Giorgio Valentini
CURRICULUM VITAE

SYNTHETIC CURRICULUM

Degree in Biology and Computer Science, PhD in Computer Science (University of Genoa). Associate Professor at the Department of Computer Science, University of Milan (confirmed in 2014). National Qualification as full professor in 2017. Member of the UNIMI Doctoral School of Informatics and scientific director for UNIMI of the European doctorate in Genomics and Bioinformatics in collaboration with the Joint Research Center of the European Union. Research activity: development and application of Artificial Intelligence methods to bio-medical problems, with ongoing collaborations with various European and American research centers and universities. He is author of over 140 scientific publications with peer-review in journals, book chapters and international conferences in the field of Machine Learning, Bioinformatics and Computational Biology.

ANALYTICAL CURRICULUM

Qualifications and courses of study

- National scientific qualification for the function of full professor - sector 01 / B1 Informatica - valid from 10/04/2017 to 10/04/2023 (article 16, paragraph 1, Law 240/10)
- PhD in Computer Science, University of Genoa (2003). Title: "Ensemble methods based on bias-variance analysis".
- Master Degree in Information Sciences, University of Genoa (1999), grade: 110/110 cum laude.
- MasterDegree in Biological Sciences, University of Genoa (1981), grade: 110/110 cum laude.
- Classical Maturity, Classical Lyceum of Savona (1977), grade: 56/60.

Actual position

Associate Professor (scientific-disciplinary sector INF / 01) at the Department of Computer Science, University of Milan since 2010.
I teach didactic courses within the Computer Science teaching coordination board of the same University. I received the confirmation in the role in 2014 from the Faculty and the Department of belonging.

Previous positions

- Permanent researcher at the Department of Computer Science, University of Milan (2005-2009)
- Post-doc researcher (research fellow) at the Department of Computer Science, University of Milan (2003-2004)
- PhD student at the University of Genoa (2000-2003)
- Contract researcher at the National Institute of Physics of Matter (1999), Genoa
- Lecturer in natural sciences, biology and chemistry and at several Italian high schools (1983-1998)

Research activity

My research activity is between Machine Learning and Bioinformatics, and is motivated by complex problems in Molecular Biology and Medicine. By modeling these problems in a Machine Learning context, I design and develop new machine learning algorithms or adopt suitable existing algorithms to address relevant problems in the field of Bioinformatics, with a particular interest, especially in recent years, in the development of Machine Learning methods for Medicine Genomics, and Precision and Personalized Medicine.
Although in my research activity the development of new Machine Learning methods is closely linked to real problems in the bio-medical field, I have also developed "pure" Machine Learning research lines, especially for the design and analysis of ensemble methods.
For this reason the general outline of my main research lines can be schematically articulated as follows:

I. *Bioinformatics*
- A. *Analysis, development and application of supervised Machine Learning methods to Bioinformatics*
  - A1. Machine Learning Methods for Personalized Medicine
  - A2. Hierarchical ensemble methods for structured prediction in biological ontologies
  - A3. Supervised ensemble methods for supporting bio-molecular diagnosis
- B. *Analysis development and application of semi-supervised Machine Learning methods to Bioinformatics*
  - B1. Methods based on kernelised score functions for the analysis of complex biomolecular networks.
  - B2. Methods based on parameterized and cost-sensitive Hopfield networks for the prediction of gene function.
  - B3. Scalable methods for the analysis of big biomolecular networks.
- C. *Analysis development and application of unsupervised Machine Learning methods to Bioinformatics*
  - C1. Stability analysis based methods for evaluating the reliability of clusters identified in complex bio-molecular data
  - C2. Ensemble clustering methods for pattern research in bio-molecular data
- D. *Methods for the integration of big biological and medical data.*
  - D1. Algorithms for the massive combination of biomolecular networks
  - D2. Supervised ensemble methods for the integration of omics data

II. *Machine Learning*
- A. *Analysis and design of ensemble methods*
  - A1. Hyper-ensemble methods for unbalanced classification problems with big data
  - A2. Multiclass, multi-label and multi-path hierarchical ensemble methods
  - A3. Ensemble methods based on the decomposition of the error into bias and variance
  - A4. Supervised ensemble methods based on randomized projections
  - A5. Methods of ensemble with error correction codes for multiclass classification.
  - A6. Ensemble clustering methods
- B. *Design and implementation of machine learning software libraries*

Description of the research lines
The research lines are summarized below with reference to the publications listed at the bottom of ththis document.

