example of two phase transitions, a first-order (red solid line) and a second-order one (dashed blue line).

Physiome is a multi-scale approach aiming to functionally synthesize models at different levels, and understand human physiology based on computational models [342]. Standardisation of models has been part of this effort, and an important number of models is now available in the physiome repository (https://models.physiomeproject.org/welcome).

A flagship project has been the cardiovascular physiome, which aimed to use integrative multi-scale modelling and link the whole heart function with small scale systems and phenomena (e.g. ion channel mutations, ischaemic tissue, drug toxicity, biochemical pathways), always with an eye towards providing tools for the clinician to investigate hypotheses and interpret experimental data. Within the physiome paradigm, the virtual physiological human (https://www.vph-institute.org/), has been a long term initiative to embrace systems medicine at

organism level, towards integrating all information available for each patient, and generating computer models to predict patient's health evolution.

Precision medicine. According to the HORIZON2020 Advisory Group (EU Health Ministers – December 2015), precision medicine is "a medical model using characterization of individual's phenotypes and genotypes (e.g., molecular profiling, medical imaging, lifestyle data) for tailoring the right therapeutic strategy for the right person at the right time, and/or to determine the predisposition to disease and/or to deliver timely and targeted prevention." Precision medicine is then an approach to patient care that promotes the idea of doctors selecting most adequate treatments for patients based on a genetic understanding of their disease. This idea does not literally mean to create the drugs or medical devices that are specific for a patient, but divide the individuals into clusters (subpopulations) that differ in their susceptibility to a particular disease, biology or prognosis of those diseases or response to specific treatments and select treatment based on that knowledge [343]. Preventive or therapeutic interventions can then be concentrated on those who will actually benefit and save expenses on unnecessary treatments and side effects in patients that do not. Older synonym for precision medicine was "personalized medicine", which was often misinterpreted as implying that unique treatments can be designed for each individual. As a result, the term "precision medicine" was created [344].

Probabilistic Risk Analysis. Probabilistic risk analysis (PRA) is aiming at quantitative measures for evaluation the risk of system failures (*e.g.*, supply of essential medicines within a healthcare system, availability of innovative drugs and active ingredients in the pharmaceutical sector, disruption of agri-food supply chains in natural disasters, security issues in the nuclear power industry), in which the common statistical analysis is very difficult or even impossible due to multiple

and disparate issues (*e.g.*, non-existence of pertinent data, the system complexity, the uncertainty about consequences) [345].

The probabilistic risk is related with the probability distributions for the losses in a given time horizon, while PRA methods also includes event trees, fault trees, and Bayesian networks. The PRA approach typically considers: *i)* identification of failure scenarios; *ii)* computation of scenarios probabilities, by combination of events probabilities and the associated random variables distributions; *iii)* the evaluation of consequences, the extension and impacts of those scenarios. The data obtained in this way can then be used to feed a robust model with multiple goals, namely, by minimizing the expectance of system failure for a given budget (and/or for a given schedule), while verifying if the probabilistic measures for risk failure are satisfactory.

Probabilistic risk analysis is also strongly connected with other concepts of interest, such as Model robustness, Model Verification and Validation, Parameter Sensitivity Analysis and Uncertainty Quantification. Difficulties are usually associated with the scenarios definition, the selection of random variables distributions and events probabilities, as well as sparsity and high-dimensionality.

Quantitative systems pharmacology. Quantitative systems pharmacology (QSP) or systems pharmacology modelling is a computational and mathematical modelling approach that simulates the mechanistic effects of drug effectiveness [346]. QSP combines pharmacokinetic/pharmacodynamic (PK/PD) modelling with systems biology and systems engineering [347, 348]. It integrates drug pharmacology, physiology, mathematics and biochemistry, and accounts for drug liberation, absorption, disposition, metabolism and excretion. QSP, which is a type of *in silico* modelling, typically makes use of differential equations to model the dynamics of the drug interacting with the biological system. More recently, QSP involves genomic, transcriptomic, metabolomic and proteomic levels, as well as regulatory and epigenomic levels. QSP is increasingly being used in

pharmaceutical research and development to help guide the discovery and development of new treatments and therapies, and to extrapolate animal data to humans [349, 350, 351]. This is in line with recent directions in stratified medicine or precision medicine, by which model parameters can be tuned to simulate specific biomedical type. The advancement in big data and data science is gradually forming an integral part of QSP, complementing its traditional mechanistic modelling.

Random Forest. In data mining, Random Forests (RFs) are classification algorithms based on combining multiple Decision Trees (DTs) models. The underlying concept is that an ensemble of models, each one independently trained on a subset of the data and each one casting a vote about a particular instance, could yield a better result than a single model, especially in problems are characterized by a large number of variables, each one of them encoding very little information. Following this idea, Random Forests are created by merging multiple DT predictors, each one trained using a different subset of the initial data [352]. Each tree in random forest is grown as follows: *i*) sample with replacement a given number of cases from the training set at random. This sample will be the training set for growing the tree; *ii*) given M input variables, randomly select $m \ll M$ of them at each node, and choose the best one to split the node; *iii*) grow the tree with no pruning. Given one new instance, the final classification corresponds to the class voted by the majority of the trees. While there is no strict rule about the optimal number of trees to be grown, studies suggest that little is gained by going over 1.000 trees [353].

Random forests have three significant advantages: first, they do not suffer from overfitting, and can thus be use in small data sets. Second, their computational cost is reduced, and are very prone to parallelization (as each tree can be created in an independent process). Finally, they have been shown to outperform most known algorithms, in terms of accuracy [354]. On the negative side, it is worth

noting that the number of trees in the model must be selected by the researcher, and that not clear rules are available to guide this process.

Random graphs. Random graphs are graphs, or networks, that are artificially constructed by creating links between nodes according to a given probability distribution [355, 356]. As such, they do not correspond to any real-world system; but they instead provide a tool for answering specific questions about how some properties may appear. Due to the lack of any pre-defined structure, except for those naturally arising from the defined probability distribution, random graphs are well suited to be used as null models.

Scale-free networks. A scale-free network is any complex network whose degree distribution approximatively follows a power law; in other words, the fraction of nodes with degree k goes as $P(k) \approx k^{-\gamma}$, with γ being a parameter usually in the range (2, 3). Many real-world networks, including biological ones [357, 358], have been found to be scale-free to some degree [359, 360], although no consensus still exists on the best way of statistically test such property [361].

Scale-free networks are of relevance for different reasons.

First of all, the degree distribution implies that most nodes have very few connections, while a (statistically significant) high number of them concentrate the majority of the links; these latter ones are thus more important for the functioning of the network, or more central, and are usually called “hub”.

Secondly, the structure induced by scale-freeness implies a great resilience against random disruptions; note that, if a node is deleted at random, there is a high probability for that node to be secondary and weakly connected. On the other

hand, a targeted attack can do much damage, as it can target a node of very high centrality [362, 363].

Finally, several models have been proposed to explain the appearance of scale-free networks [364, 365, 366, 367]; and, more generally, the presence of such structure can point towards the existence of some generative processes.

Simulated annealing. Simulated annealing (SA) is a form of optimization that is used to approximate global optimization in a large search space. This method is used in discrete space, where finding an approximate global optimum is more important than finding a precise local optimum in a fixed amount of time. In these situations, simulated annealing is often preferable to methods such as gradient descent. It is especially useful in finding global optima when large numbers of local optima are present. Simulated annealing uses the objective function of an optimization problem instead of the energy of material. Implementation of SA consists of hill-climbing and picking a random move, instead of the best move. If the selected move improves the solution, it is accepted, and when not, it moves with probability less than 1. The value of probability decreases exponentially with the amount of how much the solution is worsened [368, 369]. Beyond general optimisation problems (see for instance [370, 371, 372]), SA has extensively been used for segmenting medical images [373, 374].

Small-world network. The theory of small-world networks [375] is based on the observation of biologic or complex systems that can be represented using graphical models. The specific graph shows especial characteristics, such as having a high clustering of its elements, and a very fast association between any two different nodes that can be inferred by following the shortest path between the nodes through the graph connections.

The formulation of small-world networks was inspired by the idea that the "degree of separation" or distance between two different (unfamiliar) persons on the Earth is about five [376]. Not only social networks have been observed to follow this pattern, network of collaborators, complex systems and brain networks also follow this interesting rule.

A small-world network can be also explained as the transition from random or chaotic systems to highly regular or structured ones. For example, in a regular lattice network, where the nodes only have connections to the closest or adjacent nodes, it can be observed that by disconnecting and randomly reconnecting the nodes, the average distance between any two nodes in the network rapidly decays while maintaining the local network of closest nodes only decay slightly in density (clustering coefficient). In neural networks this property of small-worldness can be seen as critical to maintain a fast integration among distant neural population in order to process information efficiently, while the different tokens of information are locally processed in highly dense local networks.

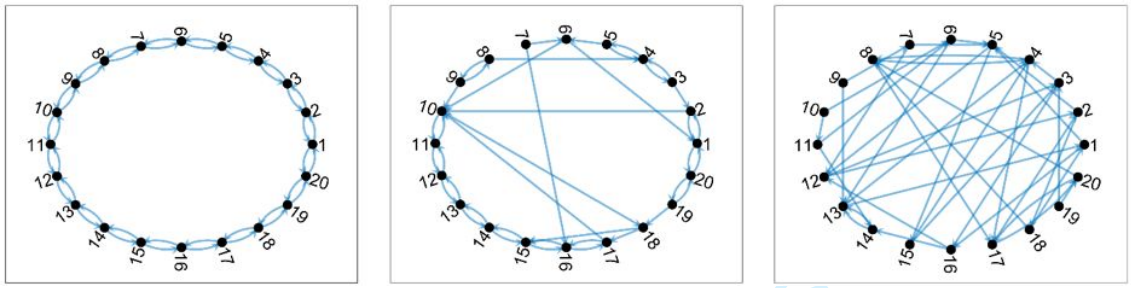

Figure 11. Example of the creation of a small-world network.

Smoothed-particle hydrodynamics. Smoothed-particle hydrodynamics (SPH) is a computational method that is used for simulating the mechanics of continuum media, such as solid mechanics and fluid flows [377]. Many fields of research have employed SPH method, such as engineering, astrophysics, ballistics, volcanology, and oceanography [378, 379, 380]. It is a meshfree Lagrangian method, meaning there is no division of domain of interest in the form of mesh (see Finite Element Method and Finite Volume Method), but rather the coordinates move with the fluid. In such way, the resolution of the method can

easily be adjusted with respect to variables such as density. Here, the computational domain is discretized by a finite set of interpolating points (particles) with invariant coordinates in the material frame. Each SPH particle represents a finite mass of the discretized continuum and carries the information about all physical variables which are evaluated at their positions. Interpolating (smoothing) function and its derivatives at surrounding particles are used to evaluate the function values and their derivatives at a specific particle [381]. SPH has been used, for instance, to model therapeutic solutions aimed at helping heart muscle to regenerate after an injury [382].

Solid-fluid interaction. Solid-fluid interaction is a numerical approach that is used to model phenomena that involve both the surrounding fluid and immersed solid objects. Using this approach, both domains are simulated concurrently, and they form a coupled mechanical system. The fluid is acting on the solid object via external forces and causes the motion and deformation of the deformable solid and vice versa – the solid is opposing the deformation and influence of fluid and this way alters the fluid flow. Solid-fluid interaction techniques have been applied, for instance, in modelling the deployment of stent within stenotic artery with deformable arterial wall [383]; in simulating the behaviour of deformable cells within a fluid flow [384, 385]; and in providing insight into the benefits of different treatment alternatives in a case of type B aortic dissection [386].

Statistical bioinformatics. Application of statistical techniques to large sets of biomedical data – mainly genomics data, but recently this has evolved to include any type of -omics data. For more information, refer to [387, 388, 389, 390].

Statistical Networks. One of the properties of a system is that it consists of interacting components at different levels. Creating a corresponding network may

be based on biology (see Biological Networks) or may be based on analytical arguments, or both. Statistical epistasis networks belong among the simplest examples of such networks, in which nodes refer to units of analysis and edges are formed via a notion of statistical significance. They have become popular tools in genome-wide association interaction studies to highlight higher-order interactions in typically underpowered studies [391]. In general, the major challenge with statistical networks is to assess and minimize statistical artefacts that may hamper network-derived biological conclusion-drawing [392].

Support Vector Machine. Binary linear classifiers based on the identification of hyperplanes in the feature space, dividing the training instances in two groups according to the training label. The model is trained by firstly constructing a feature space, i.e. a hyper-space defined by the features available in the data set, which must always be numerical. Records are mapped into this space, and the best linear separation between them is then calculated. The best separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class, as this minimises the error. Modified version of SVMs have been developed to tackle different problems, including regression problems [393], or the use of different kernels (i.e. distance functions) to obtain non-linear models [394]. Among SVM's disadvantages are a high computational cost, and the complexity of dealing with classifications with multiple labels. For more details, refer to [395, 396].

Surrogate model. Surrogate model is an engineering method that is used when an outcome of interest cannot be easily directly measured, and instead, a model of the outcome is used. In many real-world problems, one simulation can take from minutes, to hours and even days to finish the calculation. Therefore, sometimes design optimization, sensitivity analysis and what-if analysis are impossible to investigate, since that would mean running thousands or even

millions of simulations. Surrogate models, also known as metamodels, are compact, scalable analytic models that approximate the multivariate input/output behaviour of complex systems, based on only a limited set of computationally expensive simulations. In such way, surrogate models actually mimic the complex behaviour of the simulation model, and are applied in design automation, parametric studies, design space exploration, optimization and sensitivity analysis. Other synonyms for surrogate models are response surface models (RSM), emulators, auxiliary models, repro-models, metamodels, etc. [397].

