FINAL PROGRAM

&

BOOK OF ABSTRACTS

FCT FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

CEFITEC
CENTRO DE FÍSICA E INVESTIGAÇÃO TECNOLÓGICA
Centre of Physics and Technological Research

VENUE

Room 204, Building IV (E.IV), Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa, Campus Caparica, Portugal

FOREWORD

This book contains the final program and the abstracts of CIBB2018, the 15th International Conference on Computational Intelligence methods for Bioinformatics and Biostatistics, an international conference which provides a multi-disciplinary forum for researchers interested in the application of computational intelligence, in a broad sense, to open problems in bioinformatics, biostatistics, systems and synthetic biology and medical informatics. Cutting edge methodologies capable of accelerating life science discoveries will be addressed. Following its tradition and roots, this year's meeting will bring together researchers from the international scientific community interested in advancements and future perspectives in bioinformatics and biostatistics. Under this compliance, current trends and future opportunities at the edge of computer and life sciences, the application of computational intelligence to a system and synthetic biology and the consequent impact on innovative medicine will be of great interest for the conference. Theoretical and experimental biologists were also invited to participate to present novel challenges and foster multidisciplinary collaboration. This meeting was sponsored by Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa and Center of Physics and Technological Research (CEFITEC), Portugal. The Conference Program only includes oral presentations, among which we are honoured to announce plenary keynote lectures, given by Alberto Paccanaro (University of London, United Kingdom), Alexandra Carvalho (Universidade de Lisboa, Portugal), Benoit Liquet (University of Pau and Pays de l'Adour, France), Fernando L. Ferreira (Universidade Nova de Lisboa, Portugal) and Veronica Vinciotti (Brunel University London, United Kingdom).

The organization has attributed four fees grants to PhD students. These grants were conferred having into account the average punctuation given by the reviewers.

Organizers also expect and seek the event to contribute for the strengthening of collaborative research within the fields of Computational Intelligence methods for Bioinformatics and Biostatistics. A Social Dinner on the evening of 7th September 2018 at Casa da Cerca, a beautiful monument of Almada city, will contribute fruitfully to foster alliances.

We would like to express our thanks to all participants, all members of the Program Committee and to the secretariat.

Conference Chairs

Maria Raposo, CEFITEC/FCT/UNL, Portugal
Paulo A. Ribeiro, CEFITEC/FCT/UNL, Portugal
Susana Sério, CEFITEC/FCT/UNL, Portugal

ORGANIZATION

GENERAL CHAIRS

Maria Raposo, Universidade Nova de Lisboa, Portugal
Paulo Lisboa, Liverpool John Moores University, United Kingdom
Giorgio Valentini, Università degli Studi di Milano, Italy

BIOSTATISTICS TECHNICAL CHAIR

Chiara Brombin, Università Vita-Salute San Raffaele, Italy

BIOINFORMATICS TECHNICAL CHAIR

Angelo Ciaramella, Università degli Studi di Napoli "Parthenope", Italy

SPECIAL SESSION ORGANIZER

Andrea Bracciali, University of Stirling, United Kingdom
Antonino Staiano, Università degli Studi di Napoli "Parthenope", Italy

STEERING COMMITTEE

Pierre Baldi, University of California, Irvine, CA, USA
Elia Biganzoli, University of Milan, Italy
Clelia Di Serio, University Vita-Salute San Raffaele, Italy
Alexandru Floares, Oncological Institute Cluj-Napoca, Romania
Jon Garibaldi, University of Nottingham, United Kingdom
Nikola Kasabov, Auckland University of Technology, New Zealand
Francesco Masulli, University of Genova, Italy and Temple University, PA, USA
Leif Peterson, TMHRI, Houston, Texas, USA
Roberto Tagliaferri, University of Salerno, Italy

PROGRAMME COMMITTEE

Claudio Angione, Teesside University, United Kingdom
Sansanee Auephanwiriyakul, Chiang Mai University, Thailand
Gilles Bernot, University of Nice Sophia Antipolis, France
Andrea Bracciali, University of Stirling, United Kingdom
Stefan Canzar, Saarland University, Netherlands
Giulio Caravagna, University of Edinburgh, United Kingdom
Davide Chicco, Princess Margaret Cancer Centre, Canada
Angelo Ciaramella, University of Naples "Parthenope", Italy
Luisa Cutillo, University of Sheffield, United Kingdom
Angelo Facchiano, CNR - Istituto di Scienze dell'Alimentazione, Italy
Enrico Formenti, Nice Sophia Antipolis University, France
Christoph M. Friedrich, University of Applied Science and Arts Dortmund, Germany
Yair Goldberg, University of Haifa, Israel
Marco Grzegorczyk, Groningen University, Netherlands
Giosue' Lo Bosco, University of Palermo, Italy
Hassan Mahmoud, University of Genova, Italy
Anna Marabotti, University of Salerno, Italy
Elena Marchiori, Radboud University, Netherlands
Tobias Marschall, Saarland University / Max Planck Institute for Informatics, Germany
Mauri Giancarlo, University of Milano-Bicocca, Italy
Bud Mishra, Courant Institute, NYU, and School of Medicine, Mt Sinai, USA
Marianna Pensky, University of Central Florida, USA
Nadia Pisanti, Universita di Pisa, Italy & Erable Team, INRIA Italy
Vilda Purutçuoğlu, Middle East Technical University, Turkey
Paolo Romano, IRCCS University Hospital San Martino IST, Italy
Simona Ester Rombo, University of Palermo, Italy
Stefano Rovetta, University of Genova, Italy
Antonino Staiano, University of Napoli "Parthenope", Italy
Francesco Stingo, MD Anderson USA
Paolo Tieri, Consiglio Nazionale delle Ricerche, Italy
Alfonso Urso, ICAR-CNR, Italy
Filippo Utro, IBM T.J. Watson Research Center, USA
Alfredo Vellido, Universitat Politecnica de Catalunya, Spain
Pawel P. Labaj, Boku University Vienna, Austria
Maria Raposo, Universidade Nova de Lisboa, Portugal
Paulo A. Ribeiro, Universidade Nova de Lisboa, Portugal
Susana Sério, Universidade Nova de Lisboa, Portugal
Marco Frasca, Università degli Studi di Milano, Italy
Armando Blanco, Universidad de Granada, Spain
Matteo Re, Universitá degli Studi di Milano, Italy

ORGANISATION COMMITTEE

Maria Raposo, Universidade Nova de Lisboa, Portugal
Paulo A. Ribeiro, Universidade Nova de Lisboa, Portugal
Susana Sério, Universidade Nova de Lisboa, Portugal
Andrea Bracciali, University of Stirling, United Kingdom
Filipa Pires, Universidade Nova de Lisboa, Portugal
Gonçalo Magalhães-Mota, Universidade Nova de Lisboa, Portugal
Telma Marques, Universidade Nova de Lisboa, Portugal
Sara Pereira, Universidade Nova de Lisboa, Portugal
Jeniffer Farias dos Santos, Universidade de São Paulo, Brazil
Paulo Zagalo, Universidade Nova de Lisboa, Portugal
Thais P. Pivetta, Universidade Nova de Lisboa, Portugal
João Pereira da Silva, Universidade Nova de Lisboa, Portugal
Ana Cruz, Universidade Nova de Lisboa, Portugal (Secretariat)

PROGRAM

Thursday 6th September 2018

Venue: Room 204, Building IV (E.IV)

8:30-9:00	**Reception**
9:00-9:10	**Opening Ceremony**
Session A	**Plenary Lecture** Chair: *Brombin Chiara*
9:10- 9:50	**Benoit Liquet** *A Unified Regularized Group PLS Algorithm Scalable to Big Data. Application on genomics data*
Session B	**Machine explanation – Interpretation of Machine Learning Models for Medicine and Bioinformatics** Chair: *Roberto Tagliaferri*
9:50-10:10	Raul V. Casana-Eslava, Ian H. Jarman, Sandra Ortega-Martorell, Paulo J. Lisboa and José D. Martínez-Guerrero *Structure finding stabilization and optimization with the PC algorithm*
10:10-10:30	Ahsan Bilal, Alfredo Vellido and Vicent Ribas *Enabling interpretation of the outcome of a human obesity prediction machine learning analysis from genomic data*
10:30-10:50	**Coffee Break**
Session C	**Plenary Lecture** Chair: *Angelo Ciaramella*
10:50-11:30	**Veronica Vinciotti** *Sparse graphical models in genomics: an application to censored qPCR data*
Session D	**Computational Intelligence Methods for Bioinformatics and Biostatistics –I** Chair: *Usman Sumo Friend Tambuna*
11:30 -11:50	Marco Frasca, Maryam Sepehri, Alessandro Petrini, Giuliano Grossi and Giorgio Valentini *Committee-based Active Learning to Select Negative Examples for Predicting Protein Functions*
11:50 -12:10	Vanessa D'Amario, Gabriele Arnulfo, Lino Nobili and Annalisa Barla *Classification of epileptic activity through temporal and spatial characterization of intracranical recordings*
12:10 -12:30	Vincenzo Bonnici, Simone Caligola, Antonino Aparo and Rosalba Giugno. *Centrality speeds the subgraph isomorphism search up in target aware contexts*

12:30-14:00	**Lunch**
Session E	**Fast and Efficient Solutions for Computational Intelligence Methods in Bioinformatics, Systems and Computational Biology** Chair: *Leif Peterson*
14:00- 14:20	Stefano Beretta, Mauro Castelli, Ivo Gonçalves and Daniele Ramazzotti. *A quantitative assessment of the effect of different algorithmic schemes to the task of learning the structure of Bayesian Networks*
14:20- 14:40	Andrea Tangherloni, Simone Spolaor, Leonardo Rundo, Marco S. Nobile, Paolo Cazzaniga, Giancarlo Mauri, Pietro Lio, Daniela Besozzi and Ivan Merelli *GENHAP: Evolutionary Computation for Haplotype Assembly*
14:40- 15:00	Stefano Beretta, Paola Bonizzoni and Ivan Merelli *HaploVir: Inferring Viral Haplotypes from Deep Sequencing of a Mixture of Strains using a Path Decomposition Algorithm*
15:00-15:20	Hassan Pazira *Improved Predictor-Corrector Algorithm*
15:20-15:40	Jade Hind, Abir Hussain, Dhiya Al-Jumeily, Casimiro Aday Curbelo Montañez, Basma Abdulaimma, Carl Chalmers and Paulo Lisboa *Association Analysis Shift Regularisation: A Proposed Methodology for Epistasis Interaction in Genomic Studies*
15:40 -16:00	**Coffee Break**
Session F	**Engineering Bio-Interfaces and Rudimentary Cells as a Way to Develop Synthetic Biology** Chair: *Quirina Ferreira*
16:00- 16:20	Mónica Machado and Quirina Ferreira *Drug Delivery Films for Ocular Diseases Treatment*
16:20-16:40	João Pereira da Silva, Paulo Morgado Zagalo, Paulo A. Ribeiro and Maria Raposo *Adsorption of Triclosan on sensors based on PAH/PAZO thin-films: the effect of pH*
16:40-17:00	Filipa Pires, Bárbara Rodrigues, Gonçalo Magalhães-Mota, Paulo António Ribeiro and Maria Raposo *Effect of epigallocatechin-3-gallate on DMPC oxidation revealed by infrared spectroscopy*
17:00-17:20	Thais Pivetta, Filipa Pires and Maria Raposo *Effect of EGCG on the DNA in presence of UV Radiation*
17:20-17:40	Sara Pereira, Érica Pinto, Paulo Ribeiro and Susana Sério *Non-Thermal Atmospheric pressure Plasmas: Generation, Sources and Applications*

17:40-18:00 Gonçalo Magalhães-Mota, Filipa Pires, Paulo António Ribeiro and Maria Raposo
Detection of Triclosan Dioxins After UV Irradiation – A Preliminar Study

18:00-18:20 Telma Marques, Sam Eden, Małgorzata A. Śmiałek, Ilko Bald, Maria Raposo and Nigel J. Mason
Enhanced degradation of laser-irradiated DNA using Gold Nanoparticles

18:20-18:40 Catarina Delfino and Quirina Ferreira
Real-time adsorption of a Co-octaethylporphyrin monitorized with scanning tunneling microscopy

Friday 7th September 2018

Venue: Room 204, Building IV (E.IV)

Session G **Plenary Lecture**
Chair: Giorgio Valentini

9:00- 9:40 **Alberto Paccanaro**
Answering questions in biology and medicine by making inferences on networks

Session H **Soft Computing Methods for Characterizing Diseases from Omics Data**
Chair: *Marzio Pennisi*

9:40- 10:00 Antonino Fiannaca, Massimo La Rosa, Laura La Paglia, Alfonso Urso, Riccardo Rizzo and Giosue' Lo Bosco
Identification of Key miRNAs in regulation of PPI Networks

10:00- 10:20 Domenico Amato, Mattia Di Gangi, Giosue' Lo Bosco and Riccardo Rizzo
Recurrent Deep Neural Networks for Nucleosome Classification

10:20-10:40 Angelo Ciaramella, Davide Nardone and Antonino Staiano
Compressive Sensing and Hierarchical Clustering for Microarray Data with Missing Values

10:40-11:00 **Coffee Break**

Session I **Computational Methods for Neuroimaging Analysis**
Chair: *Marco Grzegorczyk*

11:00-11:20 Max Garagnani, Evgeniya Kirilina and Friedemann Pulvermüller.
Perception-action circuits for word learning and semantic grounding: a neurocomputational model and neuroimaging study