I. *Bioinformatics*

A. *Analysis, development and application of supervised Machine Learning methods to Bioinformatics*

A1. Machine Learning Methods for Personalized Medicine

The identification of genetic variants associated with human pathologies represents one of the fundamental challenges of "Personalized and Precision Medicine", and requires the development of a new generation of Machine Learning methods to select the rare potentially "deleterious" variants (i.e. causative or associated with the risk of disease) in the sea of "neutral" variants that represent the "physiological" genetic variability of each individual. To this end I developed hyperSMURF (hyper-ensemble of SMOTE under-sampled random forests), a new method that adopts learning strategies based on resampling techniques and hyper-ensembling techniques to tackle the problem of the “imbalance” that

characterize several types of genomic [R49, C77]. This Machine Learning method constitutes the "core" of Genomiser, a methodology and a software tool recently proposed in the context of an international collaboration, which uses both genotypic and phenotypic data to identify pathological variants causing Mendelian genetic diseases [R48].
HyperSMURF is a general method for the analysis of genetic variants and can be applied to the study of different pathologies, but its performance depends significantly on the tuning of learning parameters [C78]. For this reason we developed a highly parallel version of the algorithm for High Performance Computing architectures in the context of the HyperGeV project - Detection of Deleterious Genetic Variation through Hyper-ensemble Methods in order to fine-tune the parameters and analyze big data in the the context of relevant Genomic Medicine problems [paper submitted to GigaScience].
I am also developing imbalance-aware learning techniques based on mini-batch sizing and mini-batch imbalance-aware sampling techniques for predicting pathogenic genetic variants with deep neural networks [C84].

A2. Hierarchical ensemble methods for structured prediction in biological ontologies

Relevant concepts in the field of molecular biology (for example: the functions of genes and proteins) and medicine (abnormal phenotypes associated with human pathologies) are organized according to hierarchical ontologies structured as trees (eg: FunCat for the functional classification of genes) or as acyclic direct graphs (DAG) (eg: the Gene Ontology (GO) for the classification of genes and proteins and the HPO (Human Phenotype Ontology) for the classification of pathological human phenotypes.
In this context I developed hierarchical ensemble methods [R39] based on the true path rule (TPR) [R29, C44, C48, C51] and cost-sensitive Bayesian methods for probabilistic reconciliation of the output of the base learners [R25, C50, C53] and for the structured prediction of the function of genes and proteins in tree-structured ontologies. I have therefore shown that the combination of hierarchical ensemble methods, cost-sensitive learning strategies and the integration of different types of data significantly improve the performance in the prediction of gene functions at the level of the entire genome [R25, R32].
In the context of the prediction of abnormal human phenotypes according to the HPO, I have recently proposed new hierarchical ensemble methods for structured prediction based on DAG that have achieved state of the art results [R50, C76, C79, C70, C71]. Recently, new methods based on isotonic regression algorithms combined with the TPR algorithm led to state-of-the-art results in the prediction of protein function [paper in preparation].

A3. Supervised ensemble methods for supporting bio-molecular diagnosis

The bio-molecular classification of pathological phenotypes requires the development of methods well-suited to the characteristics of the "omics" data used, often characterized by high dimensionality. In this context, various methods of supervised ensembles were explored, such as methods based on error correcting codes [R3, C11, C14], on the reduction of the dimensionality of data through randomized projections [R8, C22, C23], or methods of bagging and its variants [R5, C18, C19] and ither based on the analysis of data complexity [C37, C43].
Finally, I applied univariate feature selection and cost-sensitive SVM methods to the analysis of radiographic images (classification of pulmonary nodules) with results comparable with the best ones in the literature [R9, C25].