Systems biology. Systems biology is the field devoted to the computational and mathematical modelling of complex biological systems [398, 399, 400]. It focuses on the relationships between the components of a biological system, and how these relationships give rise to its global function and behaviour. This is opposed to a reductionist paradigm.

Systems bioinformatics. A new approach to the analysis of biomedical data that is based on the application of a systems biology perspective. This includes, on one hand, a top-down view, with bioinformatics methods being used to extract and analyse information from “omics” data generated through high-throughput techniques [401], eventually integrating omics data coming from different sources [402, 403, 404]. On the other hand, this is complemented with a bottom-up approach, where information from molecular cells and tissues, alongside mathematical models, are used to elucidate the function and dynamic behaviour of cells, organs and organisms.

Systems dynamics. Systems dynamics or dynamical systems is a mathematical method or modelling approach for understanding the behaviour of complex systems with their states evolving over time. This is used in *in silico* modelling of biomedical systems. For instance, biochemical reactions (using mass action law), intracellular signalling pathways, activity of excitable/nerve cells and their networks, biological rhythms, cancer development, and population dynamics can be described by dynamical systems [405, 406, 407, 408, 409].

A system often consists of a set of interacting elements or components that forms a larger component or entity. Understanding the latter's behaviour is often not immediately clear just based on the elements or building blocks, but through the analysis of the interactions leading to "emergent" dynamical behaviour. The analysis could be performed analytically (especially for simpler systems) or computationally using various numerical methods. Often, the stability of the system is also evaluated analytically or computationally either locally e.g. around some steady state, or globally. Software are often used for numerical computation. The popular ones include XPPAUT (C programming based) [410] and MATCONT (MATLAB programming based) [411].

The elements or interactions can be linear or nonlinear. The interactions can be instantaneous or time-delayed. The system can be deterministic or stochastic (i.e. in the presence of noise). Supposed a system's state variable is described by a vector x, and the environment of system is described by parameters a, the evolution mechanism of dynamical systems can be continuous (behaving continuously over time) and described by a group of differential equations,

$$\frac{dx}{dt} = f(x, a, t),$$

or discrete (behaving over discrete time points) and described by difference equations,

$$x(t+1) = f[x(t), a],$$

or described by symbolic dynamics i.e. mathematical function mappings [409]

$$f{:}x(t){\rightarrow}x(t+1).$$

Often but not necessary, nonlinearity in the system can lead to highly non-trivial emergent dynamics. For instance, varying some parameter around its critical value can dramatically change the behaviour of the system. This is termed bifurcation [412] or phase transition, and is linked to Catastrophe Theory [413]. Some other topics related to systems dynamics or dynamical systems theory include Chaos Theory [409].

Systems Engineering. Systems Engineering is a multi/transdisciplinary field devoted to the engineering and engineering management of very large and complex socio-technical systems. It addresses all the elements within a system, their individual properties and inter-relations are considered and integrated in a holistic approach, through a combination of relationships to jointly perform a useful function as a whole. Systems Engineering combines Engineering with Management, Finance, Economics, Pure/Exact and Social Sciences, in a way to adequately design, develop, and implement the large and complex systems that are so important nowadays. It is typically used to manage the inherent complexity of societal problems, e.g., either in spacecraft design or in combination with pharmacokinetic/pharmacodynamic (PK/PD) modelling and Systems Biology [347, 348]. In this way, the Systems Engineering approaches are delimited within the Systems Theory framework [414].

Systems medicine. Systems medicine is an interdisciplinary field of study that looks at the human body as a system, composed of interacting parts, and further integrated into an environment. It considers that these complex relationships exist

on multiple levels, and that they have to be understood in light of a patient's genomics, behaviour and environment. As such, it integrates contributions from multiple research fields, including medicine, systems biology, statistics, modelling and simulation, and data science. The earliest uses of the term systems medicine appeared in 1992, in two articles independently published by B. J. Zeng [3] and T. Kamada [4].

As the name suggests, systems medicine represents the convergence of two main fields:

- *Systems biology*, the field of study that focuses on complex interactions within biological systems, using a holistic approach.
- *Medicine*, as it presents a clear focus towards medical research and medical practice. As such, systems medicine aims at having tangible benefits for the patients, with the identification of those elements that are critical for influencing the course of the system (i.e. medical conditions).

Among its objectives, it is worth highlighting:

- Systems medicine is not systems biology just in one species, but similar to the distinction between "medicine" and "biology" systems medicine needs to have to objective to achieve patient benefit, by either better or earlier diagnosis and therapy.
- Systems medicine questions and replaces the current concept of medicine, which is largely built on organ-based subfields and symptom-based disease definitions, towards a holistic-defining diseases at a mechanistic level.
- Systems medicine defines (diagnostic and therapeutic) targets not any longer as single molecules but rather perturbed networks, which form subgraphs of the interactome.
- At the application side, systems medicine will lead to precision diagnostics and therapeutics.

- Some therapeutics/drugs will not need to be developed *de novo* but repurposed/repositioned.
- Use multilayer diagnostic tools.
- Thereby systems medicine will enable predictive, personalized, preventative, participatory medicine.
- By increasing medical precision and efficacy, systems medicine ideally addresses the financial pressures on all health care providers and enables the ultimate move from an input medicine to an output medicine (see recent World Economic Forum Davos).

System of Systems. Systems of Systems can be represented as large scale, complex, distributed systems. System of Systems concept is described in terms of "Maier's criteria" [415]: operational and managerial independence, distribution, emergent behaviour as a result of component behaviour and evolutionary development. System of Systems principles can be applied in integrating health management, medical diagnosis and medical support systems [416].

Standards. The word "standard" has several different definitions. Whereas in general metrology, a standard is a reference that is used to calibrate measurements, in the systems biology field, standards have been developed through standardization initiatives (e.g. ISO, COMBINE [417]) to format and describe data and models, for exchange and understanding between scientific communities. Three types of standards have been considered [418]: standard formats for representing data and models; standard metadata for describing types of data and models; controlled vocabularies and ontologies to provide a common vocabulary.

Structural covariance networks. A technique used to reconstruct complex networks representations of brain cortical regions. The network is defined such that nodes represent brain regions, and links the Pearson's correlation of cortical thickness or volume between pairs of regions, as yielded by magnetic resonance data (MRI) [419, 420]. Structural covariance between regions can be used to construct the so-called structural covariance networks. Several studies have been conducted in which structural covariance networks have been analysed in healthy subjects [421, 422], and in groups of patients with disorders such as autism, attention deficit hyperactivity disorder, schizophrenia, or Alzheimer's disease [423, 424, 425, 426], or to assess the differences between gifted children and controls [427]. Since the SCN is at the group level, (structural) connectivity parameters are also at the group level and a permutation test will be needed to infer differences between measures. See also morphometric similarity networks.

Time-evolving networks. One major problem that was found while studying time-evolving systems through complex networks was that edges may not continuously be active. To illustrate, let us consider the network of contacts between inpatients of an hospital, which may be used to model the propagation of infectious diseases. Firstly, two people may be connected by a link even if they have been in the same room for a short time window, thus the probability of contagious should not be binarized. Secondly, the sequence of contacts is also important: if a person met patient *A* and later patient *B*, a disease cannot spread from *B* to *A*. The solution was the development of the concept of time-evolving, or temporal, networks, in which a collection of networks represent the status of the system as it evolves through time [428, 429].

Time scale separation. Dynamic mathematical models can be simplified using time scale separation approach: if part of a system operates sufficiently fast compared to the rest of the system, it may be assumed to have reached a steady-

state [430]. This allows the elimination of fastest components from the model, lumping them with slower components as they determine the speed of systems reaction. This approach can be very efficient in multiscale modelling where dynamics of very different processes are merged. Time scale separation is applied for modelling of vector-borne diseases taking where human host epidemiology is much slower than the transmission of vector from human to human by mosquitos: only human time scale is investigated assuming that human-human transmission happens instantly [431]. Time scale separation can be used to simplify modelling of biochemical processes at cellular physiology level [432].

Variation partitioning. Also called “commonality analysis”, a technique aimed at quantifying the part of the observed variation that is the shared consequence of two (or more) explanatory variables. It was initially introduced in 1992 by D. P. Borcard and co-authors in ecology [433], and has since seen some limited applications in medicine [434, 435].

Virtual physiological human. See physiome.

Acknowledgments

This article is based upon work from COST Action OpenMultiMed (CA15120), supported by COST (European Cooperation in Science and Technology).

J. L. M. thanks Instituto Politécnico de Portalegre, CERENA-Centro de Recursos Naturais e Ambiente, and the support of FCT- Fundação para a Ciência e a Tecnologia under the strategic project UID/ECI/04028/2020.

J.A.S. and J.B. obtained financial support from the Austrian Science Fund FWF (projects SFB-F54 and TCS-46).

Participation of V.S. in the project was supported by Czech Ministry of Education, Youth and Sports (LTC18074).

J. S.-C. thanks the support of the UVic-UCC (grant R0947) and the Ministry of Economic Affairs and Competitiveness of Spain (grant TEC2016-77791-C4-2-R).

D.R. thanks the support of Slovenian Research Agency (P1-0390 and MRIC-ELIXIR).

Part of the research is supported by the SILICOFCM project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777204. This article reflects only the author's view. The Commission is not responsible for any use that may be made of the information it contains.

References

[1] A. Castiglioni, A History of Medicine, London: Routledge, 2019.

[2] B. J. L. Berry and H. Kim, "Long waves 1790-1990: intermittency, chaos, and control," in *Chaos theory in the social sciences: Foundations and Applications*, University of Michigan Press, 1996, pp. 215-236.

[3] B. Z. Zeng, "On the holographic model of human body," in *1st National Conference of Comparative Studies Traditional Chinese Medicine and West Medicine (Medicine and Philosophy)*, 1992.

[4] T. Kamada, “System biomedicine: a new paradigm in biomedical engineering,” *Frontiers of medical and biological engineering: the international journal of the Japan Society of Medical Electronics and Biological Engineering,* vol. 4, no. 1, p. 1, 1992.

[5] M. Zanin, I. Chorbev, B. Stres, E. Stalidzans, J. Vera, P. Tieri, F. Castiglione, D. Groen, H. Zheng, J. Baumbach and J. Schmid, “Community effort endorsing multiscale modelling, multiscale data science and multiscale computing for systems medicine,” *Briefings in bioinformatics,* vol. 20, no. 3, pp. 1057-1062, 2019.

[6] E. Bonaneau, “Agent-based modeling: Methods and techniques for simulating human systems,” *Proceedings of the National Academy of Sciences USA,* vol. 99, no. 3, pp. 7280-7287, 2002.

[7] B. Punam and S. Chawla, “Agent Based Information Retrieval System Using Information Scent,” *Journal of Artificial Intelligence,* vol. 3, no. 4, 2010.

[8] S. Ugurlu and N. Erdogan, “An Agent-Based Information Retrieval System,” in *Proceedings of the First International Conference on Advances in Information Systems (ADVIS '00)*, London, 2000.

[9] J. L. Posadas, J. L. Poza, J. E. Simo, G. Benet and F. Blanes, “Agent-based distributed architecture for mobile robot control,” *Engineering Applications of Artificial Intelligence,* vol. 21, no. 6, pp. 805-823, 2008.

[10] S. Pennisi, F. Pappalardo and S. Motta, “Agent Based Modeling of Lung Metastasis-Immune System Competition,” *Lecture Notes in Computer Science,* vol. 5666, pp. 1-3, 2009.

[11] C. M. Glen, M. L. Kemp and E. O. Voit, “Agent-based modeling of morphogenetic systems: Advantages and challenges,” *PLOS Computational Biology,* vol. 15, no. 3, p. e1006577, 2019.

[12] F. Castiglione and F. Celada, Immune System Modeling and Simulation, Boca Raton: CRC Press, 2015.

[13] J. von Neumann, The theory of self-reproducing automata, Urbana, IL: University of Illinois Press, 1966.

[14] S. Ulam, “Random processes and transformations,” in *Proceedings of the International Congress of Mathematics*, 1952.

[15] C. Langton, Artificial Life: An Overview, MIT Press, 1995.

[16] L. Zhang, Z. Wang, J. A. Sagotsky and T. S. Deisboeck, “Multiscale agent-based cancer modeling,” *J Math Biol,* vol. 58, no. 4-5, pp. 545-559, 2009.

[17] C. Gong, O. Milberg, B. Wang, P. Vicini, R. Narwal, L. Roskos and A. S. Popel, “A computational multiscale agent-based model for simulating spatio-temporal tumour immune response to PD1 and PDL1 inhibition,” *J. R. Soc. Interface,* vol. 14, p. 20170320, 2017.

[18] J. M. Epstein, “Modelling to contain pandemics,” *Nature,* vol. 460, no. 7259, p. 687, 2009.

[19] L. Perez and S. Dragicevic, “An agent-based approach for modeling dynamics of contagious disease spread,” *International journal of health geographics,* vol. 8, no. 1, p. 50, 2009.

[20] G. An, Q. Mi, J. Dutta-Moscato and Y. Vodovotz, "Agent-based models in translational systems biology," *Wiley Interdisciplinary Reviews: Systems Biology and Medicine,* vol. 1, no. 2, pp. 159-171, 2009.

[21] M. van Gerven and S. Bohte, Artificial neural networks as models of neural information processing, 2018.

[22] I. Goodfellow, Y. Bengio and A. Courville, Deep learning, MIT Press, 2016.

[23] J. Awwalu, A. G. Garba, A. Ghazvini and R. Atuah, "Artificial Intelligence in Personalized Medicine - Application of AI Algorithms in Solving Personalized Medicine Problems," *International Journal of Computer Theory and Engineering,* vol. 7, no. 6, pp. 439-443, 2015.

[24] T. Ching, X. Zhu and L. X. Garmire, "Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data," *PLoS Computational Biology,* vol. 14, no. 4, p. e1006076, 2018.