11:20-11:40 Angela Serra, Antonio Della Pietra, Marcus Herdener, Roberto Tagliaferri and Fabrizio Esposito
Automatic discrimination of auditory stimuli perceived by the human brain

11:40-12:00 Simeon Spasov, Olaf Hauk and Seyedeh-Rezvan Farahibozorg.
Decoding Semantic Word Categories from Electro- and Magnetoencephalography data

12:00-12:20 Tiago Fernandes, João Pereira, Bruno Direito, Alexandre Sayal and Miguel Castelo-Branco
Statistical validation of State-Space Granger Causality with Time-Reverse Surrogate and its boundaries in fMRI data

12:20-12:40 Razvan E. Kusztos, Giovanna Maria Dimitri and Pietro Liò
Neural Models for Brain Networks Connectivity Analysis

12:40-14:00 **Lunch**

Session J	**Plenary Lecture** Chair: *Paulo Lisboa*
14:00- 14:40	**Alexandra Carvalho** *Model selection for temporal biomedical data*
Session K	**Networking biostatistics and bioinformatics** Chair: *Clelia di Serio*
14:40- 15:00	Monica Chiogna, Vera Djordjilovic, Chiara Romualdi and Elisa Salviato *Searching for the Source of Difference: a Graphical Model Approach*
15:00- 15:20	Federica Cugnata and Paola M.V. Rancoita *Improving Sensitivity and Specificity Calculation from Myocardial Imaging Data*
15:20- 15:40	Luca Del Core, Eugenio Montini, Clelia Di Serio and Andrea Calabria *Longitudinal Studies and Integrative Biology with Heterogeneous data: an approach using rarefaction*
15:40- 16:00	Mahdi Shafiee Kamalabad and Marco Grzegorczyk *A new partially Coupled Piece-Wise linear Regression Model for statistical network Structure Inference*
16:00 -16:20	**Coffee Break**
Session L	**Computational Intelligence Methods for Bioinformatics and Biostatistics –II** Chair: *Stefano Beretta*
16:20-16:40	Carlos Cano *Computational Annotation of Genetic Biomarkers using Topologically Associating Domains*
16:40-17:00	João Villa-Brito, Marta B. Lopes, Alexandra M. Carvalho and Susana Vinga *Unravelling breast and prostate common gene signatures by Bayesian network learning*
17:00-17:20	Pedro Ferreira, Alexandra Carvalho and Susana Vinga *Variational inference in probabilistic single-cell RNA-seq models*
17:20-17:40	Bozidar Popovic *New method for getting probability distributions with application*
17:40-18:00	Laxmi Parida and Filippo Utro *PiXora: Simultaneous Phasing of Multiple Polyploids*
18:00-18:40	**CIBB 2019 Meeting** *Chair: Roberto Tagliaferri*
19:00-22:30	**Social Dinner**

Saturday 8th September 2018

Venue: Room 204, Building IV (E.IV)

Session M	**Plenary Lecture** Chair: *Antonino Staiano*
9:00- 9:40	**Fernando Luís Ferreira** *Ethics and our moral in research, let's think about it!*
Session N	**Computational Intelligence Methods for Bioinformatics and Biostatistics –III** Chair: *Hassan Pazira*
9:40- 10:00	Marta Lovino, Gianvito Urgese, Enrico Macii, Santa Di Cataldo and Elisa Ficarra *Predicting the Oncogenic Potential of Gene Fusions using Convolutional Neural Networks*
10:00-10:20	Ivan Olier, Philippa Grace McCabe, Disha Agarwal and Sandra Ortega-Martorell *Benchmarking Multitask Learning for QSARS*
10:20-10:40	Sebastian Daberdaku *Paratope identification by classification of local antibody surface patches enriched with eight physicochemical properties*
10:40-11:00	Pietro Lio and Andrea Bracciali *Computational Intelligence to explore the link between diabetes and arthritis*
11:00-11:20	**Coffee Break**
Session O	**Modeling and Simulation Methods for System Biology and System Medicine** Chair: *Fabrizio Frasca*
11:20-11:40	Chiara Damiani, Dario Pescini and Marco S. Nobile *Global Sensitivity Analysis of Constraint-based Metabolic Models*
11:40-12:00	Ahmad Husein Alkaff, Mutiara Saragih, Mochammad Arfin Fardiansyah Nasution and Usman Sumo Friend Tambunan *Inhibition of Primed Ebola Virus Glycoprotein by Peptide Compound Conjugated to HIV-1 Tat Peptide Through a Virtual Screening Approach*
12:00-12:20	Niccolò Totis, Andrea Tangherloni, Marco Beccuti, Paolo Cazzaniga, Marco Nobile, Daniela Besozzi, Marzio Pennisi and Francesco Pappalardo *GPU Powered Parameter Estimation of a large-Scale Kinetic Metabolic Model*
12:20-12:40	Mochammad Arfin Fardiansyah Nasution, Ahmad Husein Alkaff, Ilmi Fadhilah Rizki and Usman Sumo Friend Tambunan

Pharmacophore Modelling, Virtual Screening, and Molecular Docking Simulations of Natural Product Compounds as Potential Inhibitors of Ebola Virus Nucleoprotein

12:40-14:00	**Lunch**

Session P **Machine Learning in Health Informatics and Biological Systems**
Chair: *Marco Masseroli*

14:00- 14:20 Davide Castaldi, Ilaria Giordani, Antonio Candelieri, Francesco Archetti and Roberto Mattina
The Scourge of Anti-microbial Resistance: A Machine learning Ap proach for prescription Patterns Analytics

14:20- 14:40 Fabrizio Frasca, Matteo Matteucci, Marco Morelli and Marco Masseroli
Unveiling Gene Expression Histonic Regulative Patterns by Hyperplanes Clustering

14:40-15:00 Arif Canakoglu, Luca Nanni, Artur Sokolovsky and Stefano Ceri.
Designing and Evaluating Deep Learning Methods for Cancer Classification on Gene Expression Data

15:00-15:20 Hugo Martiniano, Muhammad Asif, Astrid Vicente and Luís Correia
A Semi-supervised Learning Approach to the Identification and Prioritization of Genes Associated with Autism Spectrum Disorder

15:20-15:40 Joana Moreira, Mariana Moreira, Nuno Pombo and Nuno Garcia
Identification of real and imaginary movements in EEG using Machine Learning models

15:40 -16:00	**Coffee Break**

Session Q **Computational Intelligence Methods for Bioinformatics and Biostatistics –IV**
Chair: *Marco Grzegorczyk*

16:00 -16:20 Alessio Mancini, Claudio Angione, Pietro Liò and Sandra Pucciarelli
The ciliates as complex eukaryotic single cell models for studying molecular communication: an insight on the evolution of signaling molecules also involved in human brain metabolism

16:20 -16:40 Leif Peterson and Timothy Thompson
In Silico ADME and Toxicity Prediction Using N-grams for Chemical Fingerprinting

16:40 -17:00 Cheick Tidiane Ba, Elena Casiraghi, Marco Frasca, Jessica Gliozzo, Giuliano Grossi, Marco Mesiti, Marco Notaro, Paolo Perlasca, Alessandro Petrini, Matteo Re and Giorgio Valentini
A Graphical Tool for the Exploration and Visual Analysis of Biomolecular Networks

17:00-17:20 Manuela Ferrario, Alice Cambiaghi, Eliandre de Olivera and Roberta Pastorelli

	An innovative approach to integrate proteomics and metabolomics data in severe septic shock
17:20-17:40	Casimiro Aday Curbelo Montañez, Paul Fergus, Carl Chalmers and Jade Hind *Analysis of Extremely Obese Individuals Using Deep Learning Stacked Autoencoders and Genome-Wide Genetic Data*
17:40-18:00	Fernando Luis-Ferreira, Daniel Rodrigues, João Sarraipa and Ricardo Goncalves *Smartwatch Based Fall Detection System for Elders' Surveillance*
18:00-18:10	**Organizers & CIBB2019 Organizers End of Meeting & Farewell**

ABSTRACTS

Session A
Plenary Lecture

Thursday, 6 September 2018, 9:10-9:50

A Unified Regularized Group PLS Algorithm Scalable to Big Data. Application on genomics data

Benoit Liquet
University of Pau and Pays de l'Adour, France

Partial Least Squares (PLS) methods have been heavily exploited to analyse the association between two blocs of data. These powerful approaches can be applied to data sets where the number of variables is greater than the number of observations and in presence of high collinearity between variables. Different sparse versions of PLS have been developed to integrate multiple data sets while simultaneously selecting the contributing variables. Sparse modelling is a key factor in obtaining better estimators and identifying associations between multiple data sets. The cornerstone of the sparsity version of PLS methods is the link between the SVD of a matrix (constructed from deflated versions of the original matrices of data) and least squares minimisation in linear regression. We present here an accurate description of the most popular PLS methods, alongside their mathematical proofs. A unified algorithm is proposed to perform all four types of PLS including their regularised versions. Our methods enable us to identify important relationships between genomic expression and cytokine data from an HIV vaccination trial. We also proposed a new methodology by accounting for both grouping of genetic markers (e.g. genesets) and temporal effects. Finally, various approaches to decrease the computation time are offered, and we show how the whole procedure can be scalable to big data sets.

Session B
Machine explanation – Interpretation of Machine Learning Models for Medicine and Bioinformatics

Thursday, 6 September 2018, 9:50-10:10

Structure finding stabilization and optimization with the PC algorithm (Paper #2)

Raul V. Casana-Eslava[1], Ian H. Jarman[1], Sandra Ortega-Martorell[1], Paulo J. Lisboa[1] and José D. Martínez-Guerrero[2]
[1] *Department of Applied Mathematics, Liverpool John Moores University (LJMU)*
r.v.casanaeslava@ljmu.ac.uk, i.h.jarman@ljmu.ac.uk, s.ortegamartorell@ljmu.ac.uk, p.j.lisboa@ljmu.ac.uk

[2]*Departament d'Enginyeria Electrònica - ETSE, Universitat de València (UV), jose.d.martin@uv.es*

In the field of structure finding, the PC algorithm is a well-known constraint based algorithm used to build a Directed Acyclic Graph (DAG) from Conditional Independence maps where a major challenge is to minimize errors in the graph structure. This work presents empirical evidence for best practice: to reduce false positive errors via the False Discovery Rate (FDR), and to identify appropriate parameter settings to improve the False Negative Reduction (FNR). In addition, several node ordering policies are investigated that transform the skeleton graph into a DAG (edges orienting rules), the results show that ordering nodes by strength of mutual information recovers a representative DAG in reasonable time, although a more accurate graph can be recovered using a random order of samples at the expense of increasing the computation time.

Thursday, 6 September 2018, 10:10-10:30

Enabling interpretation of the outcome of a human obesity prediction machine learning analysis from genomic data (Paper #12)

Ahsan Bilal[1,2], Alfredo Vellido[1,3] and Vicent Ribas[2]
[1]*Universitat Politècnica de Catalunya (UPC BarcelonaTech) Barcelona 08034, Spain, avellido@cs.upc.edu*
[2] *EURECAT: Centre Tecnològic de Catalunya, Barcelona 08005, Spain, vicent.ribas@eurecat.org*
[3] *Intelligent Data Engineering and Artificial Intelligence (IDEAI) Research Center Barcelona 08034, Spain*

In this brief paper, we address the medical problem of human obesity prediction from genomic data. Genomic datasets may contain a huge number of features and they often have to be analyzed within the realm of Big Data technologies. As a medical problem, obesity prediction would welcome interpretable outcomes. Therefore, the analyst would benefit from approaches in which the problem of very high data dimensionality could be eased as much as possible. Feature selection can be an essential part of such approaches. In this context, though, traditional machine learning methods may struggle. Here, we propose a pipeline to address this problem using partitioning strategies: both vertical, by dividing the data based on gender, and horizontal, by splitting each of the analyzed chromosomes into 5,000-instances subsets. For each, Minimum Redundancy and Maximum Relevance feature selection is used to find rankings of the single nucleotide polymorphisms most relevant for classification in the medical dataset.

Session C
Plenary Lecture

Thursday, 6 September 2018, 10:50-11:30

Sparse graphical models in genomics: an application to censored qPCR data

Veronica Vinciotti
Department of Mathematics, Brunel University London, Uxbridge UB8 3PH, London, UK.

Regularized inference of networks using graphical modelling approaches has seen many applications in biology, most notably in the recovery of regulatory networks from high-dimensional gene expression data. Various extensions to the standard graphical lasso approach have been proposed, such as dynamic and hierarchical graphical models. In this talk, I will focus on a latest extension to censored graphical models in order to deal with censored data such as qPCR data. We propose a computationally efficient EM-like algorithm for the estimation of the conditional independence graph and thus the recovery of the underlying regulatory network.

Session D
Computational Intelligence Methods for Bioinformatics and Biostatistics –I

Thursday, 6 September 2018, 11:30-11:50

Committee-based Active Learning to Select Negative Examples for Predicting Protein Functions (Paper #5)

Marco Frasca, Maryam Sepehri, Alessandro Petrini, Giuliano Grossi and Giorgio Valentini
Dipartimento di Informatica, Universit´a degli Studi di Milano, Via Comelico 39, 20135, frasca@di.unimi.it, maryam.sepehri@unimi.it, alessandro.petrini@unimi.it, grossi@di.unimi.it, valentini@di.unimi.it

The Automated Functional Prediction (AFP) of proteins became a challenging problem in bioinformatics and biomedicine aiming at handling and interpreting the extremely large-sized proteomes of several eukaryotic organisms. A central issue in AFP is the absence of public repositories for protein functions, e.g. the Gene Ontology (GO), of well defined sets of negative examples to learn accurate classifiers for AFP. In this paper we investigate the Query by Committee paradigm of active learning to select the negatives most informative for the the classifier and the protein function to be inferred. We validated our approach in predicting the Gene Ontology function for the S.cerevisiae proteins.