B. *Analysis development and application of semi-supervised Machine Learning methods to Bioinformatics*

In the field of Systems Biology and Network Medicine, I developed semi-supervised graph-based learning methods for the study of biological systems as complex entities, in which the biological functions derive fundamentally from the relationshipds between the parts that make up the system.

B1. *Methods based on kernelised score functions for the analysis of complex biomolecular networks.*

The problem of the association of genes or proteins or more generally of biomolecules with a specific biological property (eg: a biomolecular function, a diseases or a drug indication) can be modeled as a problem of node label ranking on a graph. Most of the methods proposed for graph analysis use local or global learning strategies to rank the nodes or predict the arcs of the graph itself. To integrate local and global learning strategies, I proposed semi-supervised transduction methods that are able to learn both

the global topology and the local characteristics of each node of the graph [R46] through the graph kernelization.
These methods have been successfully applied to the analysis of complex biomolecular networks for protein function prediction, to gene ranking problems (disease gene prioritization), and drug repositioning problems (search for new therapeutic indications of drugs originally designed for different therapeutic purposes) [R31, R33, R35, R38, R41, C61].
Finally, I recently proposed new semi-supervised network-based methods based on graph kernels which instead of analyzing the "biomarker space", as usually is the case in the context of Network Medicine, analyze the "patient space", through the construction of patient networks based on the similarity of their biomolecular profiles (eg: considering their expression or genetic profiles). The "patient networks" can be used both for diagnostic or prognostic purposes, to stratify the patients themselves into pathological subtypes, to identify biomarkers associated with specific pathologies or to predict the response to drugs [N21, paper submitted].

B2. *Methods based on parameterized and cost-sensitive Hopfield networks for the prediction of gene function.*

In the context of the prediction of the function of proteins and genes according to the GO, a significant problem is represented by the imbalance in the annotations. In fact for the majority of GO classes only a relatively small number of annotated genes (positive examples) is available. In this context, classical learning methods (including the Hopfield Networks) tend to always predict the majority class. For this reason we designed a new class of parameterized Hopfield Networks (COSNet) able to learn their own learning parameters from the data, explicitly taking into account the imbalance between positive and negative examples [R36, R44, C58].
A COSNet variant, designed to explicitly consider "categories" of neurons known a priori in the network has been successfully applied to the prediction of the function of proteins in a multi-species context [R45].
Finally, an "imbalance-aware" method of data integration coupled with the parametric Hopfield Networks has obtained state-of-the-art results for the prediction of protein functions [R43, C67]

B3. *Scalable methods for the analysis of big biomolecular networks.*

Big data analysis in large biomolecular networks represents a significant problem in computational biology. In this context, I developed vertex-centric algorithms and used technologies such as GraphChi based on the efficient use of secondary memory to process large graphs that cannot be directly loaded into primary memory. The goal is to analyze big data built with "omics" data, with relevant applications in Molecular Biology and Medicine, using simple stand-alone workstations. Promising experimental results have been obtained in the context of multi-species prediction of protein function [R42, C65, C74].
Another line of research concerns the use of GPU technology for the massively parallel implementation of node label prediction algorithms such as COSNet for efficient processing of large graphs [R52], with applications to the prediction of GO classes, using the integrated multi-species network of the STRING database [N20] which includes millions of proteins of different species. The construction of a network consisting of tens of millions of human SNPs (single nucleotide polimorphisms) for the search for potentially deleterious or pathological SNPs through the parallel version [R52] of COSNet is being developed.

C. *Analysis development and application of unsupervised Machine Learning methods to Bioinformatics*

C1. *Stability analysis based methods for evaluating the reliability of clusters identified in complex biomolecular data*

The validation of the clusters identified by clustering algorithms is a problem of great relevance in bioinformatics: genomics and proteomics present several problems in which it is fundamental to evaluate the reliability of the structures and patterns identified in complex biomolecular data.
The research activity was articulated in the development of algorithms for the analysis of the reliability and selection of the order of the model for unsupervised problems [R14, R15, R18, C29, C34, N10], and algorithms for the analysis of the reliability of the single clusters [R11, R12, C24], using a new approach based on the analysis of the stability of the clusters obtained. We developed statistical tests based on the

chi square distribution [R15, C29] and on the classical Bernstein inequality [R18, C34] for the search for multiple structures in complex data.
These methods were applied to the validation of pathological subclasses characterized at the bio-molecular level and to the search for multiple structures in biomolecular data, using data generated by high-throughput bio-technologies [R12, R13, R15, R18, R22, C39, N9, N10].