[25] I. Bica, P. Velickovic, H. Xiao and P. Li, "Multi-omics data integration using cross-modal neural networks," in *Proceedings of the 26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2018)*, 2018.

[26] Y. Donner, S. Kazmierczak and K. Fortney, "Drug repurposing using deep embeddings of gene expression profiles," *Molecular pharmaceutics,* vol. 15, no. 10, pp. 4314-4325, 2018.

[27] M. J. Keeling and P. Rohani, Modeling infectious diseases in humans and animals, Princeton University Press, 2011.

[28] S. Särkkä, Bayesian filtering and smoothing, Cambridge University Press, 2013.

[29] P. Lucas, "Bayesian networks in medicine: a model-based approach to medical decision making," 2001.

[30] F. V. Jensen, An introduction to Bayesian networks, London: UCL Press, 1996.

[31] S. Andreassen, C. Riekehr, B. Kristensen, H. C. Schønheyder and L. Leibovici, "Using probabilistic and decision–theoretic methods in treatment and prognosis modeling," *Artificial Intelligence in Medicine,* vol. 15, no. 2, pp. 121-134, 1999.

[32] F. L. Seixas, B. Zadrozny, J. Laks, A. Conci and D. C. M. Saade, "A Bayesian s ′network decision model for supporting the diagnosis of dementia, Alzheimer disease and mild cognitive impairment," *Computers in biology and medicine,* vol. 51, pp. 140-158, 2014.

[33] M. L. P. Bueno, A. Hommersom, P. J. Lucas, M. Lappenschaar and J. G. Janzing, "Understanding disease processes by partitioned dynamic Bayesian networks," *Journal of biomedical informatics,* vol. 61, pp. 283-297, 2016.

[34] S. E. de Rooij, A. Abu-Hanna, M. Levi and E. de Jonge, "Identification of high-risk subgroups in very elderly intensive care unit patients," *Critical Care,* vol. 11, no. 2, p. R33, 2007.

[35] A. M. Kalet, J. H. Gennari, E. C. Ford and M. H. Phillips, "Bayesian network models for error detection in radiotherapy plans," *Physics in Medicine & Biology,* vol. 60, no. 7, p. 2735, 2015.

[36] L. Xing, M. Guo, X. Liu, C. Wang, L. Wang and Y. Zhang, "An improved Bayesian network method for reconstructing gene regulatory network based on candidate auto selection," *BMC genomics,* vol. 18, no. 9, p. 844, 2017.

[37] T. Nielsen and F. Jensen, Bayesian networks and decision graphs, Springer Science & Business Media, 2009.

[38] P. Dagum, A. Galper and E. Horvitz, "Dynamic network models for forecasting," in *Uncertainty in artificial intelligence*, Morgan Kaufmann, 1992, pp. 41-48.

[39] N. Friedman and M. Goldszmidt, "Discretizing continuous attributes while learning Bayesian networks," in *13th International Conference on Machine Learning*, 1996.

[40] H. Li, L. Lu, K. Manly, E. Chesler, L. Bao, J. Wang, M. Zhou, R. Williams and Y. Cui, "Inferring gene transcriptional modulatory relations: a genetical genomics approach," *Human molecular genetics,* vol. 14, no. 9, pp. 1119-1125, 2005.

[41] J. Zhu, P. Lum, J. Lamb, D. GuhaThakurta, S. Edwards, R. Thieringer, J. Berger, M. Wu, J. Thompson, A. Sachs and E. Schadt, "An integrative genomics approach to the reconstruction of gene networks in segregating populations," *Cytogenetic and genome research,* vol. 105, no. 2-4, pp. 363-374, 2004.

[42] M. J. Bayarri and J. O. Berger, "The interplay of Bayesian and frequentist analysis," *Statistical Science,* pp. 58-80, 2004.

[43] E.-J. Wagenmakers, M. Lee, T. Lodewyckx and G. J. Iverson, "Bayesian versus frequentist inference," in *Bayesian evaluation of informative hypotheses*, Springer, 2008, pp. 181-207.

[44] G. Casella and R. L. Berger, "Reconciling Bayesian and frequentist evidence in the one-sided testing problem," *Journal of the American Statistical Association,* vol. 82, no. 397, pp. 106-111, 1987.

[45] T. Sustersic, V. Rankovic, M. Peulic and A. S. Peulic, "An Early Disc Herniation Identification System for Advancement in the Standard Medical Screening Procedure based on Bayes Theorem," *IEEE journal of biomedical and health informatics,* vol. 24, no. 1, pp. 151-159, 2019.

[46] D. Ashby, "Bayesian statistics in medicine: a 25 year review," *Statistics in medicin,* vol. 25, no. 21, pp. 3589-3631, 2006.

[47] M. Zanin, S. Belkoura, J. Gomez, C. Alfaro and J. Cano, "Topological structures are consistently overestimated in functional complex networks," *Scientific reports,* vol. 8, no. 1, p. 11980, 2018.

[48] L. C. Gurrin, J. J. Kurinczuk and P. R. Burton, "Bayesian statistics in medical research: an intuitive alternative to conventional data analysis," *Journal of Evaluation in Clinical Practice,* vol. 6, no. 2, pp. 193-204, 2000.

[49] L. S. Freedman and D. J. Spiegelhalter, "The assessment of the subjective opinion and its use in relation to stopping rules for clinical trials," *Journal of the Royal Statistical Society: Series D,* vol. 32, no. 1-2, pp. 153-160, 1983.

[50] D. J. Spiegelhalter and L. S. Freedman, "A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion," *Statistics in medicine,* vol. 5, no. 1, pp. 1-13, 1986.

[51] K. Chaloner, T. Church, T. A. Louis and J. P. Matts, "Graphical elicitation of a prior distribution for a clinical trial," *Journal of the Royal Statistical Society: Series D,* vol. 42, no. 4, pp. 341-353, 1993.

[52] J. B. Kadane, J. M. Dickey, R. L. Winkler, W. S. Smith and S. C. Peters, "Interactive elicitation of opinion for a normal linear model," *Journal of the American Statistical Association,* vol. 75, no. 372, pp. 845-854, 1980.

[53] B. Efron, "Bayesians, frequentists, and scientists," *Journal of the American Statistical Association,* vol. 100, no. 469, pp. 1-5, 2005.

[54] H. Jeffreys, The theory of probability, OUP Oxford, 1998.

[55] J. M. Bernardo, "Reference posterior distributions for Bayesian inference," *Journal of the Royal Statistical Society: Series B (Methodological),* vol. 41, no. 2, pp. 113-128, 1979.

[56] J. O. Berger and J. M. Bernardo, "On the development of the reference prior method," *Bayesian statistics,* vol. 4, pp. 35-60, 1992.

[57] E. T. Jaynes, Probability theory: The logic of science, 2003: Cambridge university press.

[58] G. S. Datta and R. Mukerjee, Probability matching priors: higher order asymptotics, Springer Science & Business Media, 2012.

[59] J. O. Berger, W. Strawderman and D. Tang, "Posterior propriety and admissibility of hyperpriors in normal hierarchical models," *The Annals of Statistics,* vol. 33, no. 2, pp. 606-646, 2005.

[60] R. E. Kass and L. Wasserman, "The selection of prior distributions by formal rules," *Journal of the American Statistical Association,* vol. 91, no. 435, pp. 1343-1370, 1996.

[61] P. C. Lambert, A. J. Sutton, P. R. Burton, K. R. Abrams and D. R. Jones, "How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS," *Statistics in medicine,* vol. 24, no. 15, pp. 2401-2428, 2005.

[62] J. Herson, "Bayesian analysis of cancer clinical trials: An introduction to four papers," *Statistics in medicine,* vol. 22, no. 1, pp. 1-3, 1992.

[63] S. Wieand and S. Cha, "Description of the statistical aspects of a study for advanced colorectal cancer patients," *Statistics in medicine,* vol. 11, no. 1, pp. 5-11, 1992.

[64] L. S. Freedman and D. J. Spiegelhalter, "Application of Bayesian statistics to decision making during a clinical trial," *Statistics in medicine,* vol. 11, no. 1, pp. 23-35, 1992.

[65] J. B. Greenhouse, "On some applications of Bayesian methods in cancer clinical trials," *Statistics in medicine,* vol. 11, no. 1, pp. 37-53, 1992.

[66] D. O. Dixon and R. Simon, "Bayesian subset analysis in a colorectal cancer clinical trial," *Statistics in medicine,* vol. 11, no. 1, pp. 13-22, 1992.

[67] D. A. Berry, "Bayesian clinical trials," *Nature reviews Drug discovery,* vol. 5, no. 1, p. 27, 2006.

[68] P. Armitage, G. Berry and J. N. S. Matthews, Statistical methods in medical research, John Wiley & Sons, 2008.

[69] R. L. Winkler, An introduction to Bayesian inference and decision, Holt, Rinehart and Winston, 1972.

[70] G. Casella, "An introduction to empirical Bayes data analysis," *The American Statistician,* vol. 39, no. 2, pp. 83-87, 1985.

[71] R. E. Kass and L. Wasserman, "A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion," *Journal of the american statistical association,* vol. 90, no. 431, pp. 928-934, 1995.

[72] J. O. Berger, “An overview of robust Bayesian analysis [with discussion],” *Test,* vol. 3, pp. 5-124, 1994.

[73] C. Robert, The Bayesian choice: from decision-theoretic foundations to computational implementation, Springer Science & Business Media, 2007.

[74] P. S. Ayyaswamy, “Introduction to biofluid mechanics.,” *In Fluid Mechanics,* pp. 779- 852, 2012.

[75] Y. C. Fung, Biomechanics: circulation., Springer Science & Business Media, 2013.

[76] C. S. Peskin, “ Flow patterns around heart valves: a numerical method. Journal of computational physics,” vol. 10, no. 2, pp. 252 -271, 1972.

[77] J. E. N. J. R. M. V. G. S. T. M. L. &. O. M. Nichols, “ Production and assessment of decellularized pig and human lung scaffolds.,” *Tissue Engineering Part A,* vol. 19, no. 17-18, pp. 2045-2062, 2013.

[78] D. Liepsch, “An introduction to biofluid mechanics—basic models and applications. Journal of biomechanics,” vol. 35, no. 4, pp. 415- 435, 2002.

[79] J. W. .. Valvano, Encyclopedia of Medical Devices and Instrumentation. Bioheat transfer., John Wiley & Sons, Inc., 2006.

[80] Z. S. &. L. J. Deng, “Analytical study on bioheat transfer problems with spatial or transient heating on skin surface or inside biological bodies.,” *Journal of Biomechanical Engineering,* vol. 124, no. 6, pp. 638-649, 2002.

[81] J. C. Chato, “Fundamentals of bioheat transfer.,” *Thermal dosimetry and treatment planning,* pp. 1-56, 1990.

[82] A.-L. Barabasi and Z. N. Oltvai, “Network biology: Understanding the cell's functional organization,” *Nature Reviews Genetics,* vol. 5, no. 2, pp. 101-113, 2004.

[83] P. Tieri, L. Farina, M. Petti, L. Astolfi, P. Paci and F. Castiglione, “Network Inference and Reconstruction in Bioinformatics,” in *Encyclopedia of Bioinformatics and Computational Biology*, Oxford Academic Press, 2019, pp. 805-813.

[84] A.-L. Barabasi, N. Gulbahce and J. Loscalzo, “Network medicine: A network-based approach to human disease,” *Nature Reviews Genetics,* vol. 12, no. 1, pp. 56-68, 2011.

[85] N. Przulj, “Protein-protein interactions: making sense of networks via graph-theoretic modeling,” *Bioessays,* vol. 33, no. 2, pp. 115-123, 2011.

[86] H. Jeong, S. P. Mason, A. L. Barabási and Z. N. Oltvai, “Lethality and centrality in protein networks,” *Nature,* vol. 411, no. 6833, p. 41, 2001.

[87] V. S. Rao, K. Srinivas, G. N. Sujini and G. N. Kumar, “Protein-protein interaction detection: methods and analysis,” *Int J Proteomics,* vol. 2014, p. 147648, 2014.

[88] T. M. Cafarelli, A. Desbuleux, Y. Wang, S. G. Choi, D. De Ridder and M. Vidal, “Mapping, modeling, and characterization of protein–protein interactions on a proteomic scale,” *Current Opinion in Structural Biology,* vol. 44, pp. 201-210, 2017.

[89] V. Huynh-Thu and G. Sanguinetti, “Gene Regulatory Network Inference: An Introductory Survey,” *Gene Regulatory Networks,* pp. 1-23, 2018.

[90] A. J. Butte, P. Tamayo, D. Slonim, T. R. Golub and I. S. Kohane, “Discovering functional relationships between RNA expression and chemotherapeutic

susceptibility using relevance networks," *Proc Natl Acad Sci USA,* vol. 97, no. 22, pp. 12182-12186, 2000.

[91] S. Aibar, C. B. Gonzalez-Blas, T. Moerman, V. A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J. C. Marine, P. Geurts, J. Aerts, J. van den Oord, Z. K. Atak, J. Wouters and S. Aerts, "SCENIC: single-cell regulatory network inference and clustering," *Nat Methods,* vol. 14, no. 11, pp. 1083-1086, 2017.

[92] T. E. Chan, M. P. H. Stumpf and A. C. Babtie, "Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures," *Cell Syst,* vol. 5, no. 3, pp. 251-267, 2017.

[93] L. E. Chai, S. K. Loh, S. T. Low, M. S. Mohamad, S. Deris and Z. Zakaria, "A review on the computational approaches for gene regulatory network construction," *Comput Biol Med,* vol. 48, pp. 55-65, 2014.

[94] F. Emmert-Streib, M. Dehmer and B. Haibe-Kains, "Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks," *Front Cell Dev Biol,* vol. 2, no. 38, 2014.