Thursday, 6 September 2018, 11:50-12:10

Classification of epileptic activity through temporal and spatial characterization of intracranical recordings (Paper #6)

Vanessa D'Amario[1], Gabriele Arnulfo[1], Lino Nobili[1] and Annalisa Barla[2]
[1]*DIBRIS, Universit`a degli Studi di Genova, Genova, Italy,*
[2]*Ospedale Niguarda Ca' Granda, Milano, Italy*

Focal epilepsy is a chronic condition characterized by hyper-activity and abnormal synchronization of a specific brain region. For pharmacoresistant patients, the surgical resection of the critical area is considered a valid clinical solution, therefore, an accurate localization is crucial to minimize neurological damage. In current clinical routine the characterization of the epileptogenic zone (EZ) is performed using invasive methods, such as Stereo-Electroencephalography (SEEG). Medical experts perform the tag of neural electrophysiological recordings by visually inspecting the acquired data, a highly time consuming and subjective procedure. Here we show the results of an automatic multi-modal

classification method for the evaluation of critical areas in focal epileptic patients. The proposed method represents an attempt in the characterization of brain areas which integrates the anatomical information on neural tissue, inferred using Magnetic Resonance Imaging (MRI) in combination with spectral features extracted from SEEG recordings.

Thursday, 6 September 2018, 12:10-12:30

Centrality speeds the subgraph isomorphism search up in target aware contexts (Paper #11)

Vincenzo Bonnici, Simone Caligola, Antonino Aparo and Rosalba Giugno
University of Verona, Department of Computer Science, Strada le Grazie 15, 37134 Verona (Italy), vincenzo.bonnici@univr.it, simone.caligola@univr.it, antonino.aparo@univr.it, rosalba.giugno@univr.it

Subgraph isomorphism (SubGI) is known to be a NP-Complete problem. Several methodologies use heuristic approaches to solve it, differing into the strategy to search the occurrences of a graph into another. This choice strongly influences their computational effort requirement. We investigate seven search strategies where global and local topological properties of the graphs are exploited by means of weighted graph centrality measures. Results on benchmarks of biological networks show the competitiveness of the proposed seven alternatives and that, among them, local strategies predominate on sparse target graphs, and closeness- and eigenvector-based strategies outperform on dense graphs.

Session E
Fast and Efficient Solutions for Computational Intelligence Methods in Bioinformatics, Systems and Computational Biology

Thursday, 6 September 2018, 14:00-14:20

A quantitative assessment of the effect of different algorithmic schemes to the task of learning the structure of Bayesian Networks (Paper #15)

Stefano Beretta[1], Mauro Castelli[2], Ivo Gonçalves[3] and Daniele Ramazzotti[4]
[1]*DISCo, Universit´a degli Studi di Milano-Bicocca. 20126 Milano, Italy. stefano.beretta@disco.unimib.it*
[2]*NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal. mcastelli@novaims.unl.pt*
[3]*INESC Coimbra, DEEC, University of Coimbra, Portugal. icpg@dei.uc.pt*
[4]*Department of Pathology, Stanford University, California, USA. daniele.ramazzotti@stanford.edu*

One of the most challenging tasks when adopting Bayesian Networks (BNs) is learning their structure from data. This task is complicated by the huge search space of possible solutions and proved to be a well-known NP-hard problem and, hence, approximations are required. However, to the best of our knowledge, a quantitative analysis of the performance and characteristics of the different heuristics to solve this problem has never been done before. In this work, we provide a study of the different state-of-the-arts methods for structural learning

on simulated data, considering both BNs with discrete and continuous variables, and with different rates of noise in the data. We investigate the characteristics of different scores proposed for the inference and the statistical pitfalls within them.

Thursday, 6 September 2018, 14:20-14:40

GENHAP: Evolutionary Computation for Haplotype Assembly (Paper #27)

Andrea Tangherloni[1], Simone Spolaor[1], Leonardo Rundo[1], Marco S. Nobile[1], Paolo Cazzaniga[1], Giancarlo Mauri[1], Pietro Lio[1], Daniela Besozzi[1] and Ivan Merelli[1]
[1]*Department of Informatics, Systems and Communication University of Milano-Bicocca, Milano, Italy, andrea.tangherloni@disco.unimib.it*
[2]*Department of Human and Social Sciences, University of Bergamo, Bergamo, Italy*
[3]*Computer Laboratory, University of Cambridge, Cambridge, United Kingdom*
[4]*Institute of Biomedical Technologies, Italian National Research Council, Segrate (MI), Italy*

The reconstruction of the two distinct copies of each chromosome, called haplotypes, is an essential process for the characterization of the genome of an individual. Here we address a successful approach for haplotype assembly, called the weighted Minimum Error Correction (wMEC) problem, which consists in computing the two haplotypes that partition the sequencing reads into two disjoint sub-sets with the least number of corrections to the Single Nucleotide Polymorphisms values. To solve this problem we propose GenHap, a computational method based on Genetic Algorithms, which are able to obtain optimal solutions thanks to a global search process. To evaluate the effectiveness of GenHap, we test it on a synthetic (yet realistic) dataset based on the PacBio RS II sequencing technology. We compare the performance of GenHap against HapCol, an efficient state-of-the-art algorithm for haplotype assembly. We show that GenHap always obtains high accuracy solutions (in terms of haplotype error rate), and is up to 20x faster than HapCol on this synthetic (yet realistic) dataset.

Thursday, 6 September 2018, 14:40-15:00

HaploVir: Inferring Viral Haplotypes from Deep Sequencing of a Mixture of Strains using a Path Decomposition Algorithm (Paper #37)

Stefano Beretta[1,2], Paola Bonizzoni[1] and Ivan Merelli[2]
[1]*DISCo, Universit´a degli Studi di Milano-Bicocca. 20126 Milano, Italy. fberetta,bonizzonig@disco.unimib.it*
[2]*Istituto di Tecnologie Biomediche, Consiglio Nazionale delle Ricerche. 20090 Segrate, Italy. ivan.merelli@itb.cnr.it*

Retroviruses mutate very fast and tend to generate many potentially drug-resistant clones within infected patients. Being able to identify these different mutants is important for efficient drug administration. Blood samples from patients can be sequenced in order to identify the retroviral clones. Computationally, this problem can be regarded as a nonstandard clustering issue due to missing pairwise similarity measures between non-overlapping reads. This paper presents a new computational technique for the identification of viral haplotypes. We model this as a computational problem based on graphs, and we exploit two path reconstruction algorithms on the graph built on the computed SNP calls. These two procedures are implemented in a tool called HaploVir which exploits these techniques to compute the different clones: MaxFlow which is based on maximum flow of the graph, and MaxPaths which is a greedy approach that finds the maximal augmented paths. We perform and experimental analysis on real data to

assess the performance of HaploVir in identifying viral clones from a set of input sequences. HaploVir is implemented in Python and its source code is freely available at https://bitbucket.org/bereste/haplovir with some example datasets.

Thursday, 6 September 2018, 15:00-15:20

Improved Predictor-Corrector Algorithm (Paper #19)

Hassan Pazira
Johann Bernoulli Institute, University of Groningen, The Netherlands, h.pazira@rug.nl

From a computational point of view, the dgLARS method consists essentially in the computation of the implicitly defined solution curve. In [1] this problem is satisfactorily solved by using a predictor-corrector (PC) algorithm, that however has the drawback of becoming intractable when working with thousands of predictors. Using the PC algorithm lead to an increase in the run times needed for computing the solution curve. In this paper we explain an improved version of the PC algorithm (IPC), proposed in [2], to decrease the effects stemming from this problem for computing the solution curve. The IPC algorithm allows the dgLARS method to be implemented using less steps, greatly reducing the computational burden because of reducing the number of points of the solution curve.

[1] L. Augugliaro and A. M. Mineo and E. C. Wit. "Differential Geometric Least Angle Regression: A Differential Geometric Approach to Sparse Generalized Linear Models". Journal of the Royal Statistical Society: Series B, vol. 75, n.3, pp. 471–498, 2013.
[2] H. Pazira and L. Augugliaro and E. C. Wit. "Extended differential geometric LARS for highdimensional GLMs with general dispersion parameter". Statistics and Computing, vol.28, n.4, pp.753–774, 2018.

Thursday, 6 September 2018, 15:20-15:40

Association Analysis Shift Regularisation: A Proposed Methodology for Epistasis Interaction in Genomic Studies (Paper #24)

Jade Hind[1], Abir Hussain[1], Dhiya Al-Jumeily[1], Casimiro Aday Curbelo Montañez[1], Basma Abdulaimma[1], Carl Chalmers[1] and Paulo Lisboa[2]
[1]Department of Computer Science, Faculty of Engineering and Technology, Liverpool John Moores University, UK
[2]Department of Applied Mathematics, Faculty of Engineering and Technology, Liverpool John Moores University, UK

The main aim of the paper is to show the application of computational methods together with good practice in quality control, to identify candidate genomic combinations, using Single Nucleotide Polymorphism (SNP) data that suggest a significant relationship exists with the presence of breast cancer. We utilise a dataset from the randomised controlled trial project Discovery, Biology, and Risk of Inherited Variants in Breast Cancer; a cohort consisting of 13,477 case-control female subjects of European population after Quality Control is performed. In this paper, we outline and provide results for use of the shift regularisation methodology that presents a novel combination of variants that indicate high risk in a small cohort of the population.

Session F
Engineering Bio-Interfaces and Rudimentary Cells as a Way to Develop Synthetic Biology

Thursday, 6 September 2018, 16:00-16:20

Drug Delivery Films for Ocular Diseases Treatment (Paper #43)

Mónica Machado[1], Helena Morais[1], Luísa Mendonça[1], Jorge Morgado[1,2], Quirina Ferreira[1]
[1]*Instituto de Telecomunicações, Av. Rovisco Pais, 1049-001, Lisbon, Portugal, quirina.ferreira@lx.it.pt*
[2]*Bioengineering Department, Instituto Superior Técnico, University of Lisbon, Av. Rovisco Pais, 1049-001, Lisbon, Portugal*

This article presents biocompatible and nanostructured layer-by-layer films able to release an ocular drug during a month at precise controlled times with applications in glaucoma treatment. The growth of films was tracked by ultraviolet-visible spectroscopy showing the formation of each layer. The pharmacokinetics of drug was also controlled by UV-Vis spectroscopy during 28 days. The results show that the films are able to release precise amounts of the ocular drug during a month and at precise periods of time.

Thursday, 6 September 2018, 16:20-16:40

Adsorption of Triclosan on sensors based on PAH/PAZO thin-films: the effect of pH (Paper #34)

João Pereira da Silva, Paulo Morgado Zagalo, Gonçalo Magalhães-Mota, Paulo A. Ribeiro and Maria Raposo
CEFITEC, Departamento de Física, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal, {jvp.silva, p.zagalo, g.barreto}@campus.fct.unl.pt, {mfr, pfr}@fct.unl.pt

Triclosan (TCS) is a broad-spectrum antimicrobial, preservative agent widely used in pharmaceuticals and personal care products, considered as a troubling contaminant from the environmental point of view because of its toxicity, bacterial resistance promotion, and estrogenic effects. Under this compliance, the pernicious presence of TCS in the environment is requiring the development of molecular dedicated sensors, which in turn leads to the need to find adequate molecular systems capable of giving rise to a transduction. In this work, in order to investigate the affinity of TCS to common polyelectrolytes in an aqueous environment the adsorption of TCS on thin layer-by-layer (LbL) films of poly[1-[4-(3-carboxy-4-hydroxyphenylazo) benzene sulfonamido]-1,2ethanediyl, sodium salt] (PAZO) and poly (allylamine hydrochloride) (PAH) polyelectrolytes at different values of pH of the solution and changing the outer layer, PAZO and PAH, was investigated. Results demonstrated that the PAH layer is the most indicated to better adsorb TCS molecules. These results are of great importance for the development of TCS sensors based on LbL films, since it indicates that the outer layers of LbL films should be positive electrically charged.

Thursday, 6 September 2018, 16:40-17:00

Effect of epigallocatechin-3-gallate on DMPC oxidation revealed by infrared spectroscopy (Paper #22)

Filipa Pires, Bárbara Rodrigues, Gonçalo Magalhães-Mota, Paulo António Ribeiro and Maria Raposo
CEFITEC, Departamento de Física, Faculdade de Ciências e Tecnologia, *Universidade Nova de Lisboa, 2829-516 Caparica, Portugal*, *af.pires@campus.fct.unl.pt, mfr@fct.unl.pt*

The daily exposure of skin cells to the sun increases the rate of production of free radicals, which threatens the healthy appearance of skin and, even more worrying, damages the structural integrity of tissues and DNA, causing inflammation and carcinogenesis. This work demonstrates the feasibility of using natural agents, in particular tea catechins, in protecting lipidic membranes from oxidative stress-induced by UV radiation exposure. For that purpose, thin cast films prepared from vesicular suspensions of dimyristoylphosphatidylcholine (DMPC) and dimyristoylphosphatidylcholine + (-)-epigallocatechin-3-gallate (DMPC+EGCG) were deposited onto calcium fluoride supports and irradiated with 254 nm UV radiation. The molecular damage after irradiation with UV light was analysed by infrared (IR) together with 2D correlation spectroscopies. Results revealed that the DMPC phospholipid polar moiety is the most vulnerable and sensitive structural target of UV radiation. To check if the presence of the EGCG molecules is protecting the lipids, the principal components analysis (PCA) mathematical method was applied, allowing to conclude that EGCG slows down the cascade of the oxidant-events in the lipid, thus protecting the polar moiety of the lipid.