C2. Ensemble clustering methods for pattern research in bio-molecular data

The search for bio-molecular patterns in data characterized by high dimensionality and low cardinality (for example: DNA microarray or spectrometric data related to proteins), led to the design and development of clustering ensemble methods specific for this type of data. In particular, unsupervised methods based on randomized projections have been developed to analyze data characterized by high dimensionality [C28]. These methods were subsequently applied to the analysis of gene expression data [C31]. We also developed clustering ensemble methods based on randomized projections that use a fuzzy approach both for the basic clustering constituting the ensemble, and for combining the clustering obtained on multiple instances of the data. From the initial algorithm [C33] a more general algorithmic scheme has been developed from which different fuzzy ensemble clustering algorithms can be derived [C36] and this approach has been applied to the analysis of gene expression data for the search of pathological subclasses characterized at bio-molecular level [R20].

D. *Methods for the integration of big biological and medical data.*

D1. *Algorithms for the massive combination of biomolecular networks*

The integration of biomolecular networks constructed with different types of "omics" data is a significant problem in the field of Systems Biology and Network Medicine.
My research activity has contributed to show that the integration with non-weighted and weighted methods of different biomolecular networks plays an essential role in identifying the genes associated with over 700 pathologies [R41]. I have also proposed several methodologies for the construction and integration of biomolecular networks, with applications to the prediction of protein function, and to the problems of disease gene prioritization and drug-target association [R33, R38, R43, C63, C66, C67, C75].

D2. *Supervised ensemble methods for the integration of omics data*

My research activity has shown that even relatively simple ensemble methods such as majority voting or decision templates can achieve results comparable with state of the art in "omics" data integration [R24, R26]. My other works have confirmed the effectiveness of ensemble methods for the integration of complex biomolecular data [C41, C45, C46, C47], also showing that the ensemble methods are also able to tolerate relatively high levels of noise in the data, without a significant deterioration in performance [R27]. I have also carried out studies on the application of ensemble methods for the integration of heterogeneous data and for the prediction of the subcellular localization of proteins [C54, C57], and on the use of XML for the integration of heterogeneous biomolecular data [R23, C28 , C40].

II. *Machine Learning*

Although my research on new machine learning methods has been mostly motivated by complex problems in Molecular Biology and Medicine, I have also developed "pure" Machine Learning research lines, especially for the design and analysis of ensembles of learning machines.

A. *Analysis and design of ensemble methods*

A1. *Hyper-ensemble methods for unbalanced classification problems with big data*

Many relevant classification problems, not only in the field of Computational Biology and Genomic Medicine, are characterized by a strong imbalance of the examples between the classes. In this context classical Machine Learning methods tend to be biased towards the majority class (negative class) and fail

to learn the examples of the minority class (positive class). The hyperSMURF method, originally designed for strongly unbalanced genomic problems [R49] is general enough to be applied in other contexts characterized by a strong data imbalance. The algorithm, based on hyper-ensembling and data resampling techniques, has achieved state-of-the-art results in the field of Genomic Medicine [R48].
The parallelized version, parSMURF, is able to process big data, and further improves the performance of hyperSMURF, thanks to the automatic tuning of its learning parameters. The methods scales nicely with big data, and has been deployed in two flavours: a version for stand-alone multi-core architectures and another one for High performance computing clusters, depending on the level of complexity of the problem addressed [article submitted to GigaScience].