[95] E. A. Serin, H. Nijveen, H. W. Hilhorst and W. Ligterink, "Learning from Co-expression Networks: Possibilities and Challenges," *Front Plant Sci,* vol. 7, p. 444, 2016.

[96] P. Tieri, A. Termanini, E. Bellavista, S. Salvioli, M. Capri and C. Franceschi, "Charting the NF-κB pathway interactome map," *PLoS ONE,* vol. 7, no. 3, p. e32678, 2012.

[97] E. J. Molinelli, A. Korkut, W. Wang, M. L. Miller, N. P. Gauthier, X. Jing, P. Kaushik, Q. He, G. Mills, D. B. Solit, C. A. Pratilas, M. Weigt, A. Braunstein, A. Pagnani, R. Zecchina and C. Sander, "Perturbation biology: inferring signaling networks in cellular systems," *PLoS Comput Biol,* vol. 9, no. 12, p. e1003290, 2013.

[98] J. A. Papin, T. Hunter, B. O. Palsson and S. Subramaniam, "Reconstruction of cellular signalling networks and analysis of their properties," *Nat Rev Mol Cell Biol,* vol. 6, no. 2, pp. 99-111, 2005.

[99] E. Pitkanen, J. Rousu and E. Ukkonen, "Computational methods for metabolic reconstruction," *Curr Opin Biotechnol,* vol. 21, no. 1, pp. 70-77, 2015.

[100] Z. Nikoloski, R. Perez-Storey and L. J. Sweetlove, "Inference and Prediction of Metabolic Network Fluxes," *Plant Physiol,* vol. 169, no. 3, pp. 1443-1455, 2015.

[101] M. R. Brent, "Past roadblocks and new opportunities in transcription factor network mapping," *Trends in Genetics,* vol. 32, no. 11, pp. 736-750, 2016.

[102] T. Maniatis, S. Goodbourn and J. A. Fischer, "Regulation of inducible and tissue-specific gene expression," *Science,* vol. 236, no. 4806, pp. 1237-1245, 1987.

[103] H. Niwa, "The principles that govern transcription factor network functions in stem cells," *Development,* vol. 145, no. 6, p. dev157420, 2018.

[104] N. Bonzanni, A. Garg, K. A. Feenstra, J. Schütte, S. Kinston, D. Miranda-Saavedra, J. Heringa, I. Xenarios and B. Göttgens, "Hard-wired heterogeneity in blood stem cells revealed using a dynamic regulatory network model," *Bioinformatics,* vol. 29, no. 13, pp. i80-i88, 2013.

[105] V. Moignard, S. Woodhouse, L. Haghverdi, A. J. Lilly, Y. Tanaka, A. C. Wilkinson, F. Buettner, I. C. Macaulay, W. Jawaid, E. Diamanti and S. I. Nishikawa, "Decoding the regulatory network of early blood development from single-cell

gene expression measurements," *Nature biotechnology,* vol. 33, no. 3, p. 269, 2015.

[106] Y. Kang, H. H. Liow, E. J. Maier and M. R. Brent, "NetProphet 2.0: Mapping Transcription Factor Networks by Exploiting Scalable Data Resources," *Bioinformatics,* vol. 34, no. 2, pp. 249-257, 2017.

[107] G. K. Smyth, "Linear models and empirical bayes methods for assessing differential expression in microarray experiments," *Stat. Appl. Genet. Mol. Biol.,* vol. 1, p. 3, 2004.

[108] V. Matys, O. V. Kel-Margoulis, E. Fricke, I. Liebich, S. Land, A. Barre-Dirrie, I. Reuter, D. Chekmenev, M. Krull, K. Hornischer and N. Voss, "TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes," *Nucleic acids research,* vol. 34, no. 1, pp. D108-D110, 2006.

[109] Ž. Urlep, G. Lorbek, M. Perše, J. Jeruc, P. Juvan, M. Matz-Soja, R. Gebhardt, I. Björkhem, J. A. Hall, R. Bonneau and D. R. Littman, "Disrupting Hepatocyte Cyp51 from Cholesterol Synthesis Leads to Progressive Liver Injury in the Developing Mouse and Decreases RORC Signalling," *Scientific reports,* vol. 7, p. 40775, 2017.

[110] B. D. Ratner, A. S. Hoffman, F. J. Schoen and J. E. Lemons, Biomaterials science: an introduction to materials in medicine, Elsevier, 2004.

[111] J. Park and S. L. Roderic, Biomaterials: an introduction, Springer Science & Business Media, 2007.

[112] J. D. Bronzino, Biomedical engineering handbook, CRC Press, 1999.

[113] J. D. Bronzino, J. Y. Wong and D. R. Peterson, Biomaterials: principles and Practices, CRC Press, 2012.

[114] J. &. D. S. L. Humphrey, Introduction to Biomechanics, New York: Springer-Verlag, 2016.

[115] V. M. &. Z. V. M. Zatsiorsky, Kinetics of human motion., Human Kinetics, 2002.

[116] D. A. Winter, Biomechanics and motor control of human gait: normal, elderly and pathological., 1991.

[117] Y. C. Fung, Biomechanics: mechanical properties of living tissues., Springer Science & Business Media, 2013.

[118] C. J. De Luca, "The use of surface electromyography in biomechanics.," *Journal of applied biomechanics,* vol. 13, no. 2, pp. 135-163, 1997.

[119] N. H. &. M. J. P. Reynolds, "Single cell active force generation under dynamic loading–Part II: Active modelling insights.," *Acta biomaterialia,* vol. 27, pp. 251-263, 2015.

[120] B. Chopard and M. Droz, Cellular automata, Springer, 1998.

[121] V. Vezhnevets and V. Konouchine, "GrowCut: Interactive multi-label ND image segmentation by cellular automata," in *Graphicon*, 2005.

[122] S. Wongthanavasu and V. Tangvoraphonkchai, "Cellular Automata-based Algorithm and its Application in Medical image processing," in *2007 IEEE International Conference on Image Processing*, IEEE, 2007, pp. 1-41.

[123] A. Prieto-Langarica, H. Kojouharov, B. Chen-Charpentier and L. Tang, "A cellular automata model of infection control on medical implants," *Applications and applied mathematics: an international journal,* vol. 6, no. 1, p. 1, 2011.

[124] R. M. Z. dos Santos and S. Coutinho, "Dynamics of HIV infection: A cellular automata approach," *Physical review letters,* vol. 87, no. 16, p. 168102, 2001.

[125] A. R. Mikler, S. Venkatachalam and K. Abbas, "Modeling infectious diseases using global stochastic cellular automata," *Journal of Biological Systems,* vol. 13, no. 4, pp. 421-439, 2005.

[126] M. A. Banning, "A review of clinical decision making: Models and current research," *J. Clin. Nurs.,* vol. 17, pp. 187-195, 2008.

[127] M. G. M. Hunink, C. M. Weinstein, E. Wittenberg, M. F. Drummond, J. S. Pliskin, J. B. Wong and P. P. Glasziu, Decision making in health and medicine: Integrating evidence and values, Cambridge University Press, 2014.

[128] G. Kong, D. L. Xu and J. B. Yang, "Clinical decision support systems: a review on knowledge representation and inference under uncertainties," *International Journal of Computational Intelligence Systems,* vol. 1, no. 2, pp. 159-167, 2008.

[129] I. Sim, P. Gorman, R. A. Greenes, R. B. Haynes, B. Kaplan, H. Lehmann and P. C. Tang, "Clinical decision support systems for the practice of evidence-based medicine," *J. Am. Med. Inform. Assoc.,* vol. 8, no. 6, pp. 527-534, 2008.

[130] H. A. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, A. Kalloo, A. B. H. Hassen, L. Thomas, A. Enk and L. Uhlmann, "Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Annals of Oncology,* vol. 29, no. 8, pp. 1836-1842, 2018.

[131] S. Belkoura, M. Zanin and A. LaTorre, "Fostering interpretability of data mining models through data perturbation," *Expert Systems with Applications,* vol. 137, pp. 191-201, 2019.

[132] S. Khairat, D. Marc, W. Crosby and A. A. S. Sanousi, "Reasons for physicians not adopting clinical decision support systems: Critical analysis," *JMIR Med. Inform.,* vol. 6, no. 2, p. e24, 2018.

[133] L. Kaufman and P. J. Rousseeuw, Finding groups in data: an introduction to cluster analysis, John Wiley & Sons, 2009.

[134] P. Berkhin, "A survey of clustering data mining techniques," in *Grouping multidimensional data*, Springer, 2006, pp. 25-71.

[135] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics),* vol. 28, no. 1, pp. 100-108, 1979.

[136] D. Steinley, "K-means clustering: a half-century synthesis," *British Journal of Mathematical and Statistical Psychology,* vol. 59, no. 1, pp. 1-34, 2006.

[137] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern recognition letters,* vol. 31, no. 8, pp. 651-666, 2010.

[138] M. Ester, H.-P. Kriegel, J. Sander and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," *KDD,* vol. 96, no. 34, pp. 226-231, 1996.

[139] G. J. McLachlan and E. B. Kaye, Mixture models: Inference and applications to clustering, New York: M. Dekker, 1988.

[140] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological),* vol. 39, no. 1, pp. 1-22, 1977.

[141] R. Albert and A.-L. Barabasi, "Statistical mechanics of complex networks," *Reviews of modern physics,* vol. 74, no. 1, p. 47, 2002.

[142] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez and D.-U. Hwang, "Complex networks: Structure and dynamics," *Physics reports,* vol. 424, no. 4-5, pp. 175-308, 2006.

[143] S. H. Strogatz, "Exploring complex networks," *Nature,* vol. 410, no. 6825, p. 268, 2001.

[144] G. Kossinets and D. J. Watts, "Empirical analysis of an evolving social network," *Science,* vol. 311, no. 5757, pp. 88-90, 2006.

[145] G. Bonanno, G. Caldarelli, F. Lillo and R. N. Mantegna, "Topology of correlation-based minimal spanning trees in real and model markets," *Physical Review E,* vol. 68, no. 4, p. 046130, 2003.

[146] A. H. Y. Tong, G. Lesage, G. D. Bader, H. Ding, H. Xu, X. Xin, J. Young, G. F. Berriz, R. L. Brost and M. Chang, "Global mapping of the yeast genetic interaction network," *Science,* vol. 303, no. 5659, pp. 808-813, 2004.

[147] L. D. F. Costa, F. A. Rodrigues, G. Travieso and P. R. Villas Boas, "Characterization of complex networks: A survey of measurements," *Advances in physics,* vol. 56, no. 1, pp. 167-242, 2007.

[148] M. Zanin, D. Papo, P. A. Sousa, E. Menasalvas, A. Nicchi, E. Kubik and S. Boccaletti, "Combining complex networks and data mining: why and how," *Physics Reports,* vol. 635, pp. 1-44, 2016.

[149] P. W. Anderson, "More is different," *Science,* vol. 177, no. 4047, pp. 393-396, 1972.

[150] H. A. Simon, The organization of complex systems, Springer, 1977.

[151] Y. Bar-Yam, Dynamics of complex systems, CRC Press, 2019.

[152] D. T. Kaplan, M. I. Furman, S. M. Pincus, S. M. Ryan, L. A. Lipsitz and A. L. Goldberger, "Aging and the complexity of cardiovascular dynamics," *Biophysical journal,* vol. 59, no. 4, pp. 945-949, 1991.

[153] A. L. Goldberger and B. J. West, "Fractals in physiology and medicine," *The Yale journal of biology and medicine,* vol. 60, no. 5, p. 421, 1987.

[154] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology,* vol. 278, no. 6, pp. 2039-2049, 2000.

[155] M. Zanin, L. Zunino, O. A. Rosso and D. Papo, "Permutation entropy and its main biomedical and econophysics applications: a review," *Entropy,* vol. 14, no. 8, pp. 1553-1577, 2012.

[156] C. K. Peng, S. Havlin, H. E. Stanley and A. L. Goldberger, "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series.," *Chaos: An Interdisciplinary Journal of Nonlinear Science,* vol. 5, no. 1, pp. 82-87, 1995.

[157] S. A. Kauffman, "Antichaos and adaptation," *Scientific American,* vol. 265, no. 2, pp. 78-85, 1991.

[158] J. Langedijk, A. K. Mantel-Teeuwisse, D. S. Slijkerman and M. H. D. B. Schutjens, "Drug repositioning and repurposing: terminology and definitions in literature," *Drug Discovery Today,* vol. 20, no. 8, p. 1027–1034, 2015.

[159] J. Meslamani, J. Li, J. Sutter, A. Stevens, H. O. Bertrand and D. Rognan, "Protein–ligand-based pharmacophores: generation and utility assessment in computational ligand profiling," *Journal of chemical information and modeling,* vol. 52, no. 4, pp. 943-955, 2012.

[160] A. Aliper, S. Plis, A. Artemov, A. Ulloa, P. Mamoshina and A. Zhavoronkov, "Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data," *Molecular pharmaceutics,* vol. 13, no. 7, pp. 2524-2530, 2016.

[161] Y. A. Lussier and J. L. Chen, "The emergence of genome-based drug repositioning," *Science translational medicine,* vol. 3, no. 96, 2011.

[162] C. Andronis, A. Sharma, V. Virvilis, S. Deftereos and A. Persidis, "Literature mining, ontologies and information visualization for drug repurposing," *Briefings in bioinformatics,* vol. 12, no. 4, pp. 357-368, 2011.

[163] K. Savva, M. Zachariou, A. Oulas, G. Minadakis, K. Sokratous, N. Dietis and G. M. Spyrou, "Computational Drug Repurposing for Neurodegenerative Diseases," in *In Silico Drug Desig*, Academic Press, 2019.