Thursday, 6 September 2018, 17:00-17:20

Effect of EGCG on the DNA in presence of UV Radiation (Paper #29)

Thais Pivetta, Filipa Pires and Maria Raposo
CEFITEC, Departamento de Física, Faculdade de Ciências e Tecnologia, *Universidade Nova de Lisboa, 2829-516 Caparica, Portugal*, t.pivetta@campus.fct.unl.pt, af.pires@campus.fct.unl.pt, mfr@fct.unl.pt

The exposure to ultraviolet (UV) radiation is clearly a current concern since it damages the deoxyribonucleic acid (DNA) and increases the likelihood of developing skin cancer. On the other hand, green tea compounds such as (-)epigallocatechin-3-gallate (EGCG) present several biological properties and, are well-known for its antioxidant activity. The aim of this work is evaluate the effect of the UV radiation on DNA in presence of EGCG molecules. Results of the evolution of the UV-visible spectra with the UV irradiation suggest that EGCG act like an intercalant molecule and a micromolar concentration of EGCG is effective to induce a strong degradation on the DNA pyrimidines bases under UV radiation. This achievement can lead to a novel class of non-binding safe molecules capable of affinity interaction with the DNA as intercalant molecule which can be used as anti-tumor drugs.

Thursday, 6 September 2018, 17:20-17:40

Non-Thermal Atmospheric pressure Plasmas: Generation, Sources and Applications (Paper #31)

Sara Pereira, Érica Pinto, Paulo A. Ribeiro and Susana Sério
CEFITEC, Departamento de Física, Faculdade de Ciências e Tecnologia, *Universidade Nova de Lisboa, 2829-516 Caparica, Portugal*

Non-thermal atmospheric pressure plasmas, also known as Cold Atmospheric Plasmas (CAPs) are emerging as a potential alternative for the treatment of cancer since they can be generated at atmospheric conditions and their low temperatures allow the interaction with living tissue

without thermal damage. This article will focus on the study of the interaction between plasma and a non-cancerous cell line, particularly, VERO cells. Some in-vitro experiments were performed with a custom-made device in order to better understand how CAPs affect non-cancerous cells. It was also studied how factors such as the distance from the device (gap), the duration and the type of treatment, direct or indirect, affect cell viability after treatments. The obtained results reveal the importance of the determination of the optimal relation between gap and treatment time, since small variations in each one of them can lead to different results in the cell viability.

Thursday, 6 September 2018, 17:40-18:00

Detection of Triclosan Dioxins After UV Irradiation – A Preliminar Study (Paper #54)

Gonçalo Magalhães-Mota, Filipa Pires, Paulo António Ribeiro and Maria Raposo
CEFITEC, Departamento de Física, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal, {g.barreto, af.pires}@campus.fct.unl.pt, {pfr, mfr}@fct.unl.pt

Triclosan (TCS) by itself represents a major health and environmental problem. Also concerning are its photoproducts, various dioxins, which are even more dangerous, creating a need and opportunity to develop dedicated sensors to detect their presence in water. By treating featured data through principal component analysis (PCA), the foot-print of the dangerous TCS products after irradiation can be clearly outlined. This result allow us to conclude that a TCS sensor device based on electronic tongue concept can be envisaged.

Thursday, 6 September 2018, 18:00-18:20

Enhanced degradation of laser-irradiated DNA using Gold Nanoparticles (Paper #47)

Telma Marques[1,2], Sam Eden[2], Małgorzata A. Śmiałek[2,3], Ilko Bald[4,5], Maria Raposo[1] and Nigel J. Mason[2]
[1]*CEFITEC, Departamento de Física, Faculdade de Ciências e Tecnologia, FCT, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal, telma.marques@open.ac.uk, mfr@fct.unl.pt*
[2]*School of Physical Sciences, The Open University, Milton Keynes, United Kingdom, sam.eden@open.ac.uk, nigel.mason@open.ac.uk*
[3]*Department of Control and Power Engineering, Faculty of Ocean Engineering and Ship Technology, Gdansk University of Technology, Gdansk, Poland, smialek@pg.edu.pl*
[4]*Institute of Chemistry − Physical Chemistry, University of Potsdam, Potsdam, Germany.*
[5]*BAM Federal Institute for Materials Research and Testing, Berlin, Germany, bald@uni-potsdam.de*

An long-standing goal in the development of cancer treatments is to use visible radiation in radiotherapy. Gold nanoparticles (AuNPs) have been brought to the forefront of medical research due to their optical characteristics as well as their excellent biocompatibility. In this work DNA aqueous samples with and without AuNPs were irradiated with 532 nm laser light for different periods of time. The results demonstrate that AuNPs strongly enhance the damage of the nucleobase moieties in salmon sperm DNA. These findings suggests that there is great potential to use NPs as DNA photosensitizers in future treatments of diseases such as melanomas and other skin cancers, but further studies should be carried out to demonstrate and confirm the viability of using NPs in vivo.

Thursday, 6 September 2018, 18:20-18:40

Real-time adsorption of a Co-octaethylporphyrin monitorized with scanning tunneling microscopy (Paper #42)

Catarina L. Delfino[1], Jorge Morgado[1,2], Quirina Ferreira[1]
[1]Instituto de Telecomunicac¸ ˜oes, Av. Rovisco Pais, 1049-001, Lisbon, Portugal, quirina.ferreira@lx.it.pt
[2]Bioengineering Department, Instituto Superior T´ecnico, University of Lisbon, Av. Rovisco Pais, 1049-001, Lisbon, Portugal

This article shows the formation of a Co-octhaethylporphyrin (CoOEP) monolayer on graphite in real-time using scanning tunneling microscopy (STM) at solid/liquid interface. High-resolution images of the molecules were obtained showing the adsorption dynamics as function of time.

Session G
Plenary Lecture

Friday, 7 September 2018, 9:00-9:40

Answering questions in biology and medicine by making inferences on networks

Alberto Paccanaro
University of London, United Kingdom

An important idea that has emerged recently is that a cell can be viewed as a complex network of interrelating proteins, nucleic acids and other bio-molecules. At the same time, data generated by large-scale experiments often have a natural representation as networks such as protein-protein interaction networks, genetic interaction networks, co-expression networks. From a computational point of view, a central objective for systems biology and medicine is therefore the development of methods for making inferences and discovering structure in biological networks possibly using data which are also in the form of networks. In this talk, I'll present novel computational methods for solving biological problems which can all be phrased in terms of inference and structure discovery in large scale networks. These methods are based and extend recent developments in the areas of machine learning (particularly semi-supervised learning and matrix factorization), graph theory and network science. I will show how these computational techniques can provide effective solutions for: 1) quantifying similarity between heritable diseases at molecular level using exclusively disease phenotype information; 2) disease gene prediction; 3) drug side-effect prediction.

Session H
Soft Computing Methods for Characterizing Diseases from Omics Data

Friday, 7 September 2018, 9:40-10:00

Identification of Key miRNAs in regulation of PPI Networks (Paper #52)

Antonino Fiannaca[1], Massimo La Rosa[1], Laura La Paglia[1], Alfonso Urso[1], Riccardo Rizzo[1] and Giosue' Lo Bosco[2]
[1]CNR-ICAR, National Research Council of Italy, Via Ugo La Malfa, 153, Palermo,{antonio.fiannaca, massimo.larosa, laura.lapaglia, alfonso.urso, riccardo.rizzo}@icar.cnr.it
[2]Dipartimento di Matematica e Informatica, UNIPA, Universit´a degli Studi di Palermo, Palermo, Italy, giosue.lobosco@unipa.it

In this paper, we explore the interaction among miRNA and deregulated proteins in some pathologies. Assuming that miRNA can influence mRNA and consequently the proteins regulation, we explore this connection by using an interaction matrix derived from miRNA-target data and PPI network interactions. From this interaction matrix and the set of deregulated proteins, we search for the miRNA subset that influences the deregulated proteins with a minimum impact on the not deregulated ones. This regulation problem can be formulated as a complex optimization problem. In this paper, we have tried to solve it by using the Genetic Algorithm Heuristic. As the main result, we have found a set of miRNA that is known to be involved in the disease development.

Friday, 7 September 2018, 10:00-10:20

Recurrent Deep Neural Networks for Nucleosome Classification Networks (Paper #45)

Domenico Amato, Mattia Di Gangi, Giosue' Lo Bosco and Riccardo Rizzo
[1] Dipartmento di Matematica e Informatica - Universit`a degli studi di Palermo
Via Archirafi 34, 90123 Palermo, Italy, domen.amato91@gmail.com, giosue.lobosco@unipa.it
[2] Fondazione Bruno Kessler, Via Sommarive, 18, 38123 Trento, Italy,digangi@fbk.eu
[3] ICAR-CNR - National Research Council of Italy, Via Ugo La Malfa 153, 90146 Palermo, Italy, riccardo.rizzo@icar.cnr.it

Nucleosomes are DNA-histone complex, each wrapping about 150 pairs of double-stranded eukaryote DNA. Several biological studies have shown that the nucleosome positioning influences the regulation of cell type-specific gene activities. In addition, bioinformatic studies have shown proof of sequence specificity in the DNA fragment wrapped into nucleosomes. The main consequence has been the adoption of sequence features representation for the automatic identification of nucleosomes on a genomic scale. In this work, we propose a recurrent deep neural network for nucleosome classification. The proposed architecture stacks convolutional and Long Shortterm Memories layers to automatically extract features from short and long-range dependencies in a sequence. The adoption of this network allows avoiding the feature extraction and selection steps while improving the classification performances. Results have been computed on eight data sets of three different organisms, from Yeast to Human.

Friday, 7 September 2018, 10:20-10:40

Compressive Sensing and Hierarchical Clustering for Microarray Data with Missing Values (Paper #49)

Angelo Ciaramella, Davide Nardone and Antonino Staiano
Dept. of Science and Technology, University of Naples "Parthenope", Isola C4, Centro Direzionale, I-80143, Napoli (NA), Italy,
{angelo.ciaramella,antonino.staiano}@uniparthenope.it;davide.nardone@studenti.uniparthenope.it.

In gene expression microarray measurements, usually, multiple missing expression values are generated, and the proper handling of missing values is a critical task. To address the issue, in this paper a novel methodology, based on compressive sensing mechanism, is proposed for analyzing gene expression data on the basis of topological characteristics of gene expression time series. The approach conceives, when data are recovered, their processing through a non-linear PCA for dimensional reduction and an Hierarchical Clustering Algorithm for agglomeration and visualization. Experiments have been performed on the Yeast Saccharomyces cerevisiae dataset by considering different percentages of information loss. The approach highlights robust performance when high percentage of loss of information occur and when few sampling data are available.

Session I
Computational Methods for Neuroimaging Analysis

Friday, 7 September 2018, 11:00-11:20

Perception-action circuits for word learning and semantic grounding: a neurocomputational model and neuroimaging study (Paper #16)

Max Garagnani[1,2], Evgeniya Kirilina[3,4] and Friedemann Pulvermüller[1,5,6]
1 Department of Computing, Goldsmiths, University of London, SE14 6NW, London – United Kingdom
2 Brain Language Laboratory, Department of Philosophy and Humanities, Freie Universität Berlin Habelschwerdter Allee 45, 14195 Berlin – Germany
3 Neurocomputational Neuroimaging Unit, Freie Universität Berlin, Habelschwerdter Allee 45, 14195 Berlin – Germany
4 Department of Neurophysics, Max-Plank Institute for Cognitive and Brain Sciences, Stephanstraße 1a, Leipzig, Germany
5 Berlin School of Mind and Brain, Humboldt Universität zu Berlin, Luisenstr. 56, 10099 Berlin, Germany
6 Einstein Center for Neurosciences Berlin, Charitéplatz 1, 10117 Berlin – Germany

A neurocomputational architecture of the left-hemispheric areas of the brain is presented which was used to simulate and explain neural correlates of word learning and semantic grounding. The model's main distinguishing features are that (i) it replicates connectivity and anatomical structure of the relevant brain areas, and (ii) it implements only functional mechanisms reflecting known cellular- and synaptic-level properties of the cerebral cortex. Stimulation of the "sensorimotor" model areas (mimicking early stages of word acquisition) leads to the spontaneous formation of cell assemblies (CAs), network correlates of memory traces for meaningful words. Preliminary results of a recent functional Magnetic Resonance Imaging study confirm the model's predictions, and, for the first time, localise the neural correlates of semantic grounding of novel spoken items in primary visual cortex. Taken together, these results provide strong support for perceptual accounts of word meaning acquisition in the brain, and point to a unifying theory of cognition based on action-perception circuits whose emergence, dynamics and interactions are grounded in known neuroanatomy and neurobiological learning mechanisms.

Friday, 7 September 2018, 11:20-11:40

Automatic discrimination of auditory stimuli perceived by the human brain (Paper #1)

Angela Serra[1], Antonio Della Pietra[1], Marcus Herdener[2], Roberto Tagliaferri[1] and Fabrizio Esposito[3]
[1] NeuRoNeLab, DISA-MIS University of Salerno, Fisciano (Sa), 84084, Italy
[2] Department of Psychiatry, Psychotherapy and Psychosomatics,University Hospital for Psychiatry Zurich, Lenggstr. 31, 8032 Zurich, Switzerland.
[3] Department of Medicine, Surgery and Dentistry "Scuola Medica Salernitana", University of Salerno, Via S. Allende, 84081,Baronissi, Salerno, Italy

Humans are able to perceive small difference of sound frequency but it is still unknown how the difference in frequency information is represented at the level of the primary sensory cortex. Indeed, analysis of fMRI imaging identified tonotopic maps through the auditory pathways to the primary sensory cortex. These maps are unfortunately too coarse to show ultra-fine discrimination. Then, the hypothesis is that these small frequency differences are recognised thanks to the information coming from a large set of auditory neurons. To investigate this possibility, a multi-voxel pattern discriminating analysis of the response of BOLD-fMRI in the bilateral auditory cortex to tonal stimuli with different shift in frequency was performed. Our results suggest that small shift in the frequency are easily classified compared with big shift and that multiple areas of the auditory cortex are involved in the tone recognition.