A2. *Multiclass, multi-label and multi-path hierarchical ensemble methods*

The problem of gene clssification has stimulated the research and development of multiclass classification algorithms (the functional classes of genes are of the order of hundreds or thousands), multi-label (a gene can belong to several classes) and multi-path (classes are structured according to trees or DAGs). The hierarchical ensemble algorithms that we developed for this type of problem [R25, R29, R32, R50] have a broader value and can be applied in other contexts characterized by hierarchical tree or DAG-structured taxonomies. These algorithms adopt a two-step learning strategy: at first the individual classes are learned in a "flat" manner by the base learners, and then the predictions of the trained models are combined according to the class hierarchy. Recently we developed an algorithm for DAG-structured taxonomies based on the integration of isotonic regression and bottom-up combination techniques (from the most specific classes to the most general classes) which guarantees the consistency of the predictions and systematically improves the predictions of the “flat” learning algorithms [article in preparation].

A3. *Ensemble methods based on the decomposition of the error into bias and variance*

In this line of research the decomposition of the error into bias and variance is used as a tool to analyze the property and characteristics of the learning algorithms.
Based on the Domingos theory, which generalizes to the 0/1 loss function the classical analysis based on the quadratic loss function, I analyzed the relationships between learning and decomposition of the error in bias and variance in the case of Support Vector Machines [R7].
The characterization of the learning processes of SVMs in terms of the decomposition of the error into bias and variance also offers a rationale for the development of new ensemble methods [R7, C16].
Taking advantage of the SVM bias-variance analysis, I proposed a new ensemble algorithm, called Lobag (Low Bias Bagging), which estimates SVM bias, selects SVMs with minor bias and then combines them through bootstrap-based aggregation mechanisms. This approach jointly reduces the bias and variance of the error, and can be interpreted as a "low-bias" variant of bagging [C20]. The method has been successfully applied to the classification of bio-molecular tumor diseases [C19]. The bias-variance error analysis was subsequently extended to resampling-based ensemble methods, showing the relationships and differences in the learning mechanisms of bagging methods, random aggregation and Lobag [R10, C21].

A4. *Supervised ensemble methods based on randomized projections*

In relation to the bio-molecular diagnosis of cancer, ensemble methods based on random subspaces have been developed, using SVM as a learner base [R8, C22]. An extension of the model, which includes a feature selection stage to eliminate the less relevant features for classification, followed by the application of the random subspace method on the remaining features, showed competitive results with state-of-the-art ensemble methods published in the literature [C23].

A5. *Methods of ensemble with error correction codes for multiclass classification.*

ECOC (Error Correcting Output Coding) ensemble methods allow to improve the reliability of the prediction for multiclass classification problems through the redundant coding of the class labels, realized with the decomposition of a multiclass problem into a series of dichotomous problems solved by a classifier ensemble.
In this context I analyzed the effectiveness of ECOC methods for multi-class problems in both ensembles of learning machines and single “monolithic” learning machines [R4, C2] and subsequently I experimentally analyzed the dependency between errors at the level of the single bits of the ECOC

codewords through measurements based on mutual information [R6], in order to compare different types of ECOC codings and different architectures of machine learning systems based on ECOC [C5, C7, C8].
In addition to applications in the bioinformatics field [R3, C11, C12], the ECOC ensembles (together with boosting methods) have been successfully applied also to multi-class problems with electronic noses [R1, C3, C6, C15].

A6. *Ensemble clustering methods*

Unsupervised ensemble methods based on randomized projections, motivated by clustering problems in spaces of high dimensionality and reduced cardinality that characterize different problems in the bioinformatics field, represent an unsupervised extension of the methods of supervised random subspace [C28].
In [C24] I showed how the random projections induced by the random subspace method can produce significant distortions in the gene expression data, while using randomized projections that obey the Johnson and Lindenstrauss lemma we can generate data of reduced dimensionality whose metric features are similar to those of the original space [R13]. In accordance with this analysis, we proposed clustering ensemble methods based on randomized projections [C31] that have been successfully applied to the analysis of DNA microarray data [C28].
A fuzzy extension of the ensemble method proposed in [C31] has been developed in [C36]: from the combination of different "crispization" mechanisms of basic fuzzy clustering and different types of fuzzy aggregation of basic clustering, we can derive an algorithmic scheme from which several fuzzy ensemble clustering methods can be obtained [C36]. These methods have been applied to the analysis of gene expression data [R20, C32, C33].