[164] H. Cruse, "Constraints for joint angle control of the human arm," *Biological Cybernetics,* vol. 54, no. 2, pp. 125-132, 1986.

[165] E. Stalidzans, A. Seiman, K. Peebo, V. Komasilovs and A. Pentjuss, "Model-based metabolism design: constraints for kinetic and stoichiometric models," *Biochemical Society Transactions,* vol. 46, no. 2, pp. 261-267, 2018.

[166] A. K. Dey, "Understanding and using context," *Personal and ubiquitous computing,* vol. 5, no. 1, pp. 4-7, 2001.

[167] A. Padovitz, S. W. Loke, A. Zaslavsky, B. Burg and C. Bartolini, "An approach to data fusion for context awareness," in *International and Interdisciplinary Conference on Modeling and Using Context* , Springer, 2005, pp. 353-367.

[168] J. H. Jahnke, Y. Bychkov, D. Dahlem and L. Kawasme, "Context-aware information services for health care," in *Proceedings of the KI-04 Workshop on Modeling and Retrieval of Context*, 2004.

[169] R. Wirth and J. Hipp, "CRISP-DM: Towards a standard process model for data mining," in *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, 2000, pp. 29-39.

[170] C. Shearer, "The CRISP-DM model: the new blueprint for data mining," *Journal of data warehousing,* vol. 5, no. 4, pp. 13-22, 2000.

[171] D. Pyle, Data preparation for data mining, Morgan Kaufmann, 1999.

[172] R. Cooley, B. Mobasher and J. Srivastava, "Data preparation for mining world wide web browsing patterns," *Knowledge and information systems,* vol. 1, no. 1, pp. 5-32, 1999.

[173] S. Zhang, C. Zhang and Q. Yang, "Data preparation for data mining," *Applied artificial intelligence,* vol. 17, no. 5-6, pp. 375-381, 2003.

[174] M. W. Browne, "Cross-validation methods," *Journal of mathematical psychology,* vol. 44, no. 1, pp. 108-32, 2000.

[175] M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kötter, T. Meinl, P. Ohl, K. Thiel and B. Wiswedel, "KNIME-the Konstanz information miner: version 2.0 and beyond," *AcM SIGKDD explorations Newsletter,* vol. 11, no. 1, pp. 26-31, 2009.

[176] T. Wendler and S. Gröttrup, Data mining with SPSS modeler: theory, exercises and solutions, Springer, 2016.

[177] M. Hofmann and R. Klinkenberg, RapidMiner: Data mining use cases and business analytics applications, CRC Press, 2013.

[178] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving and M. Isard, "Tensorflow: a system for large-scale machine learning," in *OSDI*, 2016, pp. 265-283.

[179] H. Chen, R. H. L. Chiang and V. C. Storey, "Business intelligence and analytics: from big data to big impact," *MIS quarterly,* pp. 1165-1188, 2012.

[180] S. Huang, K. Chaudhary and L. Garmire, "More Is Better: Recent Progress in Multi-Omics Data Integration Methods," *Frontiers in genetics,* vol. 8, p. 84, 2017.

[181] L. Breiman, Classification and regression trees, Routledge, 2017.

[182] M. Mehta, R. Agrawal and J. Rissanen, "SLIQ: A fast scalable classifier for data mining," in *International Conference on Extending Database Technology*, 1996, pp. 18-32.

[183] J. Shafer, R. Agrawal and M. Mehta, "SPRINT: A scalable parallel classi er for data mining," in *Proc. 1996 Int. Conf. Very Large Data Bases*, 1996, pp. 544-555.

[184] J. R. Quinlan, "Induction of decision trees," *Machine learning,* vol. 1, no. 1, pp. 81-106, 1986.

[185] J. R. Quinlan, C4.5: programs for machine learning, Elsevier, 2014.

[186] V. Podgorelec, P. Kokol, B. Stiglic and I. Rozman, "Decision trees: an overview and their use in medicine," *Journal of medical systems,* vol. 26, no. 5, pp. 445-463, 2002.

[187] A. T. Azar and S. M. El-Metwally, "Decision tree classifiers for automated medical diagnosis," *Neural Computing and Applications,* vol. 23, no. 7-8, pp. 2387-2403, 2013.

[188] E. Turban, Decision support and expert systems: management support systems, Prentice Hall, 1993.

[189] A. Barbosa-Póvoa, A. C. Subias and De Miranda, J. L., Optimization and decision support systems for supply chains, Zurich: Springer, 2017.

[190] K. Gurney, An introduction to neural networks, CRC Press, 2014.

[191] Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," *Nature,* vol. 521, no. 7553, p. 436, 2015.

[192] D. Lupton, "Critical perspectives on digital health technologies," *Sociology compass,* vol. 8, no. 12, pp. 1344-1359, 2014.

[193] D. Lupton, "The digitally engaged patient: Self-monitoring and self-care in the digital health era," *Social Theory & Health,* vol. 11, no. 3, pp. 256-270, 2013.

[194] F. Birnbaum, D. M. Lewis, R. Rosen and M. L. Ranney, "Patient engagement and the design of digital health," *Academic emergency medicine: official journal of the Society for Academic Emergency Medicine,* vol. 22, no. 6, p. 754, 2015.

[195] F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang and F. Sui, "Digital twin-driven product design, manufacturing and service with big data," *The International Journal of Advanced Manufacturing Technology,* vol. 94, no. 9-12, pp. 3563-3576, 2018.

[196] K. Bruynseels, F. Santoni de Sio and J. van den Hoven, "Digital twins in health care: ethical implications of an emerging engineering paradigm," *Frontiers in genetics,* vol. 9, p. 31, 2018.

[197] J. Sabater-Mir, "Towards a Healthcare Digital Twin," in *Artificial Intelligence Research and Development: Proceedings of the 22nd International Conference of the Catalan Association for Artificial Intelligence*, IOS Press, 2019, p. 312.

[198] E. Moeendarbary, T. Y. Ng and M. Zangeneh, "Dissipative particle dynamics: introduction, methodology and complex fluid applications - a review," *International Journal of Applied Mechanics,* vol. 1, no. 4, pp. 737-763, 2009.

[199] N. Filipovic, M. Zivic, M. Obradovic, T. Djukic, Z. Markovic and M. Rosic, "Numerical and experimental LDL transport through arterial wall," *Microfluidics and nanofluidics,* vol. 16, no. 3, pp. 455-464, 2014.

[200] R. Vulović, M. Nikolić and N. Filipović, "Smart platform for the analysis of cupula deformation caused by otoconia presence within SCCs," *Computer methods in biomechanics and biomedical engineering,* vol. 22, no. 2, pp. 130-138, 2019.

[201] N. Filipovic, A. Jovanovic, D. Petrovic, M. Obradovic, S. Jovanovic, D. Balos and M. Kojic, "Modelling of self-healing materials using discrete and continuum methods," *Surface Coatings International,* vol. 95, no. 2, pp. 74-79, 2012.

[202] S. Wandelt, X. Sun, E. Menasalvas, A. Rodríguez-González and M. Zanin, "On the use of random graphs as null model of large connected networks," *Chaos, Solitons & Fractals,* vol. 119, pp. 318-325, 2019.

[203] P. Erdős and A. Rényi, "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci,* vol. 5, no. 1, pp. 17-60, 1960.

[204] C. P. Wild, "Complementing the genome with an "exposome": The outstanding challenge of environmental exposure measurement in molecular epidemiology," *Cancer epidemiology, biomarkers & prevention,* vol. 14, no. 8, p. 1847, 2005.

[205] G. M. B. Louis, M. M. Smarr and C. J. Patel, "The Exposome Research Paradigm: an Opportunity to Understand the Environmental Basis for Human Health and Disease," *Current environmental health reports,* vol. 4, no. 1, pp. 89-98, 2017.

[206] G. M. B. Louis, E. Yeung, K. Kannan, J. Maisog, C. Zhang, K. L. Grantz and R. Sundaram, "Patterns and Variability of Endocrine-disrupting Chemicals During Pregnancy: Implications for Understanding the Exposome of Normal Pregnancy," *Epidemiology,* vol. 30, pp. S65-S75, 2019.

[207] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J. W. Boiten, L. B. da Silva Santos, P. E. Bourne and J. Bouwman, "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific data,* vol. 3, 2016.

[208] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial intelligence,* vol. 97, no. 1-2, pp. 245-271, 1997.

[209] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of machine learning research,* vol. 3, pp. 1157-1182, 2003.

[210] J. Reunanen, "Overfitting in making comparisons between variable selection methods," *Journal of Machine Learning Research,* vol. 3, pp. 1371-1382, 2003.

[211] T. N. Lal, O. Chapelle, J. Weston and A. Elisseeff, "Embedded methods," in *Feature extraction*, Springer, 2006, pp. 137-165.

[212] D. L. Logan, A first course in the finite element method, Cengage Learning, 2011.

[213] J. N. Reddy, Introduction to the Finite Element Method, McGraw-Hill, 2017.

[214] A. Vulović and N. Filipović, "Computational Analysis of Hip Implant Surfaces," *Journal of the Serbian Society for Computational Mechanics,* vol. 13, no. 1, pp. 109-119, 2019.

[215] M. Nikolic, P. D. Teal, V. Isailovic and N. Filipović, "Finite element cochlea box model – Mechanical and electrical analysis of the cochlea," *AIP Conference Proceedings,* vol. 1703, no. 1, p. 070012, 2015.

[216] F. Auricchio, M. Conti, A. Ferrara, S. Morganti and A. Reali, "Patient-specific finite element analysis of carotid artery stenting: a focus on vessel modeling," *International Journal for Numerical Methods in Biomedical Engineering,* vol. 29, no. 6, pp. 645-664, 2013.

[217] S. Djorovic, I. Saveljic and N. Filipovic, "Computational Simulation of Carotid Artery: From Patient-Specific Images to Finite Element Analysis," *Journal of the Serbian Society for Computational Mechanics,* vol. 13, no. 1, pp. 120-129, 2019.

[218] A. Redaelli, F. Boschetti and F. Inzoli, "The assignment of velocity profiles in finite element simulations of pulsatile flow in arteries," *Computers in biology and medicine,* vol. 27, no. 3, pp. 233-247, 1997.

[219] M. Robnik-Šikonja, M. Radović, S. Đorović, B. Anđelković-Ćirković and N. Filipović, "Modeling ischemia with finite elements and automated machine learning," *Journal of computational science,* vol. 29, pp. 99-106, 2018.

[220] T. Šušteršič, L. Liverani, A. R. Boccaccini, S. Savić, A. Janićijević and N. Filipović, "Numerical simulation of electrospinning process in commercial and in-house software PAK," *Materials Research Express,* vol. 6, no. 2, p. 025305, 2018.

[221] R. Eymard, T. Gallouët and R. Herbin, "Finite volume methods," in *Handbook of numerical analysis*, Elsevier, 2000, pp. 713-1018.

[222] A. Vulović, T. Šušteršič, S. Cvijić, S. Ibrić and N. Filipović, "Coupled in silico platform: Computational fluid dynamics (CFD) and physiologically-based pharmacokinetic (PBPK) modelling," *European Journal of Pharmaceutical Sciences,* vol. 113, pp. 171-184, 2018.

[223] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature reviews neuroscience,* vol. 10, no. 3, p. 186, 2009.

[224] K. J. Friston, "Functional and effective connectivity: a review," *Brain connectivity,* vol. 1, no. 1, pp. 13-36, 2011.

[225] J. H. Hung, T. H. Yang, Z. Hu, Z. Weng and C. DeLisi, "Gene set enrichment analysis: performance evaluation and usage guidelines," *Briefings in bioinformatics,* vol. 13, no. 3, pp. 281-291, 2011.

[226] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander and J. P. Mesirov, "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles," *Proceedings of the National Academy of Sciences,* vol. 102, no. 43, pp. 15545-15550, 2005.

[227] A. Liberzon, A. Subramanian, R. Pinchback and H. Thorvaldsdóttir, "Molecular signatures database (MSigDB) 3.0," *Bioinformatics,* vol. 27, no. 12, pp. 1739-1740, 2011.

[228] C. Granger, "Investigating causal relations by econometric models and cross-spectral methods.," *Econometrica: Journal of the Econometric Society,* pp. 424-438, 1969.

[229] N. Wiener, "The theory of prediction," in *Modern Mathematics for Engineers*, New York, McGraw-Hill, 1956, pp. 165-190.

[230] S. L. Bressler and A. K. Seth, "Wiener–Granger causality: a well established methodology," *Neuroimage,* vol. 58, no. 2, pp. 323-329, 2011.

[231] M. Zanin and D. Papo, "Detecting switching and intermittent causalities in time series," *Chaos: An Interdisciplinary Journal of Nonlinear Science,* vol. 27, no. 4, p. 047403, 2017.

[232] L. Schiatti, G. Nollo, G. Rossato and L. Faes, "Extended Granger causality: a new tool to identify the structure of physiological networks," *Physiological measurement,* vol. 36, no. 4, p. 827, 2015.

[233] E. Bose, M. Hravnak and S. M. Sereika, "Vector autoregressive (VAR) models and granger causality in time series analysis in nursing research: dynamic changes among vital signs prior to cardiorespiratory instability events as an example," *Nursing research,* vol. 66, no. 1, p. 12, 2017.

[234] E. Erdil and I. H. Yetkiner, "The Granger-causality between health care expenditure and output: a panel data approach," *Applied Economics,* vol. 41, no. 4, pp. 511-518, 2009.

[235] P. Goyal and E. Ferrara, "Graph embedding techniques, applications, and performance: A survey," *Knowledge-Based Systems,* vol. 151, pp. 78-94, 2018.

[236] H. Cai, V. Zheng and K. Chang, "A Comprehensive Survey of Graph Embedding: Problems, Techniques, and Applications," *IEEE Transactions on Knowledge and Data Engineering,* vol. 30, no. 9, pp. 1616-1637, 2018.