Friday, 7 September 2018, 11:40-12:00

Decoding Semantic Word Categories from Electro and Magnetoencephalography data (Paper #23)

Simeon Spasov[1], Olaf Hauk[2] and Seyedeh-Rezvan Farahibozorg[3]
[1]University of Cambridge, Department of Computer Science and Technology, William Gates Building, 15 J J, Thomson Ave, Cambridge CB3 0FD, United Kingdom, ses88@cam.ac.uk
[2]University of Cambridge, MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom, Olaf.Hauk@mrc-cbu.cam.ac.uk
[3]University of Cambridge, MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom, Rezvan.Farahibozorg@mrc-cbu.cam.ac.uk

How the brain processes language is not completely understood. Alternative hypotheses exist which attempt to explain the phenomenon. In order to provide a statistical analysis on the latency of deciphering semantic categories in the human brain, we used Electroencephalography (EEG) and Magnetoencephalography (MEG) data from 17 subjects, reading abstract or concrete words. We have applied support vector machine classification to differentiate between the brain states stemming from processing the two word categories. Our study suggests that the latency from semantic processing is in the ~80-250ms range with EEG data giving the earliest indication. We found that brain state separability was most easily discernible from MEG sensors and reaches a maximum at ~400-600ms. We hypothesize the early effects reflect semantic information retrieval, while the later effect may reflect mental imagery or decision-related processes.

Friday, 7 September 2018, 12:00-12:20

Statistical validation of State-Space Granger Causality with Time- Reverse Surrogate and its boundaries in fMRI data (Paper #36)

Tiago Fernandes[1], João Pereira[1], Bruno Direito[1,2], Alexandre Sayal[1,3] and Miguel Castelo-Branco[1,2,3]

[1]*Coimbra Institute for Biomedical Imaging and Translational Research, University of Coimbra, CIBIT Edifício do ICNAS, Polo 3, Azinhaga de Santa Comba, 3000-548 Coimbra, tiagofernandes@icnas.uc.pt*
[2]*Faculty of Medicine of the University of Coimbra, University of Coimbra, R. Larga, 3004-504 Coimbra*
[3] *ICNAS – Instituto de Ciências Nucleares Aplicadas à Saúde University of Coimbra, Pólo das Ciências da Saúde Azinhaga de Santa Comba, 3000-548 Coimbra*

Mapping effective connectivity in the human brain represents a major step for the understanding of the human brain connectome. More specifically, Granger Causality (GC) has been widely applied in functional magnetic resonance imaging (fMRI) studies: GC is a data-driven method that assesses the existence of directed influence between regions of interests (ROIs), based on the temporal precedence of information. We investigated a new framework, combining state-space (SS-GC) with statistical validation using time-reverse surrogates (TRS). We investigated the performance in simulated datasets, exploring different model parameters such as the number of samples, sampling frequency and signal-to-noise ratio (SNR). We have also explored the framework on an experimental real time fMRI Neurofeedback (NF) dataset. In conclusion, the work supports the notion that the framework is suitable for fMRI data application provided that a set of experimental specifications is met and that the inherent limitations are known.

Friday, 7 September 2018, 12:20-12:40

Neural Models for Brain Networks Connectivity Analysis (Paper #13)

Razvan E. Kusztos, Giovanna Maria Dimitri and Pietro Liò
Computer Laboratory, University of Cambridge, UK
gmd43@cam.ac.uk

Functional MRI (fMRI) attracts huge interest for the machine learning community nowadays. In this work we propose a novel data augmentation procedure through analysing the inherent noise in fMRI. We then use the novel augmented dataset for the classification of subjects by age and gender, showing a significant improvement in the accuracy performance of Recurrent Neural Networks. We test the new data augmentation procedure in the fMRI dataset belonging to one international consortium of neuroimaging data for healthy controls: the Human Connectome Projects (HCP). From the analysis of this dataset, we also show how the differences in acquisition habits and preprocessing pipelines require the development of representation learning tools. In the present paper we apply autoencoder deep learning architectures and we present their uses in resting state fMRI, using the novel data augmentation technique proposed. This research field, appears to be unexpectedly undeveloped so far, and could potentially open new important and interesting directions for future analysis.

Session J
Plenary Lecture

Friday, 7 September 2018, 14:00-14:40

Model selection for temporal biomedical data

Alexandra Carvalho
Universidade de Lisboa, Portugal

Human health care is changing rapidly, pressing the development of machine learning techniques for automatic diagnoses and prognosis, as well as personalized therapies for

individual patients. The emerging availability of temporal data, namely via electronic medical records, is triggering this line of research. One of the main problems is to model the dynamic process underlying the data evolution. We detail how to learn efficiently Markovian data, when the dependencies can be expressed as a dynamic Bayesian network. We follow a score-based approach, and guarantee that the learned model is optimal according to several model selection criteria. Finally, we address the problem of early classification, which is essential in time-sensitive applications, such as personalized therapies.

Session K
Networking biostatistics and bioinformatics

Friday, 7 September 2018, 14:40-15:00

Searching for the Source of Difference: a Graphical Model Approach (Paper #4)

Monica Chiogna, Vera Djordjilovic, Chiara Romualdi and Elisa Salviato
[1]*Department of Biostatistics, University of Oslo, Norway,vera.djordjilovic@medisin.uio.no*
[2]*Department of Statistical Sciences, University of Padova, Italy,monica.chiogna@unipd.it*
[3]*Department of Biology, University of Padova, Italy, chiara.romualdi@unipd.it*
[4]*IFOM, Milan, Italy, elisa.salviato@ifom.eu*

A growing body of evidence shows that, when performing differential analysis, it is highly beneficial to go beyond differences in the level of individual genes, and consider differences in their interactions as well. We propose an original statistical approach which, by studying conditional distributions, allows to identify the set of variables driving the difference between two conditions under study. Our proposal, set within the framework of Gaussian graphical models, has desirable theoretical properties, and low computational cost, it can complement standard methods, and find its place in the routine exploratory analysis of two-sample omics data.

Friday, 7 September 2018, 15:00-15:20

Improving Sensitivity and Specificity Calculation from Myocardial Imaging Data (Paper #56)

Federica Cugnata[1] and Paola M.V. Rancoita[2]
[1]*Vita-Salute San Raffaele University, CUSSB (University Centre of Statistics in the Biomedical Sciences), Via Olgettina 58, 20132 Milano, Italy, email: cugnata.federica@unisr.it*
[2]*Vita-Salute San Raffaele University, CUSSB (University Centre of Statistics in the Biomedical Sciences), Via Olgettina 58, 20132 Milano, Italy, email: rancoita.paolamaria@unisr.it*

The computation of measures of diagnostic test performance with multiple observations per subject needs advanced statistical methods in order to account for the dependency among observations. In myocardial imaging data, this dependency is related to the physical spatial proximity of the observed diagnostic units (segments). The mixed-effects logistic regression or the generalized estimating equation can be used to account for this dependency, but they are usually employed without specifically model the spatial correlation structure of the observations. Only in few studies, the correlation was modeled but based on the 2D-segmental model of the left ventricular myocardium. Nevertheless, this kind of modeling induces the definition of a spatial correlation which does not totally reflect the real setting. In order to

overcome this issue, in this work, we propose to model the correlation structure of the random effects for the segments as a function of the Euclidean distances of the centers of the segments in the 3D-segmental model. Results on simulated data showed that our methods outperformed other alternatives even in case of low spatial correlation.

Friday, 7 September 2018, 15:20-15:40

Longitudinal Studies and Integrative Biology with Heterogeneous data: an approach using rarefaction (Paper #55)

Luca Del Core[1], Eugenio Montini[1], Clelia Di Serio[2] and Andrea Calabria[1]
[1]*San Raffaele Telethon Institute for Gene Therapy (SR-Tiget), IRCCS San Raffaele Scientific Institute, Via Olgettina 58, 20132, Milano, Italy, delcore.luca@hsr.it, montini.eugenio@hsr.it, calabria.andrea@hsr.it*
[2]*University Vita-Salute San Raffaele,University Centre of Statistics in the Biomedical Sciences, Via Olgettina 58, 20132, Milano, Italy, diserio.clelia@unisr.it*

Data heterogeneity and unreliability negatively influence the effectiveness and reproducibility of the results in all fields involving sampling techniques. Data heterogeneity may occur when integrating data from sources under technological upgrades thus affecting the resolution of the measurements. Unreliability or under-representativeness in data may be due to machine/software or human variances/errors, or other unidentifiable external factors. In the era of large and big data, technological evolution, and continuous data integration, scientists are increasingly facing with the problems of (1) identifying and filtering-out unreliable data, and (2) harmonizing samples gauged with different platforms improved over time. This work is aimed at developing a new statistical framework to address both issues, showing results in real case scenarios, with impact and benefits on reproducibility.

Friday, 7 September 2018, 15:40-16:00

A new partially Coupled Piece-Wise linear Regression Model for statistical network Structure Inference (Paper #3)

Mahdi Shafiee Kamalabad and Marco Grzegorczyk
Bernoulli Institute, Groningen University, Nijenborgh 9, 9747AG Groningen, Netherlands, m.shafiee.kamalabad@rug.nl, m.a.grzegorczyk@rug.nl

In this paper, we propose a new non-homogeneous dynamic Bayesian network model with partially sequentially coupled network parameters. The idea is to segment a time series of network data using multiple changepoint processes, and to model the data in each individual segment by linear Bayesian regression models. The conventional uncoupled models infer the network interaction parameters for each segment separately, without any systematic information-sharing among segments. More recently, it was proposed to couple the network interaction parameters sequentially among segments. The idea is to enforce the parameters of any segment to stay similar to those of the previous segment. This coupling mechanism can be disadvantageous, as it enforces coupling and does not feature the option to uncouple. In this paper, we propose a new model that infers for each individual segment whether it should be coupled to (or uncoupled from) the preceding one. Our new partially coupled model can thus be seen as a consensus model between the uncoupled and the (fully) coupled model.

Session L
Computational Intelligence Methods for Bioinformatics and Biostatistics –II

Friday, 7 September 2018, 16:20-16:40

Computational Annotation of Genetic Biomarkers using Topologically Associating Domains (Paper #21)

Carlos Cano[1*], Michela Verbeni[1*], Carmen Navarro[1], Maria S. Benítez-Cantos[1], Antonio González-Aguilar[2], Gema Durán-Ogalla[3], Manuel Benavides[3], Susana Pedrinaci[4], Mercedes López de Hierro-Ruiz[4], Pilar Martínez-Tirado[5], Francisco Ruiz-Cabello[4], José Luis Martín-Ruiz[5], Armando Blanco[1], Paul Lizardi[1,6]

[1] Department of Computer Science and Artificial Intelligence, University of Granada, Spain, ccano@decsai.ugr.es, michelav@decsai.ugr.es,

[2]Instituto de Parasitología y Biomedicina Lopez-Neyra CSIC, Granada, Spain.

[3]Hospital Carlos Haya, Oncología Médica, Málaga, Spain.

[4]Hospital Virgen de las Nieves, Granada, Spain.

[5]Hospital Clínico San Cecilio, Granada, Spain.

[6]GENYO. Centre for Genomics and Oncological Research: Pfizer / University of Granada / Andalusian Regional Government, Granada, Spain, paul.lizardi@me.com.

[*]These authors contributed equally to this work.

The study of the 3D organization of nuclear DNA is attracting increasing interest since it has been shown to have a direct impact in the regulatory machinery of the cell. In particular, topologically associating domains (TADs) are structural units of chromatin regions proven to be highly self-interacting. A TAD can span from hundreds of kilobases to few megabases, thus potentially including a set of different genes together with promoters and regulatory regions. Improvements in sequencing and computational technologies are continuously delivering biomarker signatures for many different diseases. These signatures often involve hundreds or thousands of different genomic loci, thus transforming in a challenge interpreting and identifying the underlying regulatory mechanisms for the target condition. We propose a tool for computational and statistical analysis of biomarker signatures and topologically associating domains, with the aim of determining DNA domains significantly enriched with loci of interest. We believe that this approach eases the interpretation and further study of large sets of biomarker loci. In this paper, we show the potential of this tool in a case study of methylation biomarkers for Lynch Syndrome. However, the proposed tool is of general purpose and can be run on any set of loci of interest to identify enriched DNA domains.

