

Bioinformatics (BioInf)

Information Technologies (IT) proposal for the MAP-I PhD Program

A – Programmatic Component

1. Theme and motivation

In **Bioinformatics** [2], [3] computer based **algorithms, techniques and software tools** are used to solve highly complex problems in the domain of Molecular Biology, including Genomics, Proteomics and Metabolomics. With the development of new generation sequencing (NGS) the amount of genomic data is overwhelming and can only be analysed by computers. Even so, in recent projects like the sequencing of single cells project, only very powerful computers are able to “do the job” and produce in the order of several peta-bytes of data.

Another good news for Bioinformatics is that there are a large amount of publicly available data and knowledge (ontologies for example) that enable computer scientists to analyse the available data.

Developments in Genomic help to shed light into the connection between genomics and cancers. Furthermore new developments in the pharmaceutical industry enable the testing of hundreds of drugs and doses of drugs in thousands of “cell lines” which produce, again, a huge amount of data.

To profit from the above conditions of valuable data and knowledge, Data Mining research start to develop new algorithms and methodologies to be able to analyse the huge amounts of available data.

In this course we will study the basics of Molecular Biology and Chemo informatics that will enable the students to understand the major problems of current Molecular Biology. Using Data Mining algorithms and tools students will be able to do hands-on projects using real data to solve complex problems.

2 Objectives and Learning Outcomes

- To acquire basic concepts in Molecular Biology, Genomics, Proteomics, Metabolomics and Cheoinformatics;
- To be knowledgeable in Genome sequencing techniques;
- To acquire basic concepts and knowledge in Data Mining;
- To master the use of Data Mining techniques to solve Bioinformatics problems;
- To master the use of Data Mining to problems in Rational Drug Design essays where new drugs are applied to cell lines;

3 Detailed Program

1 - Introductory Molecular Biology concepts: basic concepts in Molecular Biology, Genomics, Proteomics, Metabolomics and Bioinformatics.

2 – Use Data Mining techniques to solve Bioinformatics problems: application of data mining techniques to genomic, proteomic and metabolic problems. Apply Data Mining and Chemoinformatcis to drug efficacy assessments using cell lines.

3 - Conducting Bioinformatics projects and case studies: Using real data and domain knowledge to enrich the datasets.

4 - Exploration of Web available data, knowledge and tools: Combining available tools with different sources and types of data and knowledge. Handling heterogeneous data sets in complex problems.

4 Teaching Methodologies and Evaluation

Four teaching methodologies will be applied:

- 1 - Lecture exposition of key concepts and techniques.
- 2 - Active learning.
- 3 - Case-based learning.
- 4 - Project based learning.

Evaluation will include two elements:

- A** - review of Bioinformatics research article(s) (30%, individual essay); and
- B** – a Bioinformatics project that describes the application of Bioinformatics and Data Mining tools to real-world Genomics datasets (70%, group project of 2 students).

Block Schedule Request

The lectures should be in a single block per week.

5 Bibliography

References:

- [1] Bramer,Max: Principles of data mining, Springer, 2020.
- [2] ed. K. G. Srinivasa, G. M. Siddesh, S. R. Manisekhar : Statistical modelling and machine learning principles for bioinformatics techniques, tools, and applications, Springer, 2020.
- [3] ed. Ravi Bhramaramba, Akula Chandra Sekhar : Application of computational intelligence to biology.
- [4] Ka-Chun Wong : Big data analytics in genomics, 2016.
- [5] Jure Zupan and Johann Gasteiger : Neural networks in chemistry and drug design, 1999
- [6] Shan-Hwei Nienhuys-Cheng, Ronald de Wolf ; Foundations of inductive logic programming, 1997.

B Lecture Team

1. Summary

The course is lectured by **Rui Camacho** from the Laboratory of Artificial Intelligence and Decision Support (LIAAD) R&D centre, <http://www.liaad.up.pt/>, of INESC-TEC and Faculty of Engineering University of Porto (FEUP). R. Camacho researches in Inductive Logic Programming and Data Mining. The following domain applications are used to apply the research developments: Bioinformatics, Biomedicine and Chemoinformatics. For more details, see the CVs in section B.3.

2. Coordinator

Rui Camacho (RC)

3. CV

3.1 Rui Camacho

Biography: Rui Camacho received his Ph.D. in Electrical Engineering and Computers from the University of Porto (UP), Portugal, in 2000. He is Associate Professor at the Informatics Engineering Department of the Faculty of Engineering at UP, teaching undergraduate and graduate classes of Machine Learning and Data Mining. He is also researcher at the Laboratory of Artificial Intelligence and Decision Support (LIAAD) that is part of INESC-TEC, with the current research interests:

- Inductive Logic Programming;
- Parallel Algorithms for Data Mining
- Data Mining and Machine Learning;
- Relational Data Mining;
- Applications of Chemoinformatics;
- Applications of Bioinformatics;
- Applications of Biomedicine.