[237] C. Seshadhri, A. Sharma, A. Stolman and A. Goel, "The impossibility of low-rank representations for triangle-rich complex networks," *Proceedings of the National Academy of Sciences,* vol. 117, no. 11, pp. 5631-5637, 2020.

[238] M. Pellegrini, D. Haynor and J. M. Johnson, "Protein interaction networks," *Expert review of proteomics,* vol. 1, no. 2, pp. 239-249, 2004.

[239] E. T. Bullmore and D. S. Bassett, "Brain graphs: graphical models of the human brain connectome," *Annual review of clinical psychology,* vol. 7, pp. 113-140, 2011.

[240] M. J. Keeling and K. T. Eames, "Networks and epidemic models," *Journal of the Royal Society Interface,* vol. 2, no. 4, pp. 295-307, 2005.

[241] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai and A. L. Barabási, "The large-scale organization of metabolic networks," *Nature,* vol. 407, no. 6804, p. 651, 2000.

[242] A. D. Perkins and M. A. Langston, "Threshold selection in gene co-expression networks using spectral graph theory techniques," *BMC bioinformatics,* vol. 10, no. 11, p. S4, 2009.

[243] X. Yue, Z. Wang, J. Huang, S. Parthasarathy, S. Moosavinasab, Y. Huang, S. M. Lin, W. Zhang, P. Zhang and H. Sun, "Graph Embedding on Biomedical Networks: Methods, Applications, and Evaluations," *arXiv,* p. arXiv:1906.05017, 2019.

[244] R. Ietswaart, B. M. Gyori, J. A. Bachman, P. K. Sorger and L. S. Churchman , "GeneWalk identifies relevant gene functions for a biological context using network representation learning," *bioRxiv,* p. 755579, 2019.

[245] G. Rosenthal, F. Váša, A. Griffa, P. Hagmann, E. Amico, J. Goñi, G. Avidan and O. Sporns, "Mapping higher-order relations between brain structure and function with embedded vector representations of connectomes," *Nature communications,* vol. 9, no. 1, p. 2178, 2018.

[246] A. Quattoni, S. Wang, L.-P. Morency, M. Collins and T. Darrell, "Hidden conditional random fields," *IEEE Transactions on Pattern Analysis \& Machine Intelligence,* vol. 10, pp. 1848-1852, 2007.

[247] J. F. D. Saa and M. Cetin, "Hidden conditional random fields for classification of imaginary motor tasks from eeg data," in *2011 19th European Signal Processing Conference*, IEEE, 2011, pp. 171-175.

[248] Y. Liu, J. Carbonell, P. Weigele and V. Gopalakrishnan, "Protein fold recognition using segmentation conditional random fields (SCRFs)," *Journal of Computational Biology,* vol. 13, no. 2, pp. 394-406, 2006.

[249] I. R. White, P. Royston and A. M. Wood, "Multiple imputation using chained equations: issues and guidance for practice," *Statistics in medicine,* vol. 30, no. 4, pp. 377-399, 2011.

[250] A. L. McCutcheon, Latent class analysis, Sage, 1987.

[251] M. P. Gleeson, S. Modi, A. Bender, R. L. Marchese Robinson, J. Kirchmair, M. Promkatkaew, S. Hannongbua and R. C. Glen, "The challenges involved in modeling toxicity data in silico: a review," *Current pharmaceutical design,* vol. 18, no. 9, pp. 1266-1291, 2012.

[252] L. B. Edelman, J. A. Eddy and N. D. Price, "In silico models of cancer," *Wiley Interdisciplinary Reviews: Systems Biology and Medicine,* vol. 2, no. 4, pp. 438-459, 2010.

[253] T. Martonen, J. Fleming, J. Schroeter, J. Conway and D. Hwang, "In silico modeling of asthma," *Advanced drug delivery reviews,* vol. 55, no. 7, pp. 829-849, 2003.

[254] Y. Vodovotz and T. R. Billiar, " In Silico Modeling: Methods and Applications to Trauma and Sepsis," *Critical care medicine,* vol. 41, no. 8, p. 2008, 2013.

[255] R. B. Colquitt, D. A. Colquhoun and R. H. Thiele, "In silico modelling of physiologic systems," *Best practice & research Clinical anaesthesiology,* vol. 25, no. 4, pp. 499-510, 2011.

[256] J. S. Hamid, P. Hu, N. M. Roslin, V. Ling, C. M. Greenwood and J. Beyene, "Data integration in genetics and genomics: methods and challenges," *Human genomics and proteomics,* vol. 2009, p. 869093, 2009.

[257] L. J. Lu, Y. Xia, A. Paccanaro, H. Yu and M. Gerstein, "Assessing the limits of genomic data integration for predicting protein networks," *Genome research,* vol. 15, no. 7, pp. 945-953, 2005.

[258] K. Van Steen and N. Malats, "Perspectives on Data Integration in Human Complex Disease Analysis," in *Big Data Analytics in Bioinformatics and Healthcare*, IGI Global, 2015, pp. 284-322.

[259] S. N. Thorsen and M. E. Oxley, "Fusion or Integration: What's the difference?," in *Fusion 2004: Seventh International Conference on Information Fusion*, 2004.

[260] O. Vermesan, P. Friess, P. Guillemin, H. Sundmaeker, M. Eisenhauer, K. Moessner, F. Le Gall and P. Cousin, "Internet of things strategic research and

innovation agenda," in *Internet of things: converging technologies for smart environments and integrated ecosystems*, River Publishers, 2013, pp. 7-152.

[261] D. Repta, M. A. Moisescu, I. S. Sacala, I. Dumitrache and A. M. Stanescu, "Towards the development of semantically enabled flexible process monitoring systems," *International Journal of Computer Integrated Manufacturing,* vol. 30, no. 1, pp. 96-108, 2017.

[262] H. Zheng, J. T. Wassan, M. A. Moisescu, L. Stoicu-Tivadar, J. Miranda, M. Crisan-Vida, I. S. Sacala, A. Badnjevic, I. Chorbev and B. Jakimovski, "Multiscale Computing in Systems Medicine: a Brief Reflection," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, 2018, pp. 2190-2195.

[263] G. Manogaran, N. Chilamkurti and C. H. Hsu, "Emerging trends, issues, and challenges in Internet of Medical Things and wireless networks," *Personal and Ubiquitous Computing,* vol. 22, no. 5-6, pp. 879-882, 2018.

[264] S. R. Islam, D. Kwak, M. H. Kabir, M. Hossain and K. S. Kwak, "The internet of things for health care: a comprehensive survey," *IEEE Access,* vol. 3, pp. 678-708, 2015.

[265] Y. I. N. Yuehong, Y. Zeng, X. Chen and Y. Fan, "The internet of things in healthcare: An overview," *Journal of Industrial Information Integration,* vol. 1, pp. 3-13, 2016.

[266] D. A. Wolf-Gladrow, Lattice-gas cellular automata and lattice Boltzmann models: an introduction, Berlin: Springer, 2000.

[267] S. Succi, The Lattice Boltzmann Equation for uid dynamics and beyond, Oxford University Press, 2001.

[268] M. Sukop and D. T. Thorne, Lattice Boltzmann Modeling, Heidelberg: Springer, 2006.

[269] A. A. Mohamad, Lattice Boltzmann Method: Fundamentals and Engineering Applications with Computer Codes, London: Springer, 2011.

[270] K. Timm, H. Kusumaatmaja, A. Kuzmin, O. Shardt, G. Silva and E. Viggen, The lattice Boltzmann method: principles and practice, Springer, 2017.

[271] Y. T. Feng, K. Han and D. R. J. Owen, "Coupled lattice Boltzmann method and discrete element modelling of particle transport in turbulent fluid flows: Computational issues," *International Journal for Numerical Methods in Engineering,* vol. 72, no. 9, pp. 1111-1134, 2007.

[272] O. Malaspinas, N. Fietier and M. Deville, "Lattice Boltzmann method for the simulation of viscoelastic fluid flows," *Journal of Non-Newtonian Fluid Mechanics,* vol. 165, no. 23-24, pp. 1637-1653, 2010.

[273] T. Djukic and N. Filipovic, "Numerical modeling of the cupular displacement and motion of otoconia particles in a semicircular canal," *Biomechanics and modeling in mechanobiology,* vol. 16, no. 5, pp. 1669-1680, 2017.

[274] T. Djukic, I. Saveljic and N. Filipovic, "Numerical modeling of the motion of otoconia particles in the patient-specific semicircular canal," *Computational Particle Mechanics,* vol. 6, no. 4, pp. 767-780, 2019.

[275] N. Filipovic, M. Zivic, M. Obradovic, T. Djukic, Z. Markovic and M. Rosic, "Numerical and experimental LDL transport through arterial wall," *Microfluidics and nanofluidics,* vol. 16, no. 3, pp. 455-464, 2014.

[276] C. M. Bishop, Pattern recognition and machine learning, Springer Science & Business Media, 2006.

[277] E. Alpaydin, Introduction to machine learning, MIT Press, 2009.

[278] L. Richiardi, R. Bellocco and D. Zugna, "Mediation analysis in epidemiology: methods, interpretation and bias," *International journal of epidemiology,* vol. 42, no. 5, pp. 1511-1519, 2013.

[279] K. J. Preacher, "Advances in mediation analysis: a survey and synthesis of new developments," *Annual review of psychology,* vol. 66, pp. 825-852, 2015.

[280] M. A. Musen and J. H. van Bemmel, Handbook of medical informatics, 1997.

[281] W. R. Hersh, "Medical informatics: improving health care through information," *JAMA,* vol. 288, no. 16, pp. 1955-1958, 2002.

[282] H. U. Prokosch and T. Ganslandt, "Perspectives for medical informatics," *Methods of information in medicine,* vol. 48, no. 1, pp. 38-44, 2009.

[283] R. Haux, "Medical informatics: past, present, future," *International journal of medical informatics,* vol. 79, no. 9, pp. 599-610, 2010.

[284] A. Benis, R. Barak Barkan and T. Sela, "Communication Behavior Changes Between Patients With Diabetes And Healthcare Providers Over 9 Years," *J Med Internet Res,* p. Epub ahead of print, 2020.

[285] K. W. Goodman and R. A. Miller, "Ethics and health informatics: Users, standards, and outcomes," in *Medical Informatics*, Springer, 2001, pp. 257-281.

[286] J. Chong, M. Yamamoto and J. Xia, "MetaboAnalystR 2.0: From Raw Spectra to Biological Insights," *Metabolites,* vol. 9, no. 3, p. 57, 2019.

[287] J. K. Nicholson, J. C. Lindon and E. Holmes, "'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data.," *Xenobiotica,* vol. 29, no. 11, pp. 1181-1189, 1999.

[288] O. Fiehn, J. Kopka, P. Dörmann, T. Altmann, R. N. Trethewey and L. Willmitzer, "Metabolite profiling for plant functional genomics," *Nature biotechnology,* vol. 18, no. 11, pp. 1157-1161, 2000.

[289] R. Ramautar, R. Berger, J. van der Greef and T. Hankemeier, "Human metabolomics: strategies to understand biology," *Current opinion in chemical biology,* vol. 17, no. 5, pp. 841-846, 2013.

[290] J. C. Lindon, E. Holmes and J. K. Nicholson, "Metabonomics techniques and applications to pharmaceutical research & development," *Pharmaceutical research,* vol. 23, no. 6, pp. 1075-1088, 2006.

[291] A. Dhariwal, J. Chong, S. Habib, I. King, L. B. Agellon and J. Xia, "MicrobiomeAnalyst - a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data," *Nucleic Acids Research,* vol. 45, pp. 180-188, 2017.

[292] D. Bertsimas and M. Sim, "The Price of Robustness," *Operations Research,* vol. 52, no. 1, pp. 35-53, 2004.

[293] R. Schultz, L. Stougie and M. H. Van Der Vlerk, "Two-stage stochastic integer programming: a survey," *Statistica Neerlandica,* vol. 50, no. 3, pp. 404-416, 1996.

[294] D. Hendrycks, K. Lee and M. Mazeika, "Using Pre-Training Can Improve Model Robustness and Uncertainty," *arXiv preprint,* p. arXiv:1901.09960, 2019.

[295] D. Tsipras, S. Santurkar, L. Engstrom, A. Turner and A. Madry, "Robustness May Be at Odds with Accuracy," *arXiv preprint,* p. arXiv:1805.12152, 2018.

[296] C. Cobelli, E. R. Carson, L. Finkelstein and M. S. Leaning, "Validation of simple and complex models in physiology and medicine," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology,* vol. 246, no. 2, pp. 259-266, 1984.

[297] G. Antonelli, A. Padoan, A. Aita, L. Sciacovelli and M. Plebani, "Verification or validation, that is the question," *Journal of Laboratory and Precision Medicine,* vol. 2, no. 8, 2017.

[298] J. Seidlitz, F. Váša, M. Shinn, R. Romero-Garcia, K. J. Whitaker, P. E. Vértes, K. Wagstyl, P. K. Reardon, L. Clasen, S. Liu and A. Messinger, "Morphometric similarity networks detect microscale cortical organization and predict inter-individual cognitive variation," *Neuron,* vol. 97, no. 1, pp. 231-247, 2018.

[299] S. E. Morgan, J. Seidlitz, K. J. Whitaker, R. Romero-Garcia, N. E. Clifton, C. Scarpazza, T. van Amelsvoort, M. Marcelis, J. van Os, G. Donohoe and D. Mothersill, "Cortical patterning of abnormal morphometric similarity in psychosis is associated with brain expression of schizophrenia-related genes.," *Proceedings of the National Academy of Sciences,* vol. 116, no. 19, pp. 9604-9609, 2019.