Friday, 7 September 2018, 16:40-17:00

Unravelling breast and prostate common gene signatures by Bayesian network learning (Paper #40)

Carlos João Villa-Brito[1], Marta B. Lopes[2], Alexandra M. Carvalho[3] and Susana Vinga[4]

[1]Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 Lisboa, Portugal, joao.v.brito@tecnico.ulisboa.pt

[2]IDMEC, Instituto de Engenharia Mecânica, Instituto Superior Técnico, Av. Rovisco Pais, 1040-001 Lisboa, Portugal, marta.lopes@tecnico.ulisboa.pt

[3]Instituto de Telecomunicações, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1049-001 Lisboa, Portugal, alexandra.carvalho@tecnico.ulisboa.pt

[4] *INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, R. Alves Redol 9, 1000-029 Lisboa, Portugal, susanavinga@tecnico.ulisboa.pt*

Breast invasive carcinoma (BRCA) and prostate adenocarcinoma (PRAD) are two of the most common types of cancer in women and men, respectively. As hormone-dependent tumours, BRCA and PRAD share considerable underlying biological similarities worth being exploited. The disclosure of gene networks regulating both types of cancers would potentially allow the development of common therapies, greatly contributing to disease management and health economics. A methodology based on Bayesian network learning is proposed to unravel breast and prostate common gene signatures. BRCA and PRAD RNA-Seq data from The Cancer Genome Atlas (TCGA) measured over ~20000 genes were used. A prior dimensionality reduction step based on sparse logistic regression with elastic net penalisation was employed to select a set of relevant genes and provide more interpretable results. The Bayesian networks obtained were validated against information from STRING, a database containing known gene interactions.

Friday, 7 September 2018, 17:00-17:20

Variational inference in probabilistic single-cell RNA-seq models (Paper #39)

Pedro Ferreira[1], Alexandra Carvalho[1,2] and Susana Vinga[1,3]
[1]Instituto Superior Técnico, ULisboa, Av. Rovisco Pais, 1049-001 Lisboa, Portugal pedro.fale@tecnico.ulisboa.pt
[2]Instituto de Telecomunicações, Av. Rovisco Pais, 1049-001 Lisboa, Portugal, alexandra.carvalho@tecnico.ulisboa.pt
[3] INESC-ID, R. Alves Redol 9, 1000-029 Lisboa, Portugal, susanavinga@tecnico.ulisboa.pt

Single-cell sequencing technology holds the promise of unravelling cell heterogeneities hidden in ubiquitous bulk-level analyses. However, limitations of current experimental methods also pose new obstacles that prevent accurate conclusions from being drawn. To overcome this, researchers have developed computational methods which aim at extracting the biological signal of interest from the noisy observations. In this paper we focus on probabilistic models designed for this task. Particularly, we describe how variational inference constitutes a powerful inference mechanism for different sample sizes, and critically review two recent scRNA-seq models which use it.

Friday, 7 September 2018, 17:20-17:40

A new method for getting probability distributions with application (Paper #51)

Bozidar Popovic
University of Montenegro, Faculty of Science and Mathematics, Podgorica, Montenegro

Azzalini's method is used for obtaining skewed distribution by assumption of independence of two random variables. Our aim is generalizing of Azzalini's method by avoiding independence assumption. We will suppose that X1 and X2 are two dependent random variables, and its dependence is modeled using F-G-M copula. By means of this approach we get generalized weighted distributions. These new models have more flexible hazard rate functions that allows wider applicability.

Friday, 7 September 2018, 17:40-18:00

PiXora: Simultaneous Phasing of Multiple Polyploids (Paper #8)

Laxmi Parida and Filippo Utro
Computational Biology Center, IBM T. J. Watson Research, Yorktown Heights, NY, 10598, USA. {parida, futro}@us.ibm.com

We address the problem of phasing polyploids specifically with polyploidy larger than two. We consider the scenario where the input is the genotype of samples along a genic chromosomal segment. In this setting, instead of NGS reads of the segments of a sample, genotype data from multiple individuals is available for simultaneous phasing. For this mathematically interesting problem, with application in plant genomics, we design and test two algorithms under a parsimony model. The first is a linear time greedy algorithm and the second is a carefully crafted algebraic algorithm. We show that both the methods work reasonably well (with accuracy on an average larger than 80%). The former is very time-efficent and the latter improves the accuracy further to about 90%.

Session M
Plenary lecture

Saturday, 8 September 2018, 9:00-9:40

Ethics and our moral in research, let's think about it!

Fernando Luís Ferreira
Universidade Nova de Lisboa, Portugal

As researchers, it is our will is to pursue knowledge, to contribute to society and to open new roads for the Future. Ethics is a theme always present in our minds but probably remains outside the central concerns of researchers while main subjects are developed. Sometimes we come across a formal consent or an ethics statement seen mostly as a bureaucratic task. However, lately with the so called exponential technologies, we find ourselves dueling with a variety of controversial questions resulting from the different branches of artificial intelligence as those applied to self-driving cars' decisions the exposure of privacy and Decision support systems in medicine, etc. Some are arguing that risks become clear and, one of this days, we may face a singularity and, eventually, becoming too late to stop. This short talk aims to rise some questions about present ethical issues aiming at promoting the intervention and discussion among participants at this Conference.

Session N
Computational Intelligence Methods for Bioinformatics and Biostatistics –III

Saturday, 8 September 2018, 9:40-10:00

Predicting the Oncogenic Potential of Gene Fusions using Convolutional Neural Networks (Paper #14)

Marta Lovino, Gianvito Urgese, Enrico Macii, Santa Di Cataldo and Elisa Ficarra

Politecnico di Torino, Dept. of Control and Computer Engineering, Corso Duca Degli Abruzzi 24, Torino, Italy, {marta.lovino, gianvito.urgese, enrico.macii, santa.dicataldo, elisa.ficarra}@polito.it

Predicting the oncogenic potential of a gene fusion transcript is an important and challenging task in the study of cancer development. To this date, the available approaches mostly rely on protein domain analysis to provide a probability score explaining the oncogenic potential of a gene fusion. In this paper, a Convolutional Neural Network model is proposed to discriminate gene fusions into oncogenic or non-oncogenic, exploiting only the protein sequence without protein domain information. Our proposed model obtained accuracy value close to 90% on a dataset of fused sequences.

Saturday, 8 September 2018, 10:00-10:20

Benchmarking Multitask Learning for QSARS (Paper #38)

Ivan Olier[1], Philippa Grace McCabe[1], Disha Agarwal[2] and Sandra Ortega-Martorell[1]
[1]Department of Applied Mathematics, Liverpool John Moores University, Liverpool, UK, I.A.OlierCaparroso@ljmu.ac.uk,P.McCabe@2014.ljmu.ac.uk,S.OrtegaMartorell@ljmu.ac.uk
[2]Department of Information and Communication Technology, Manipal Institute of Technology, Manipal, Karnataka, India. Disha.97.agarwal@gmail.com

Learning Quantitative Structure Activity Relationships (QSARs) is a key step in drug development. Traditionally, learning small QSAR datasets has been challenging for traditional methods. Recently, multitask learning has been applied to QSAR learning. Our aim with this study is to further understand the capabilities of multitask learning for QSAR learning. Therefore, we applied multitask learning to eighty QSAR datasets, and compared its performance with conventional single-task learning. Our results showed that multitask learning works well especially with the smaller datasets.

Saturday, 8 September 2018, 10:20-10:40

Paratope identification by classification of local antibody surface patches enriched with eight physicochemical properties (Paper #10)

Sebastian Daberdaku
Department of Information Engineering – University of Padova, Via Gradenigo 6A, 35131, Padova (PD), Italy, sebastian.daberdaku@unipd.it

This paper presents a method for antibody interface prediction starting from experimentally-solved 3D structures. Because of their ability to specifically bind to a virtually infinite number of antigens, antibodies currently represent the most important category of biopharmaceuticals. The identification of the antigen-binding site (the paratope) could help in understanding the complex mechanisms that govern antibody–antigen interactions and is crucial for antibody computational design. In this work, 3D Zernike descriptors are computed from circular patches of the antibody surface enriched with the physicochemical properties from the HQI8 amino acid index set, and are used as samples for a binary classification problem. An SVM classifier is used to distinguish interface surface patches from non-interface ones. Experimental results show that the presented method outperforms other paratope prediction software.

Saturday, 8 September 2018, 10:40-11:00

Computational Intelligence to explore the link between diabetes and arthrites (Paper #58)

Pietro Lio[1] and Andrea Bracciali[2]
[1]*Cambridge University, UK, pl219@cam.ac.uk*
[2]*Stirling University, UK, abb@cs.stir.ac.uk*

There is more awareness of the importance and diffusion of comorbidies but physicians get very little help from theoreticians. The complex patterns of comorbidities make very difficult to establish the mapping between genotype and phenotype. This work focuses on developing methodologies to evaluate the dynamical insurgence of comorbidity by integrating molecular, cellular and epidemiological data. Type-2 Diabetes (T2D) and rheumatoid arthritis (RA) are considered as model diseases. We first define novel stochastic models for T2D and RA, extending continuous models from literature based on available gene expression and cell networks data. Then, we define a joint model to study the insurgence of mutual direct comorbidities (T2D to RA and RA to T2D). Although the precise molecular and cellular "codynamics" of the two diseases is not clearly understood, we embed plausible reciprocal effects of the diseases on one another in the model. Furthermore, we consider epidemiological data on the co-evolution of diseases at the population level to benchmarking and making more accurate our model. Finally, we validate the model on case studies, which have the ultimate, long-term goal to contribute to the stratification of patient's comorbidity severity. Our work highlights aspects such as considering comorbidity state not as just the additive effect of the contributed morbidities and considers the problem of stratifying the patients on the basis of the severity of the overall medical condition.

Session O
Modeling and Simulation Methods for System Biology and System Medicine

Saturday, 8 September 2018, 11:20-11:40

Global Sensitivity Analysis of Constraint-based Metabolic Models (Paper #26)

Chiara Damiani[1,2], Dario Pescini[2,3] and Marco S. Nobile[1,2]
[1]*University of Milano-Bicocca, Department of Informatics, Systems and Communication, 20126 Milano, Italy, damiani/nobile@disco.unimib.it*
[2]*SYSBIO.IT Centre of Systems Biology, 20126 Milano, Italy*
[3]*University of Milano-Bicocca, Department of Statistics and Quantitative Methods, 20126 Milano, Italy*

In the latter years, detailed genome-wide metabolic models have been proposed, paving the way to thorough investigations of the connection between genotype and phenotype in human cells. Nevertheless, classic modeling and dynamic simulation approaches—based either on differential equations integration, Markov chains or hybrid methods—are still unfeasible on genome-wide models due to the lack of detailed information about kinetic parameters and initial molecular amounts. By relying on a steady state assumption and constraints on extracellular fluxes, constraint-based modeling provides an alternative means—computationally less

expensive than dynamic simulation—for the investigation of genome-wide biochemical models. Still, the predictions provided by constraint-based analysis methods (e.g., flux balance analysis) are strongly dependent on the choice of flux boundaries. To contain possible errors induced by erroneous boundary choices, a rational approach suggests to focus on the pivotal ones. In this work we propose a novel methodology for the automatic identification of the key fluxes in large-scale constraint-based models, exploiting variance-based sensitivity analysis and distributing the computation on massively multi-core architectures. We show a proof-of-concept of our approach on core models of relatively small size (up to 314 reactions and 256 chemical species), highlighting the computational challenges.

Saturday, 8 September 2018, 11:40-12:00

Inhibition of Primed Ebola Virus Glycoprotein by Peptide Compound Conjugated to HIV-1 Tat Peptide through a Virtual Screening Approach Models (Paper #9)

Ahmad Husein Alkaff, Mutiara Saragih, Mochammad Arfin Fardiansyah Nasution and Usman Sumo Friend Tambunan
[1]*Universitas Indonesia, Bioinformatics Research Group, Department of Chemistry, Faculty of Mathematics and Natural Sciences, Kampus UI Depok, 16424, Indonesia, ahmad.husein@sci.ui.ac.id, mutiara.saragih@sci.ui.ac.id*
marfin.f@sci.ui.ac.id, usman@ui.ac.id

A higher prevalence of Ebola hemorrhagic fever is caused by Ebola virus (EBOV). It enters into the host cell through macropinocytosis mechanism. During the entry process, the primed viral glycoprotein (GPcl) interacts with a lysosomal cholesterol transporter, Niemann Pick C1 (NPC1), leading to the fusion of the viral envelope and the host endosomal membrane. Hence, disrupting the interaction between EBOV GPcl and host NPC1 is a promising way to prevent the viral nucleocapsid content entering the cytoplasm. In this study, a virtual screening approach has been used to investigate peptide compounds conjugated to HIV-1 Tat peptide as drug lead candidate inhibiting EBOV GPcl. About 50,261 peptides from NCBI PubChem database, which acts as ligands, were subjected to initial toxicological screening to omit ligands with undesired properties. The remaining ligands underwent a pharmacophore search, rigid docking, and flexible docking simulation to discover ligands with favorable inhibition activities. Calfluxin, SNF 8906, grgesy, phosporamidon, and endothelin (16-21) were five ligands which have lower ΔGbinding value compared to the standard ligand. They were conjugated to HIV-1 Tat peptide to accumulate them inside the endosome. Subsequently, the inhibition activity was reevaluated by the second flexible molecular docking simulation. Only C-Calfluxin showed better affinity and minimal conformational changes in protein-peptide interaction compared to its respective unconjugated ligand. Further analysis through the computational pharmacological test and molecular dynamics simulation need to be performed to assess its properties before going to in vitro and in vivo analysis.