Rui Camacho was been collaborating with researchers from I3S for long time and has more than 5 conference and journal publications with such researchers as a result of this collaboration (see the Relevant publications section).

Relevant publications in the last 5 years:

His most significant publications for the field in the last 5 years are:

- [1] Raquel Teixeira; Carina Rodrigues; Carla Moreira; Henrique Barros; Rui Camacho; “Machine learning methods to predict attrition in a population-based cohort of very preterm infants”; 2022; Scientific Reports, VOLUME: 12, ISSUE: 1
- [2] Goncalves, CAO; Camacho, R ; Goncalves, CT; Vieira, AS; Diz, LB; Iglesias, EL; “Classification of Full Text Biomedical Documents: Sections Importance Assessment”, 2021, APPLIED SCIENCES-BASEL, vol. 11, N. 6
- [3] Pedro Ferreira; João Ladeiras; Rui Camacho ; “Assessing the Impact of Data Set Enrichment to Improve Drug Sensitivity in Cancer”; 2021; 15th International Conference on Practical Applications of Computational Biology and Bioinformatics, PACBB 2021 in Practical Applications of Computational Biology & Bioinformatics, 15th International Conference (PACBB 2021), Salamanca, Spain, 6-8 October, 2021., VOLUME: 325
- [4] Michael A P Domingues; Rui Camacho ; Pedro Pereira Rodrigues; “CMIID: A comprehensive medical information identifier for clinical search harmonization in Data Safe Havens”; 2021; J. Biomed. Informatics, vol 114
- [5] Cavadas, B; Leite, M; Pedro, N; Magalhaes, AC; Melo, J; Correia, M; Maximo, V ; Camacho, R ; Fonseca, NA ; Figueiredo, C ; Pereira, L; “Shedding Light on the African Enigma: In Vitro Testing of Homo sapiens-Helicobacter pylori Coevolution”; 2021; MICROORGANISMS, vol 9, N. 2

- [6] Bruno Cavadas; Rui Camacho ; Joana C Ferreira; Rui M Ferreira; Ceu Figueiredo ; Alvis Brazma; Nuno A Fonseca ; Luisa Pereira; “Gastric Microbiome Diversities in Gastric Cancer Patients from Europe and Asia Mimic the Human Population Structure and Are Partly Driven by Microbiome Quantitative Trait Loci”; 2020, MICROORGANISMS, vol 8,N. 8
- [7] Mafalda Falcão Ferreira, MSc.; Rui Camacho; Luís Filipe Teixeira, “Using Autoencoders as a Weight Initialization Method on Deep Neural Networks for Disease Detection”, BMC Supplements, BMC MEDICAL INFORMATICS AND DECISION MAKING, vol 20, N. 5; 2020
- [8] Empowering Distributed Analysis Across Federated Cohort Data Repositories Adhering to FAIR Principles; Artur Rocha ; José Pedro Ornelas ; João Correia Lopes ; Rui Camacho ; 2019, ERCIM NEWS, vol 2019, N. 118
- [9] *Carlos Adriano Gonçalves, Eva Lorenzo Iglesias, Lourdes Borrajo Rui Camacho, A. Seara Vieira, Célia Talma Gonçalves: LearnSec: A Framework for Full Text Analysis. HAIS 2018: 502-513*
- [10] Artur Rocha, Rui Camacho, Jeroen Ruwaard e Heleen Ripper, “Using multi-relational techniques to discriminate blended therapy efficiency on patients based on log data”, ISII, Vol. N. pp , 2018
- [11] *Noé Vázquez, Sara Rocha, Hugo López-Fernández, André Torres, Rui Camacho, Florentino Fdez-Riverola, Jorge Vieira, Cristina P. Vieira, Miguel Reboiro-Jato: EvoPPI: A Web Application to Compare Protein-Protein Interactions (PPIs) from Different Databases and Species. PACBB 2018: 149-156*
- [12] “Co-expression networks between protein encoding mitochondrial genes and all the remaining genes in human tissues”, João Almeida, Joana Ferreira, Rui Camacho, Luisa Pereira, BIBM 2017

Participation in R&D projects in the last 5 years:

He participated in various R&D projects:

- Safe Cities - Inovação para Construir Cidades Seguras: referência POCI-01-0247-FEDER-041435
- EUCAN_CONNECT – Connecting Europe and Canada in personalized & preventive health care, H2020 project
- RECAP - “Research on European Children and Adults born Preterm“, H2020 project :
- E-COMAPERD - “European Comparative Effectiveness Research on Internet-based Depression Treatment”, H2020 project
- NanoStima (Norte2020) project

Principal Investigator of the project: **ILP-Web-Service: An Inductive Logic Programming based Web service**

Supervision of Graduate Students:

Has supervised 57 MSc theses, 8 PhD theses and 3 Post-doc students. Currently supervises 7 PhD and 9 MSc students.