[300] G. E. Doucet, D. A. Moser, A. Rodrigue, D. S. Bassett, D. C. Glahn and S. Frangou, "Person-Based Brain Morphometric Similarity is Heritable and Correlates With Biological Features," *Cerebral Cortex,* vol. 29, no. 2, pp. 852-862, 2018.

[301] J. G. Michopoulos, C. Farhat and J. Fish, "Modeling and simulation of multiphysics systems," *Journal of Computing and Information Science in Engineering,* vol. 5, no. 3, pp. 198-213, 2005.

[302] D. E. Keyes, L. C. McInnes, C. Woodward, W. Gropp, E. Myra, M. Pernice, J. Bell, J. Brown, A. Clo and J. Connors, "Multiphysics simulations: Challenges and opportunities," *The International Journal of High Performance Computing Applications,* vol. 27, no. 1, pp. 4-83, 2013.

[303] J. G. White, E. Southgate, J. N. Thomson and S. Brenner, "The structure of the nervous system of the nematode Caenorhabditis elegans," *Philos Trans R Soc Lond B Biol Sci,* vol. 314, no. 1165, pp. 1-340, 1986.

[304] S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gomez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang and M. Zanin, "The structure and dynamics of multilayer networks," *Physics Reports,* vol. 544, no. 1, pp. 1-122, 2014.

[305] J. A. McCammon, B. R. Gelin and M. Karplus, "Dynamics of folded proteins," *Nature,* vol. 267, pp. 585-590, 1977.

[306] V. Rokhlin, "Rapid solution of integral equations of classical potential theory," *Journal of Computational Physics,* vol. 60, no. 2, pp. 187-207, 1985.

[307] S. Karabasov, D. Nerukh, A. Hoekstra, B. Chopard and P. V. Coveney, Multiscale modelling: approaches and challenges, The Royal Society, 2014.

[308] B. Stres and L. Kronegger, "Shift in the paradigm towards next-generation microbiology," *FEMS Microbiology Letters,* vol. 366, no. 15, 2019.

[309] E. Weinan, Principles of multiscale modeling, Cambridge University Press, 2011.

[310] A. Hagberg, P. Swart and D. Chult, “Exploring network structure, dynamics, and function using NetworkX,” 2008.

[311] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, “Cytoscape: a software environment for integrated models of biomolecular interaction networks,” *Genome research,* vol. 13, no. 11, pp. 2498-2504, 2003.

[312] M. E. Smoot, K. Ono, J. Ruscheinski, P.-L. Wang and T. Ideker, “Cytoscape 2.8: new features for data integration and network visualization,” *Bioinformatics,* vol. 27, no. 3, pp. 431-432, 2010.

[313] M. Bastian, S. Heymann and M. Jacomy, “Gephi: an open source software for exploring and manipulating networks,” *Icwsm,* vol. 8, pp. 361-362, 2009.

[314] V. Batagelj and A. Mrvar, “Pajek-program for large network analysis,” *Connections,* vol. 21, no. 2, pp. 47-57, 1998.

[315] Z. Hu, J. Mellor, J. Wu and C. DeLisi, “VisANT: an online visualization and analysis tool for biological interaction data,” *BMC bioinformatics,* vol. 5, no. 1, p. 17, 2004.

[316] G. Zhou, O. Soufan, J. Ewald, R. E. Hancock, N. Basu and J. Xia, “NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis.,” *Nucleic acids research,* vol. 47, no. 1, pp. 234-241, 2019.

[317] J. Xia, M. J. Benner and R. E. Hancock, “NetworkAnalyst - integrative approaches for protein-protein interaction network analysis and visual exploration,” *Nucleic Acids Research,* vol. 42, no. 1, pp. 167-174, 2014.

[318] A.-L. Barabasi, “Network Medicine - From Obesity to the "Diseasome",” *The New England Journal of Medicine,* vol. 357, no. 4, pp. 404-407, 2007.

[319] S. Maslov and K. Sneppen, “Specificity and stability in topology of protein networks,” *Science,* vol. 296, no. 5569, pp. 910-913, 2002.

[320] W. Zhu, M. Baust, Y. Cheng, S. Ourselin, M. J. Cardoso and A. Feng, “Privacy-Preserving Federated Brain Tumour Segmentation,” in *Machine Learning in Medical Imaging: 10th International Workshop, MLMI 2019*, Springer Nature, 2019, p. 133.

[321] G. Kovalenko, A. L. Ducluzeau, L. Ishchenko, M. Sushko, M. Sapachova, N. Rudova, O. Solodiankin, A. Gerilovych, R. Dagdag, M. Redlinger and M. Bezymennyi, “Complete Genome Sequence of a Virulent African Swine Fever Virus from a Domestic Pig in Ukraine,” *Microbiology Resource Announcements,* vol. 8, no. 42, pp. e00883-19, 2019.

[322] T. Cvitanović, M. C. Reichert, M. Moškon, M. Mraz, F. Lammert and D. Rozman, “Large-scale computational models of liver metabolism: How far from the clinics?,” *Hepatology,* vol. 66, no. 4, pp. 1323-1334, 2017.

[323] E. Larsdotter Nilsson and P. Fritzson, “BioChem - A Biological and Chemical Library for Modelica,” in *3rd International Modelica Conference*, Linköping, 2003.

[324] A. Belič, J. Ačimovič, A. Naik and M. Goličnik, “Analysis of the Steady-State Relations and Control-Algorithm Characterisation in a Mathematical Model of Cholesterol Biosynthesis,” *Simulation Modelling Practice and Theory,* vol. 33, pp. 18-27, 2013.

[325] A. Naik, D. Rozman and A. Belič, "SteatoNet: the first integrated human metabolic model with multi-layered regulation to investigate liver-associated pathologies," *PLoS computational biology,* vol. 10, no. 12, p. e1003993, 2014.

[326] T. Cvitanović Tomaš, Ž. Urlep, M. Moškon, M. Mraz and D. Rozman, "LiverSex Computational Model: Sexual Aspects in Hepatic Metabolism and Abnormalities," *Frontiers in physiology,* vol. 9, p. 360, 2018.

[327] D. L. Rubin, N. H. Shah and N. F. Noy, "Biomedical ontologies: a functional perspective," *Briefings in bioinformatics,* vol. 9, no. 1, pp. 75-90, 2007.

[328] A. Groß, C. Pruski and E. Rahm, "Evolution of biomedical ontologies and mappings: overview of recent approaches," *Computational and structural biotechnology journal,* vol. 14, pp. 333-340, 2016.

[329] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. J. Goldberg, K. Eilbeck, A. Ireland, C. J. Mungall and N. Leontis, "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration," *Nature biotechnology,* vol. 25, no. 11, p. 1251, 2007.

[330] G. Clermont, C. Auffray, Y. Moreau, D. Rocke, D. Dalevi, D. Dubhashi, D. Marshall, P. Raasch, F. Dehne, P. Provero and J. Tegner, "Bridging the gap between systems biology and medicine," *Genome medicine,* vol. 1, no. 9, p. 88, 2009.

[331] N. R. Council, Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease, National Academies Press, 2011.

[332] S. Köhler, N. Vasilevsky, M. Engelstad, E. Foster, J. McMurry, S. Aymé, G. Baynam, S. Bello, C. Boerkoel, K. Boycott and M. Brudno, "The Human Phenotype Ontology in 2017," *Nucleic acids research,* vol. 45, no. D1, pp. D865-D876, 2017.

[333] M. Haendel, C. Chute and P. Robinson, "Classification, Ontology, and Precision Medicine," *New England Journal of Medicine,* vol. 379, no. 15, pp. 1452-1462, 2018.

[334] M. Ashyraliyev, Y. Fomekong-Nanfack, J. A. Kaandorp and J. G. Blom, "Systems biology: parameter estimation for biochemical models," *The FEBS journal,* vol. 276, no. 4, pp. 886-902, 2009.

[335] E. Stalidzans, K. Landmane, J. Sulins and S. Sahle, "Misinterpretation risks of global stochastic optimisation of kinetic models revealed by multiple optimisation runs," *Mathematical biosciences,* vol. 307, pp. 25-32, 2019.

[336] O.-T. Chis, J. R. Banga and E. Balsa-Canto, "Structural identifiability of systems biology models: a critical comparison of methods," *PloS one,* vol. 6, no. 11, p. e27755, 2011.

[337] Z. Zi, "Sensitivity analysis approaches applied to systems biology models," *IET systems biology,* vol. 5, no. 6, pp. 336-346, 2011.

[338] A. Kiparissides, S. S. Kucherenko, A. Mantalaris and E. N. Pistikopoulos, "Global sensitivity analysis challenges in biological systems modeling," *Industrial & Engineering Chemistry Research,* vol. 48, no. 15, pp. 7168-7180, 2009.

[339] K. H. Cho, S. Y. Shin, W. Kolch and O. Wolkenhauer, "Experimental design in systems biology, based on parameter sensitivity analysis using a monte carlo

method: A case study for the tnfα-mediated nf-κ b signal transduction pathway," *Simulation,* vol. 79, no. 12, pp. 726-739, 2003.

[340] T. A. Knijnenburg, L. F. Wessels, M. J. Reinders and I. Shmulevich, "Fewer permutations, more accurate P-values," *Bioinformatics,* vol. 25, no. 12, pp. 161-168, 2009.

[341] R.-R. Liu, W.-X. Wang, Y.-C. Lai and B.-H. Wang, "Cascading dynamics on random networks: Crossover in phase transition," *Physical Review E,* vol. 85, no. 2, p. 026110, 2012.

[342] S. W. Omholt and P. J. Hunter, "The Human Physiome: a necessary key for the creative destruction of medicine," *Interface Focus,* vol. 6, no. 2, 2016.

[343] G. S. Ginsburg and H. F. Willard, Genomic and Precision Medicine: Foundations, Translation, and Implementation, Academic Press, 2016.

[344] A. Katsnelson, Momentum grows to make 'personalized' medicine more 'precise', Nature Publishing Group, 2013.

[345] T. Bedford and R. Cooke, Probabilistic risk analysis: foundations and methods, Cambridge University Press, 2001.

[346] M. Danhof, "Systems pharmacology - Towards the modeling of network interactions," *Eur. J. Pharm. Sci.,* vol. 94, pp. 4-14, 2016.

[347] H. Geerts, A. Spiros, P. Roberts and R. Carr, "Quantitative systems pharmacology as an extension of PK/PD modeling in CNS research and development," *J. Pharmacokinetics and Pharmacodynamics,* vol. 40, no. 3, pp. 257-265, 2013.

[348] P. H. van der Graaf and N. Benson, "Systems pharmacology: bridging systems biology and pharmacokinetics-pharmacodynamics (PKPD) in drug discovery and development," *Pharm. Res.,* vol. 28, no. 7, pp. 1460-1464, 2011.

[349] T. Leil and R. Bertz, "Quantitative Systems Pharmacology can reduce attrition and improve productivity in pharmaceutical research and development," *Front. Pharmacol.,* vol. 5, no. 247, 2014.

[350] T. Hart and L. Xie, "Providing data science support for systems pharmacology and its implications to drug discovery," *Expert. Opin. Drug Discov.,* vol. 11, no. 3, pp. 241-256, 2016.

[351] M. C. Peterson and M. M. Riggs, "FDA advisory meeting clinical pharmacology review utilizes a quantitative systems pharmacology (QSP) model: A watershed moment," *CPT: Pharmacometrics Syst. Pharmacol.,* vol. 4, no. 3, pp. 189-192, 2015.

[352] L. Breiman, "Random forests," *Machine learning,* vol. 45, no. 1, pp. 5-32, 2001.

[353] T. M. Oshiro, P. S. Perez and J. A. Baranauskas, "How many trees in a random forest?," in *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, Springer, 2012, pp. 154-168.

[354] A. Verikas, A. Gelzinis and M. Bacauskiene, "Mining data with random forests: A survey and results of new tests," *Pattern recognition,* vol. 44, no. 2, pp. 330-349, 2011.

[355] B. Bollobás, "Random graphs," in *Modern graph theory*, Springer, 1998, pp. 215-252.

[356] S. Janson, T. Luczak and A. Rucinski, Random graphs, John Wiley & Sons, 2011.

[357] R. Albert, "Scale-free networks in cell biology," *Journal of cell science,* vol. 118, no. 21, pp. 4947-4957, 2005.

[358] E. Fox Keller, "Revisiting "scale-free" networks," *BioEssays,* vol. 27, no. 10, pp. 1060-1068, 2005.

[359] G. Caldarelli, Scale-free networks: complex webs in nature and technology, Oxford University Press, 2007.

[360] A.-L. Barabási, "Scale-free networks: a decade and beyond," *Science,* vol. 325, no. 5939, pp. 412-413, 2009.

[361] R. Khanin and E. Wit, "How scale-free are biological networks," *Journal of computational biology,* vol. 13, no. 3, pp. 810-818, 2006.

[362] P. Crucitti, V. Latora, M. Marchiori and A. Rapisarda, "Efficiency of scale-free networks: error and attack tolerance," *Physica A: Statistical Mechanics and its Applications,* vol. 320, pp. 622-642, 2003.

[363] L. K. Gallos, R. Cohen, P. Argyrakis, A. Bunde and S. Havlin, "Stability and topology of scale-free networks under attack and defense strategies," *Physical review letters,* vol. 94, no. 18, p. 188701, 2005.

[364] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science,* vol. 286, no. 5439, pp. 509-512, 1999.

[365] A.-L. Barabási, E. Ravasz and T. Vicsek, "Deterministic scale-free networks," *Physica A: Statistical Mechanics and its Applications,* vol. 299, no. 3-4, pp. 559-564, 2001.

[366] G. Caldarelli, A. Capocci, P. De Los Rios and M. A. Munoz, "Scale-free networks from varying vertex intrinsic fitness," *Physical review letters,* vol. 89, no. 25, p. 258702, 2002.