Saturday, 8 September 2018, 12:00-12:20

GPU Powered Parameter Estimation of a large-Scale Kinetic Metabolic Model (Paper #44)

Niccolò Totis[1], Andrea Tangherloni[2], Marco Beccuti[1], Paolo Cazzaniga[3], Marco Nobile[2], Daniela Besozzi[2], Marzio Pennisi[4] and Francesco Pappalardo[5]
[1]*Department of Computer Science, University of Torino, Torino, Italy*

[2]*Department of Informatics, Systems and Communication, University of Milano-Bicocca, Milano, Italy*
[3]*Department of Human and Social Sciences, University of Bergamo, Bergamo, Italy*
[4]*Department of Mathematics and Computer Science, University of Catania, Catania, Italy*
[5]*Department of Drug Sciences, University of Catania, Catania, Italy*

Kinetic modeling is a powerful tool to describe biochemical reaction systems. Considering the biological complexity of metabolic systems, it is often challenging to represent models and to efficiently analyze their dynamics, as in the case of models that explicitly take into account different isoforms of metabolic enzymes. In this case, computationally expensive Parameter Estimation (PE) procedures are typically required since the kinetic characterization of the different isoforms is most of the times unavailable. In this work we tackle these issues with an approach that combines the descriptive power of Stochastic Symmetric Nets—a parametric and compact extension of the Petri Net formalism—with FST-PSO, an efficient optimization method suitable for the PE problem. To execute the large number of simulations required by the PE we exploit LASSIE, a GPU-powered deterministic simulator that offloads the calculations onto the GPU cores. LASSIE achieves around 30x speed-up with respect to CPU to carry out a PE of an intracellular large-scale kinetic model of human metabolism.

Saturday, 8 September 2018, 12:20-12:40

Pharmacophore Modelling, Virtual Screening, and Molecular Docking Simulations of Natural Product Compounds as Potential Inhibitors of Ebola Virus Nucleoprotein (Paper #33)

Mochammad Arfin Fardiansyah Nasution, Ahmad Husein Alkaff, Ilmi Fadhilah Rizki and Usman Sumo Friend Tambunan
Universitas Indonesia, *Bioinformatics Research Group, Department of Chemistry, Faculty of Mathematics and Natural Science, Kampus UI Depok, 16424, Indonesia, marfin.f@sci.ui.ac.id*, *ahmad.husein@sci.ui.ac.id*, *ilmi.fadhilah@sci.ui.ac.id*, *bakri@ui.ac.id*, *usman@ui.ac.id*

Ebola virus (EBOV) prevails as a serious public health issue which infected at least 27,000 people and claimed the life of about 11,000 people in the latest Ebola outbreak in 2014. Although the virus has been known for almost 40 years, currently there is no approved drug for this virus. Hence, the development of new drug candidate for Ebola is required to anticipate the future outbreak that may happen. In this research, about 229,538 natural product (NP) compounds were retrieved and screened using computational approach against EBOV nucleoprotein (NP). In the beginning, All NP compounds were screened throughout computational toxicity and druglikeness prediction tests, followed by pharmacophore-based virtual screening and molecular docking simulation to identify their binding affinity and molecular interaction in the RNA-binding groove of EBOV NP. All of these results were compared to 18β-glycyrrhetinic acid, the standard molecule of EBOV NP. In the end, about five NP compounds were identified to have exciting activities against EBOV NP. Therefore, based on the results of this study, these compounds appeared to have potential inhibition activities against EBOV NP and can be proposed for further in silico and in vitro studies.

Session P
Machine Learning in Health Informatics and Biological Systems

Saturday, 8 September 2018, 14:00-14:20

The Scourge of Anti-microbial Resistance: A Machine learning Ap proach for prescription Patterns Analytics (Paper #48)

Davide Castaldi[1], Ilaria Giordani[2], Antonio Candelieri[2], Francesco Archetti[3] and Roberto Mattina[2]
[1]Consorzio Milano Ricerche, Via Roberto Cozzi 53, 20125, Milan, Italy, castaldi@milanoricerche.it, archetti@milanoricerche.it
[2]University of Milano-Bicocca, Department of Computer Science, Systems and Communication, Viale Sarca 336, U14 Building, 20126, Milan, Italy ilaria.giordani@disco.unimib.it, antonio.candelieri@unimib.it, francesco.archetti@unimib.it
[3]University of Milano, Department of Biomedical, Surgical and Dental Sciences, Via Pascal, 36/38, 20133, Milan, Italy, roberto.mattina@unimi.it

Antimicrobial resistance (AMR) has become a health emergency worldwide: bacteria are mutating and exchanging their genes at an increasing rate, with 5% growth for some bacteria resistant to some classes of antibiotics in some EU countries [1]. Data analytics can give a major contribution to tackle this health and socioeconomic challenge: by means of insights originated from analytical results health authorities can plan informed actions.This study summarizes relevant insights obtained by the analysis of data related to about 500 general practitioners (GPs) and around one million patients, during the period 2011-2016. After specific data cleaning, the sample was reduced to just over 500'000 patients, balanced in terms of gender, age, and therapeutic indication, and related to 140 different types of anti-bacterial agents. We employed a time series clustering approach to first identify the typical prescription patterns and then discriminating between GPs with changing and constant prescription behaviors. More precisely, spherical k-means clustering, based on cosine similarity, was used to obtain clusters consisting of prescription patterns sharing occurrence of peaks and bursts. Finally, it was also possible to identify possible changes in prescription patterns, from one year to the next, for every GP as well as for similar groups of GPs. The approach proposed in this paper provides useful insights to policy makers and health care actors about local trend or individual behaviors enabling the design and implementation of consistent and targeted corrective interventions. The overuse, underuse or misuse of antibiotics molecules results in wastage of relevant resources to challenge infections, contributing to health hazards spread related to AMR.

Saturday, 8 September 2018, 14:20-14:40

Unveiling Gene Expression Histonic Regulative Patterns by Hyperplanes Clustering (Paper #32)

Fabrizio Frasca[1], Matteo Matteucci[1], Marco Morelli[2] and Marco Masseroli[1]
[1]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo Da Vinci 32, 20133 Milan, Italy, [fabrizio.frasca@mail., matteo.matteucci@, marco.masseroli@]polimi.it
[2]Center for Genomic Science of IIT@SEMM, Istituto Italiano di Tecnologia (IIT), 20139 Milan, Italy, morelli.marco@hsr.it

In targeted cancer therapy, great relevance is assumed by data-driven investigations on the fundamental mechanisms by which epigenetic modifications cooperate to regulate the

transcriptional status of genes. At the high resolution level of genome-wide studies, only general, mean regulative motifs are drawn, with possible multi-functional co-regulative roles remaining concealed. In order to retrieve sharper and more reliable regulative patterns, in this work we propose the application of K-plane regression to partition the set of protein coding genes into clusters with shared regulative mechanisms. Completely data-driven, the approach has computed clusters of genes significantly better fitted by specific linear models than by single regression, and characterized by distinct histonic input patterns and mean measured expression values.

Saturday, 8 September 2018, 14:40-15:00

Designing and Evaluating Deep Learning Methods for Cancer Classification on Gene Expression Data (Paper #46)

Arif Canakoglu, Luca Nanni, Artur Sokolovsky and Stefano Ceri
Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Italy
first.last@polimi.it

Gene expression levels, measuring the transcription activity, are widely used to predict abnormal gene activities, in particular for distinguishing between normal and tumor cells. This problem has been addressed by a variety of machine learning methods; more recently, this problem has been approached using deep learning methods, but they typically failed in meeting the same performance as machine learning. In this paper, we show specific deep learning methods that can achieve similar performance as the best machine learning methods.

Saturday, 8 September 2018, 15:00-15:20

A Semi-supervised Learning Approach to the Identification and Prioritization of Genes Associated with Autism Spectrum Disorder (Paper #25)

Hugo Martiniano[1,2], Muhammad Asif[1,2], Astrid Vicente[1,2] and Luís Correia[1].
[1]University of Lisbon, Faculty of Sciences, BioISI - Biosystems & Integrative Sciences Institute, Campo Grande, 1749-016 Lisboa, Portugal
[2]Instituto Nacional de Sa´ude Doutor Ricardo Jorge, Avenida Padre Cruz, 1649-016 Lisboa, Portugal

Autism Spectrum Disorder (ASD) is an etiologically and clinically heterogeneous set of neurodevelopmental disorders with more than 800 associated risk genes. This heterogeneity, coupled with the low penetrance of most ASD-associated mutations presents a challenge in identifying the genetic determinants of ASD. We developed a machine learning semi-supervised gene scoring and classification method based on network propagation using a variant of the random walk with restart algorithm to identify and rank genes according to their association to know ASD-related genes. The method combines information from protein-protein interactions and positive (disease-related) and negative (disease-unrelated) genes. Our results indicate that the proposed method can classify held-out known disease genes in a cross-validation setting with good performance (area under the receiver operating curve ~ 0.85, area under the precision-recall curve ~ 0.8 and Matthews correlation coefficient 0.57). We found a set of top-ranking novel candidate genes identified by the method to be significantly enriched for pathways related to synaptic transmission and ion transport and specific neurotransmitter

associated pathways previously shown to be associated with ASD.

Saturday, 8 September 2018, 15:20-15:40

Identification of real and imaginary movements in EEG using Machine Learning models (Paper #50)

Joana Moreira[1], Mariana Moreira[1], Nuno Pombo[2] and Nuno Garcia[1]
[1]Health Science Department, Universidade da Beira Interior, Covilhã, Portugal
[2]Instituto de Telecomunicações, Universidade da Beira Interior, Covilhã, Portugal, and Universidade Lusófona de Humanidades e Tecnologias, Lisbon, Portugal

The neural activity of the brain may be observed by means of an electroen-cephalogram (EEG) whose analysis and/or interpretation may lead to the recognition of human activities and behaviors. In line with this, study aims to differentiate real movements from imagined ones, through EEG signals. The signals were treated and three classification models were used to classify these two events, one representing real movement and the oth-er imagined movement of opening and closing the left fist. The preliminary results, ob-tained based on 15 volunteers, revealed that the Logistic Regression was the best classifier into the proposed model with accuracy rates varying from 36.8% to 90%. Finally, comple-mentary studies should be addressed in order to optimize not only the accuracy but also to assure uniform accuracy among the different volunteers.

Session Q
Computational Intelligence Methods for Bioinformatics and Biostatistics –IV

Saturday, 8 September 2018, 16:00-16:20

The ciliates as complex eukaryotic single cell models for studying molecular communication: an insight on the evolution of signaling molecules also involved in human brain metabolism (Paper #57)

Alessio Mancini[1], Claudio Angione[2], Pietro Liò[3] and Sandra Pucciarelli[1]
[1]School of Biosciences and Veterinary Medicine, University of Camerino, Italy
[2]School of Computing, Teesside University, UK
[3]Computer Laboratory, University of Cambridge, UK

In this paper, we propose ciliates as models for studying molecular communication via genomescale models (GEMs) and simulation tools. Many cells types in eukaryotic organisms possess cilia, that may be either motile or non-motile. Non-motile cilia, also known as primary cilia, are sensory organelles involved in at least four different signaling pathways acting on signal transduction. Pioneer study using the ciliate Paramecium demonstrated the avoiding reaction in these cells stimulating the reversion of the direction of the swimming by cilia when in the presence of obstacle. Even though no direct evidence on the presence of primary (sensory) cilia in ciliates have been reported, these organisms appear promising for studying this topic. Ciliates can communicates with pheromones. In the ciliate Blepharisma, pheromones are represented by a tryptophan-related molecule, like serotonin, a biochemical messenger in the human nervous system and other organism. Tetrahymena whole-genome sequencing also revealed that this ciliate has genes encoding tryptophan hydroxylase (TPH), which is a rate-

limiting enzyme for serotonin biosynthesis. Serotonin synthesis and metabolism affects timing of conjugation and cilia formation in Tetrahymena. In the extended version we present the phylogenetic analysis of multiomic information (protein sequences, protein structure and gene expression) of serotonin pathways in the ciliates and human. The integration of signaling and metabolic pathways of Tetrahymena may provide interesting clues for the evolution of serotonin pathway. Serotonin role in ciliates can offer clues on the evolution of this signaling molecule in the human brain.

Saturday, 8 September 2018, 16:20-16:40

In Silico ADME and Toxicity Prediction Using N-grams for Chemical Fingerprinting (Paper #61)

Leif E. Peterson and Timothy C. Thompson
[1]*Department of HealthCare Policy and Research, Weill Cornell Medical College, Cornell University, New York City, New York 10065 USA, Center for Biostatics, Institute for Academic Medicine, Houston Methodist Research Institute, 6565 Fannin Street, Houston, Texas 77030 USA, e-mail: lepeterson@houstonmethodist.org*
[2]*Prostate Cancer Basic Science Research, Department of Genitourinary Medical Oncology – Research, Division of Cancer Medicine, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Houston Texas 77030 USA, e-mail: timthomp@mdanderson.org*

There is growing worldwide demand for rapid safety assessment of novel drugs and chemicals by industry and regulatory agencies. In silico computational methods can accelerate drug safety predictions for lead compounds as an adjunct to animal studies, can reduce costs, and minimize late-stage drug failure through early toxicity prediction. This investigation compared ADME and toxicity prediction of known test compounds for human intestinal absorption, blood-brain barrier penetration, cytochrome P-450 enzyme inhibition, fathead minnow toxicity, Tetrahymena pyriformis toxicity, and honey bee toxicity using several class prediction methods. Class prediction input features based on FP2 chemical fingerprints and N-gram text analysis of SMILES strings were compared. Results indicate that use of N-gram analysis on chemical SMILES strings can reveal predictive accuracy similar to results based on FP2 chemical fingerprints. Additional research is required to evaluate the utility and computational time-complexity for using N-grams for drug ADME and toxicity prediction in the presence of physical parameters, pharmacophores, and other molecular features.