[367] J. Saramäki and K. Kaski, "Scale-free networks generated by random walkers," *Physica A: Statistical Mechanics and its Applications,* vol. 341, pp. 80-86, 2004.

[368] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, Numerical Recipes in C - 2nd Edition, Cambridge University Press, 1992.

[369] V. Granville, M. Krivánek and J.-P. Rasson, "Simulated annealing: A proof of convergence," *IEEE transactions on pattern analysis and machine intelligence,* vol. 16, no. 6, pp. 652-656, 1994.

[370] S. Webb, "Optimisation of conformal radiotherapy dose distribution by simulated annealing," *Physics in Medicine & Biology,* vol. 34, no. 10, p. 1349, 1989.

[371] E. Lessard and J. Pouliot, "Inverse planning anatomy-based dose optimization for HDR-brachytherapy of the prostate using fast simulated annealing algorithm and dedicated objective function," *Medical physics,* vol. 28, no. 5, pp. 773-779, 2001.

[372] M. Langer, S. Morrill, R. Brown, O. Lee and R. Lane, "A comparison of mixed integer programming and fast simulated annealing for optimizing beam weights in radiation therapy," *Medical Physics,* vol. 23, no. 6, pp. 957-964, 1996.

[373] N. Friedland and D. Adam, "Automatic ventricular cavity boundary detection from sequential ultrasound images using simulated annealing," *IEEE transactions on medical imaging,* vol. 8, no. 4, pp. 344-353, 1989.

[374] A. Alexandridis and E. Chondrodima, "A medical diagnostic tool based on radial basis function classifiers and evolutionary simulated annealing," *Journal of biomedical informatics,* vol. 49, pp. 61-72, 2014.

[375] D. J. Watts and S. H. Strogatz, "Collective Dynamics of Small-World Networks," *Nature,* vol. 393, no. 6684, p. 440, 1998.

[376] F. Karinthy, Chains. Everything is different., Budapest: Atheneum Press, 1929.

[377] G. R. Liu and M. B. Liu, Smoothed particle hydrodynamics: a meshfree particle method, World Scientific, 2003.

[378] J. J. Monaghan, "Smoothed particle hydrodynamics," *Annual review of astronomy and astrophysics,* vol. 30, no. 1, pp. 543-574, 1992.

[379] P. W. Cleary and M. Prakash, "Discrete–element modelling and smoothed particle hydrodynamics: potential in the environmental sciences," *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences,* vol. 362, no. 1822, pp. 2003-2030, 2004.

[380] Z. Zhang, H. Qiang and W. Gao, "Coupling of smoothed particle hydrodynamics and finite element method for impact dynamics simulation," *Engineering Structures,* vol. 33, no. 1, pp. 255-264, 2011.

[381] L. Lobovský and J. Křen, "Smoothed particle hydrodynamics modelling of fluids and solids," *Applied and Computational Mechanics,* vol. 1, pp. 512-530, 2007.

[382] A. Sofla, B. Cirkovic, A. Hsieh, J. W. Miklas, N. Filipovic and M. Radisic, "Enrichment of live unlabelled cardiomyocytes from heterogeneous cell populations using manipulation of cell settling velocity by magnetic field," *Biomicrofluidics,* vol. 7, no. 1, p. 014110, 2013.

[383] T. Djukic, I. Saveljic, G. Pelosi, O. Parodi and N. Filipovic, "Numerical simulation of stent deployment within patient-specific artery and its validation against clinical data," *Computer methods and programs in biomedicine,* vol. 175, pp. 121-127, 2019.

[384] T. Djukic, M. Topalovic and N. Filipovic, "Numerical simulation of isolation of cancer cells in a microfluidic chip," *Journal of Micromechanics and Microengineering,* vol. 25, no. 8, p. 084012, 2015.

[385] T. R. Djukic, S. Karthik, I. Saveljic, V. Djonov and N. Filipovic, "Modeling the behavior of red blood cells within the caudal vein plexus of zebrafish," *Frontiers in physiology,* vol. 7, p. 455, 2016.

[386] N. Filipovic, D. Nikolic, I. Saveljic, T. Djukic, O. Adjic, P. Kovacevic, N. Cemerlic-Adjic and L. Velicki, "Computer simulation of thromboexclusion of the complete aorta in the treatment of chronic type B aneurysm," *Computer Aided Surgery,* vol. 18, no. 1-2, pp. 1-9, 2013.

[387] W. J. Ewens and G. R. Grant, Statistical methods in bioinformatics: an introduction, Springer Science & Business Media, 2006.

[388] S. K. Mathur, Statistical bioinformatics with R, Academic Press, 2009.

[389] J. K. Lee, Statistical bioinformatics: for biomedical and life science researchers, John Wiley & Sons, 2011.

[390] H. H. Lu, B. Schölkopf and H. Zhao, Handbook of statistical bioinformatics, Springer Science & Business Media, 2011.

[391] T. Hu, N. A. Sinnott-Armstrong, J. W. Kiralis, A. S. Andrew, M. R. Karagas and J. H. Moore, "Characterizing genetic interactions in human disease association studies using statistical epistasis networks," *BMC bioinformatics,* vol. 12, no. 1, p. 364, 2011.

[392] K. Van Steen and J. H. Moore, "How to increase our belief in discovered statistical interactions via large-scale association studies?," *Human genetics,* vol. 138, no. 4, pp. 293-305, 2019.

[393] D. Basak, S. Pal and D. C. Patranabis, "Support vector regression," *Neural Information Processing-Letters and Reviews,* vol. 11, no. 10, pp. 203-224, 2007.

[394] X. F. Yan, H. W. Ge and Q. S. Yan, "SVM with RBF kernel and its application research," *Computer Engineering and Design,* vol. 27, no. 11, pp. 1996-1997, 2006.

[395] N. Cristianini and J. Shawe-Taylor, An introduction to support vector machines and other kernel-based learning methods, Cambridge university press, 2000.

[396] I. Steinwart and A. Christmann, Support vector machines, Springer Science & Business Media, 2008.

[397] D. Gorissen, I. Couckuyt, P. Demeester, T. Dhaene and K. Crombecq, "A surrogate modeling and adaptive sampling toolbox for computer based design," *Journal of Machine Learning Research,* vol. 11, pp. 2051-2055, 2010.

[398] H. Kitano, Foundations of systems biology, The MIT Press, 2001.

[399] H. Kitano, "Systems biology: a brief overview," *Science,* vol. 295, no. 5560, pp. 1662-1664, 2002.

[400] F. Boogerd, F. J. Bruggeman, J. H. S. Hofmeyr and H. V. Westerhoff, Systems biology: philosophical foundations, Elsevier, 2007.

[401] A. Oulas, G. Minadakis, K. Sokratous, M. Zachariou, M. M. Bourdakou and G. M. Spyrou, "Systems Bioinformatics: increasing precision of computational diagnostics and therapeutics through network-based approaches," *Briefings in Bioinformatics,* vol. 20, no. 3, pp. 806-824, 2017.

[402] A. Singh, C. P. Shannon, B. Gautier, F. Rohart, M. Vacher, S. J. Tebbutt and K. A. Lê Cao, "DIABLO: an integrative approach for identifying key molecular drivers from multi-omic assays," *Bioinformatics,* 2019.

[403] A. Conesa and S. Beck, "Making multi-omics data accessible to researchers," *Scientific data,* vol. 6, no. 1, pp. 1-4, 2019.

[404] C. Wu, F. Zhou, J. Ren, X. Li, Y. Jiang and S. Ma, "A selective review of multi-level omics data integration using variable selection," *High-throughput,* vol. 8, no. 1, p. 4, 2019.

[405] C. P. Fall, E. S. Marland, J. M. Wagner and J. J. Tyson, Computational cell biology, Interdisciplinary Applied Mathematics, 2002.

[406] E. M. Izhikevich, Dynamical systems in neuroscience, MIT press, 2007.

[407] A. Goldbeter, Biochemical oscillations and cellular rhythms: the molecular bases of periodic and chaotic behaviour, Cambridge university press, 1997.

[408] L. Preziosi, Cancer modelling and simulation, CRC Press, 2003.

[409] S. H. Strogatz, Nonlinear Dynamics and Chaos, Westview Press, 2014.

[410] B. Ermentrout, Simulating, Analyzing, and Animating Dynamical Systems: A Guide to XPPAUT for Researchers and Students, SIAM, 2002.

[411] A. Dhooge, W. Govaerts and Y. A. Kuznetsov, MatCont: A MATLAB package for numerical bifurcation analysis of ODEs, ACM TOMS, 2003.
[412] Y. A. Kunznetsov, Elements of Applied Bifurcation Theory, Springer-Verlag, 1998.
[413] V. I. Arnold, Elements of Applied Bifurcation Theory, Springer-Verlag, 1992.
[414] B. Thomé, Systems engineering: principles and practice of computer-based systems engineering, John Wiley and Sons Ltd., 1993.
[415] M. W. Maier, "Architecting principles for systems-of-systems," *Systems Engineering: The Journal of the International Council on Systems Engineering,* vol. 1, no. 4, pp. 267-284, 1998.
[416] Y. Hata, S. Kobashi and H. Nakajima, "Human health care system of systems," *IEEE Systems Journal,* vol. 3, no. 2, pp. 231-238, 2009.
[417] M. Hucka, D. P. Nickerson, G. D. Bader, F. T. Bergmann, J. Cooper, E. Demir, A. Garny, M. Golebiewski, C. J. Myers, F. Schreiber and D. Waltemath, "Promoting coordinated development of community-based information standards for modeling in biology: the COMBINE initiative," *Frontiers in bioengineering and biotechnology,* vol. 3, p. 19, 2015.
[418] N. J. Stanford, K. Wolstencroft, M. Golebiewski, R. Kania, N. Juty, C. Tomlinson, S. Owen, S. Butcher, H. Hermjakob, N. Le Novère and W. Mueller, "The evolution of standards and data management practices in systems biology," *Molecular systems biology,* vol. 11, no. 12, 2015.
[419] A. Alexander-Bloch, J. N. Giedd and E. Bullmore, "Imaging structural co-variance between human brain regions," *Nat Rev Neurosci.,* vol. 14, no. 5, pp. 322-336, 2013.
[420] A. C. Evans, "Networks of anatomical covariance," *NeuroImage,* vol. 80, p. 489–504, 2013.
[421] B. S. Khundrakpam, A. Reid, J. Brauer, F. Carbonell, J. Lewis, S. Ameis, S. Karama, J. Lee, Z. Chen, S. Das, A. C. Evans and The Brain Development Cooperative Group, "Developmental changes in organization of structural brain networks," *Cerebral Cortex,* vol. 23, no. 9, p. 2072–2085, 2013.
[422] B. A. Zielinski, E. D. Gennatas, J. Zhou and W. W. Seeley, "Network-level structural covariance in the developing brain," *PNAS,* vol. 107, no. 42, p. 18191–18196, 2010.
[423] D. S. Bassett, E. Bullmore, B. A. Verchinski, V. S. Mattay, D. R. Weinberger and A. Meyer-Lindenberg, "Hierarchical organization of human cortical networks in health and schizophrenia," *Journal of Neuroscience,* vol. 28, no. 37, p. 9239–9248, 2008.
[424] R. A. I. Bethlehem, R. Romero-Garcia, E. Mak, E. T. Bullmore and S. Baron-Cohen, "Structural covariance networks in children with autism or ADHD," *Cerebral Cortex,* vol. 27, no. 8, p. 4267–4276, 2017.
[425] Y. He, Z. Chen, G. Gong and A. Evans, "Neuronal networks in Alzheimer's disease," *The Neuroscientist,* vol. 15, no. 4, p. 33–350, 2009.
[426] M. Sharda, B. S. Khundrakpam, A. C. Evans and N. C. Singh, "Disruption of structural covariance networks for language in autism is modulated by verbal ability," *Brain Structure and Function,* vol. 221, no. 2, p. 1017–1032, 2016.
[427] J. Solé-Casals, J. M. Serra-Grabulosa, R. Romero-Garcia, G. Vilaseca, A. Adan, N. Vilaró, N. Bargalló and E. T. Bullmore, "Structural brain network of gifted

children has more integrated and versatile topology," *Brain Structure and Function,* vol. 224, no. 7, p. 2373–2383, 2019.

[428] P. Holme and J. Saramäki, "Temporal networks," *Physics reports,* vol. 519, no. 3, pp. 97-125, 2012.

[429] P. Holme and J. Saramäki, Temporal networks, Springer, 2013.

[430] J. Gunawardena, "Time-scale separation - Michaelis and Menten's old idea, still bearing fruit," *The FEBS journal,* vol. 281, no. 2, pp. 473-488, 2014.

[431] F. Rocha, M. Aguiar, M. Souza and N. Stollenwerk, "Time-scale separation and centre manifold analysis describing vector-borne disease dynamics," *International Journal of Computer Mathematics,* vol. 90, no. 10, pp. 2105-2125, 2013.

[432] J. Gunawardena, "A linear framework for time-scale separation in nonlinear biochemical systems," *PLoS one,* vol. 7, no. 5, p. e36321, 2012.

[433] D. P. Borcard, P. Legendre and P. Drapeau, "Partialling out the spatial component of ecological variation," *Ecology,* vol. 73, no. 3, pp. 1045-1055, 1992.

[434] A. Duchene, R. E. Graves and P. Brugger, "Schizotypal thinking and associative processing: a response commonality analysis of verbal fluency," *Journal of Psychiatry and Neuroscience,* vol. 23, no. 1, p. 56, 1998.

[435] M. Stellefson, J. F. Yannessa and G. F. Martel, "Using canonical commonality analysis to examine the predictive quality of aging and falls efficacy on balance functioning in older adults," *Evaluation & the health professions,* vol. 35, no. 2, pp. 239-255, 2012.