Saturday, 8 September 2018, 16:40-17:00

A Graphical Tool for the Exploration and Visual Analysis of Biomolecular Networks (Paper #35)

Cheick Tidiane Ba[1], Elena Casiraghi[1], Marco Frasca[1], Jessica Gliozzo[2], Giuliano Grossi[1], Marco Mesiti[1], Marco Notaro[1], Paolo Perlasca[1], Alessandro Petrini[1], Matteo Re[1] and Giorgio Valentini[1]
[1]*Department of Computer Science, Università degli Studi di Milano, Via Comelico 39, 20135 Milano*
[2]*Fondazione IRCCS Cá Granda - Ospedale Maggiore Policlinico,*
Universit`a degli Studi di Milano

Many interactions among bio-molecular entities, e.g. genes, proteins, metabolites, can be easily represented by means of property graphs, i.e. graphs that are annotated both on the vertices (e.g. entity identifier, Gene Ontology or Human Phenotype Ontology terms) and on the edges (the

strength of the relationship, the evidence of the source from which the weight has been taken, etc.). These graphs contain a relevant information that can be exploited for conducting different kinds of analysis, such as automatic function prediction [Valentini et al., 2016], disease gene prioritization [Schubach et al., 2017], drug repositioning [Re et al., 2013]. However, the number and size of the networks are becoming quite large and there is the need of tools that allow the biologists to manage the networks, graphically explore their structures, and organize the visualization and analysis of a specific graph according to different perspectives. In this paper we introduce the web service that we have developed for the visual analysis of biomolecular networks. Specifically we show the different functionalities for exploring large networks (that do not fit in the current canvas) starting from a specific vertex, for changing the view perspective of the network, and for navigating the network and thus identifying new relationships. The proposed system extends the functionalities of off-the-shelf graphical visualization tools (e.g. GraphViz and Gene-Mania). In particular, the proposed system limits the production of big cloud of points and allows further customized visualizations of the network by introducing their vertex centric exploration.

Saturday, 8 September 2018, 17:00-17:20

An innovative approach to integrate proteomics and metabolomics data in severe septic shock (Paper #17)

Manuela Ferrario[1], Alice Cambiaghi[1], Eliandre de Olivera[2] and Roberta Pastorelli[3]
[1]Politecnico di Milano, DEIB, P.zza Leonardo da Vinci 32, Milano, manuela.ferrario@polimi.it
[2]Plataforma de Protemica, Parc Cientfic de Barcelona, c/ Baldiri Reixac, 10-12, Barcelona, eoliveira@pcb.ub.es
[3]IRCCS Istituto di Ricerche Farmacologiche Mario Negri, Via La Masa 19, Milano, roberta.pastorelli@marionegri.it

In this work, we examined plasma metabolome and proteome in patients with severe septic shock enrolled in the multicenter ALBIOS study. The objective was to identify changes in the levels of metabolites involved in septic shock progression and to integrate this information with the variation occurring in proteins. Mass spectrometry based targeted metabolomics and untargeted proteomics allowed us to quantify absolute metabolites concentration and relative proteins abundance. We computed the ratio D7/D1 to take into account their variation from day 1 (D1) to day 7 (D7) after shock diagnosis. Patients were divided into two groups according to 28-day mortality. Two different elastic net logistic regression models were built: one on metabolites only and one on metabolites and proteins. Linear discriminant analysis and Partial least squares Discriminant Analysis were also implemented. All the obtained models correctly classified the observations in the testing set. By looking at the variable importance (VIP) and the selected features, the integration of metabolomics with proteomics data showed the importance of circulating lipids and coagulation cascade in septic shock progression, thus capturing a further layer of biological information complementary to metabolomics information.

Saturday, 8 September 2018, 17:20-17:40

Analysis of Extremely Obese Individuals Using Deep Learning Stacked Autoencoders and Genome-Wide Genetic Data (Paper #30)

Casimiro Aday Curbelo Montañez, Paul Fergus, Carl Chalmers and Jade Hind
[1]Department of Computer Science, Faculty of Engineering and Technology, Liverpool John Moores University, UK

Genetic predisposition has been identified as one of the components contributing to the obesity epidemic in modern societies. The aetiology of polygenic obesity is multifactorial, which indicates that lifestyle and environmental factors may influence multiples genes to aggravate this disorder. Several low-risk single nucleotide polymorphisms (SNPs) have been associated with BMI. However, identified loci only explain a small proportion of the variation observed for this phenotype. The linear nature of genome wide association studies (GWAS) used to identify associations between genetic variants and the phenotype have had limited success in explaining the heritability variation of BMI and shown low predictive capacity in classification studies. GWAS ignores the epistatic interactions that less significant variants have on the phenotypic outcome. In this paper we utilise a novel deep learning-based methodology to reduce the high dimensional space in GWAS and find epistatic interactions between SNPs for classification purposes. SNPs were filtered based on the effects associations have with BMI. The classifier is fine-tuned to classify extremely obese and non-obese individuals. The performance of classifications using progressively smaller compressed layers was compared and the results reported. The best results were obtained with 2,000 compressed units, AUC=0.97497. Using 50 compressed units it was possible to achieve AUC=0.85178.

Saturday, 8 September 2018, 17:40-18:00

Smartwatch Based Fall Detection System for Elders' Surveillance (Paper #41)

Fernando Luis-Ferreira, Daniel Rodrigues, João Sarraipa and Ricardo Goncalves
CTS, UNINOVA, Dep.º de Eng.ª Electrotécnica Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

Ageing comes with associated risks for elders' health and safety beyond pathologies or other clinical states. Stumble and fall are considered one of the major causes of mortality and morbidity in the elderly. Those risks are present at home and outside where unbalance and cognitive limitations, sometimes associated with dementia, may expose elder citizens to those situations when emergency rescue is crucial. Technology can provide systems that help tackling those problems by monitoring activity and determine if a risky situation occurred. The hereby presented research work aims at detecting stumble and fall situations that could be associated to critical health events. The proposed solution uses accelerometers, gyroscope and heart rate measurements, all sensors in a commercial smartwatch. The analysis of the different parameters allows an identification of risky situations associated with fall and triggers a request for emergency when needed.

PARTICIPANTS

First Name	Last Name	Country	e-mail	Page
Ahmad Husein	Alkaff	Indonesia	ahmad.husein@sci.ui.ac.id	36,37
Muhammad	Asif	Portugal	muhasif123@gmail.com	39
Tiago	Azevedo	United Kingdom	tiago.azevedo@cst.cam.ac.uk	
Marco	Benetollo	Italy	marco.ben92@hotmail.it	
Stefano	Beretta	Italy	stefano.beretta@disco.unimib.it	18,19
Ahsan	Bilal	Spain	ahsan924@gmail.com	16
Vincenzo	Bonnici	Italy	Vincenzo.bonnici@univr.it	18
Chiara	Brombin	Italy	brombin.chiara@hsr.it	
Andrea	Calabria	Italy	calabria.andrea@hsr.it	30
Simone	Caligola	Italy	simone.caligola@univr.it	18
Arif	Canakoglu	Italy	arif.canakoglu@polimi.it	39
Carlos	Cano	Spain	ccano@decsai.ugr.es	31
Alexandra	Carvalho	Portugal	asmcarvalho@gmail.com	
Monica	Chiogna	Italy	monica@stat.unipd.it	29
Angelo	Ciaramella	Italy	angelo.ciaramella@uniparthenope.it	25
Federica	Cugnata	Italia	cugnata.federica@hsr.it	29
Casimiro Aday	Curbelo	UK	c.a.curbelomontanez@ljmu.ac.uk	20,43
Sebastian	Daberdaku	Italy	sebastian.daberdaku@unipd.it	35
Vanessa	D'Amario	Italia	vanessa.damario@dibris.unige.it	17
Catarina	Delfino	Portugal	catldelfino@gmail.com	24
Clelia	Di Serio	Italia	diserio.clelia@unisr.it	30
Giovanna Maria	Dimitri	United Kingdom	gmd43@cam.ac.uk	28
Vera	Djordjilović	Norway	vera.djordjilovic@medisin.uio.no	29
Tiago	Fernandes	Portugal	tiagopaisb@hotmail.com	27
Manuela	Ferrario	Italy	manuela.ferrario@polimi.it	42
Pedro	Ferreira	Portugal	pedro.fale@tecnico.ulisboa.pt	32
Quirina	Ferreira	Portugal	quirinatf@gmail.com	21,24
Fabrizio	Frasca	Italy	fabrizio.frasca@mail.polimi.it	17,38,41
Ilaria	Giordani	Italy	giordani@disco.unimib.it	38
Marco	Grzegorczyk	Netherlands	m.a.grzegorczyk@rug.nl	30
Benoit	Liquet	France	benoit.liquet@univ-pau.fr	15
Paulo	Lisboa	United Kingdom	P.J.Lisboa@ljmu.ac.uk	20
Marta	Lovino	Italy	marta.lovino@polito.it	33
Fernando	Luis-Ferreira	Portugal	flf@fct.unl.pt	33,43
Mónica	Machado	Portugal	monica_mac@hotmail.com	21
Gonçalo	Magalhães-Mota	Portugal	g.barreto@campus.fct.unl.pt	21,22,23
Marco	Masseroli	Italy	masseroli@elet.polimi.it	38
Telma	Marques	Portugal	td.silva@campus.fct.unl.pt	23
Giancarlo	Mauri	Italy	giancarlo.mauri@unimib.it	19
Ivan	Merelli	Italy	ivan.merelli@itb.cnr.it	19
Paulo	Morgado Zagalo	Portugal	p.zagalo@campus.fct.unl.pt	21
Alessandro	Nonis	Italia	nonis.alessandro@hsr.it	
Ivan	Olier	United Kingdom	I.A.OlierCaparroso@ljmu.ac.uk	34
Alberto	Paccanaro	United Kingdom	Alberto.Paccanaro@rhul.ac.uk	24
Hassan	Pazira	The Netherlands	pazira.b@gmail.com	20

Marzio	Pennisi	Italy	mpennisi@dmi.unict.it	36
Sara	Pereira	Portugal	sr.pereira@campus.fct.unl.pt	22
João	Pereira da Silva	Portugal	jvp.silva@campus.fct.unl.pt	21
Érica	Pinto	Portugal	ek.pinto@campus.fct.unl.pt	22
Filipa	Pires	Portugal	af.pires@campus.fct.unl.pt	22,23
Thais	Pivetta	Portugal	t.pivetta@campus.fct.unl.pt	22
Maria	Raposo	Portugal	mfr@fct.unl.pt	21,22,23
Matteo	Re	Italy	re@di.unimi.it	41
Paulo	Ribeiro	Portugal	pfr@fct.unl.pt	21,22,23
Leonardo	Rundo	Italy	leonardo.rundo@disco.unimib.it	19
Jeniffer	Santos	Portugal	jfd.santos@campus.fct.unl.pt	
Susana	Sério	Portugal	susana.serio@fct.unl.pt	23
Mahdi	Shafiee Kamalabad	Netherlands	m.shafiee.kamalabad@rug.nl	30
Simeon	Spasov	United Kingdom	ses88@cam.ac.uk	27
Antonino	Staiano	Italy	antonino.staiano@uniparthenope.it	25
Roberto	Tagliaferri	Italy	robtag@unisa.it	27
Usman Sumo F.	Tambunan	Indonesia	usman@ui.ac.id	36,37
Giorgio	Valentini	Italy	valentini@di.unimi.it	17,41
Alfredo	Vellido	España	avellido@cs.upc.edu	16
João	Villa de Brito	Portugal	joao.v.brito@tecnico.ulisboa.pt	31
Veronica	Vinciotti	United Kingdom	Veronica.Vinciotti@brunel.ac.uk	16

HOW TO GET TO FCT

The campus is served by a wide transport network, including combined services (bus, train, ferry) and by tram, which terminus ("University") is located along one of the FCT entrances. Here are some suggestions, although there are other possibilities, combining public transportation.

From Lisbon (Portela Airport)

There is a special shuttle bus (Carris nr 91), that takes passengers to Lisbon downtown in 20 minutes. This bus also stops at “Cais do Sodré” Train Station, next to the Ferry Station, where you can take a ferryboat to “Cacilhas”. Once in “Cacilhas” you can take a tram to “Universidade”, hopping off on its terminus (located near one of the campus entrances). In Cacilhas you can also take one of the following TST buses: “Marisol”, “Fonte da Telha” or “Costa de Caparica – via Almada”. They all stop next to the FCT main entrance. (In the airport there are other town bus lines: 5, 22, 44, 45 and 83 to Lisbon).

From Lisbon (Praça de Espanha)

You can use the TST bus to “Costa da Caparica” and get out at the bus stop located right in front of FCT main entrance.

From Lisbon (Cais do Sodré)

In the “Cais do Sodré” Ferry Station, you can take a ferryboat to “Cacilhas”. Once in Cacilhas you take a tram to “Universidade”, hopping off on its terminus (located near one of the campus entrances). In Cacilhas you can also take one of the following TST buses: “Marisol”, “Fonte da Telha” or “Costa de Caparica – via Almada”. They all stop next to the FCT main entrance.

From Lisbon (Santa Apolónia Train Station)

Metro to “Amadora Este” (blue line), get out in “Jardim Zoológico” station. Take the Fertagus train to “Coina” or “Setúbal” and get out in “Pragal”. Once there, take the tram to “Universidade”, hopping off on its terminus (located near one of the campus entrances).

From Lisbon (Belém)

Take the ferry to “Trafaria”. Once there, take the TST bus to “Cacilhas”, getting out on the bus stop next to FCT.

From Costa de Caparica

Take a TST bus to “Cacilhas – via Almada” and get out in the bus stop right in front of FCT main entrance.

SPONSORS


FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA


CEFITEC
CENTRO DE FÍSICA E INVESTIGAÇÃO TECNOLÓGICA
Centre of Physics and Technological Research