B. Design and implementation of machine learning software libraries

My research activity both in Machine Learning and in Bioinformatics has always been accompanied by the design and implementation of software libraries.
Several ensemble methods that I developed were made available and implemented in the C++ library, NEURObjects, initially conceived for implementing neural networks [R2, C1].
Later, in parallel with the increase in research activities in the bioinformatics field, I designed and implemented open source R libraries for the analysis of complex bio-molecular data. In particular the clusterv library [R11] allows to analyze the reliability of single clusters in high-dimensional bio-molecular data, the mosclust library [R14] allows to determine the "optimal" number of clusters and to identify multiple structures present in complex bio-molecular data, while the hcgene library [R17] allows to analyze the acyclic direct graphs of the Gene Ontology and the FunCat trees to support the functional classification of proteins. The RANKS library [R46] provides graph-based algorithms for node label ranking and classification problems. The COSNet library implements new models of parametric Hopfield networks [R44], while HEMDAG [R50] hierarchical ensemble methods and hyperSMURF [R49] "imbalance-aware" hyper-ensembling methods for the analysis of strongly unbalanced data, and its parallel extension (parSMURF) in C ++ is being released for the analysis of genomic big data [paper submitted to an international journal].

Indices of scientific productivity

The productivity indexes extracted on 7/4/2019 from the Scopus and Google Scholar database are presented below:

Index	Scopus	Google Scholar
Number of citations	1642	3261
H - index	23	32

Index	Scopus	Google Scholar
i10 - index	50	75

Management or participation in the activities of a research group characterized by national or international collaborations

A) Direction of AnacletoLab, Laboratory of Computational Biology and Bioinformatics of the Department of Computer Science, University of Milan
http://anacletolab.di.unimi.it/.

The laboratory brings together a group of professors / researchers, grant holders and PhD students and collaborates with different national and international research groups in the field of Computational Biology, Molecular Biology and Medicine, including:
a) Computational Biology Group of the Charite - University of Humboldt Universitat Berlin,
b) the Computational Biology group of the Berlin Institute of Health,
c) the Computer Science dept. of the Royal Holloway, University of London,
d) the Jackson Laboratory for Genomic Medicine, CT, USA,
e) The Artificial Intelligence department of the University of Granada, Spain
f) The Computer Science Dept of Aalto University, Helsiniki, Finland,
g) the Wellcome Trust Sanger Institute and the European Bioinformatics Institute (EBI) of Hinxton, UK,
h) The Division of Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA, USA,
i) the European Center for Living Technologies of Venice,
l) Other national research groups both in the medical field (National Cancer Institute, and S. Raffaele Hospital of Milan, National Institute of Molecular Genetics) and computer science (University of Salerno and University of Cagliari).
(Period: from 2010 to today)

B) Participation as director of the AnacletoLab laboratory of UNIMI to the international challenge CAFA2 (Critical Assessment of Functional Annotation) within the Special Interest Group "Protein Function Prediction" of the ISCB (International Society of Computational Biology). The Special Interest Group brings together the main international research groups for the prediction of the function of proteins with computational methods (http://biofunctionprediction.org/).
In my position as head of the AnacletoLab research group, I participated in the research activity of the Special Interest Group "Protein Function Prediction" which led to the publication in Genome Biology (one of the main journal of computational biology) of a collective work involving the research activity of over 50 research groups from all over the world. In this context the laboratory was placed in the first 3 in the challenge CAFA2 for the prediction of human genes associated with pathological phenotypes.
AnacletoLab participated in the new challenge CAFA3 for the function prediction of human proteins and of the main model organisms and in the prediction of genes associated with abnormal human phenotypes.
The preliminary results of this challenge (the evaluation is still in progress) place AnacletoLab again among the "top ranked" groups at the international level for the prediction of the function of proteins.
(Period: from 2013 to today).

C) Participation in the activity of an international research group for the search of mutations associated with genetic diseases in non-coding regions of the genome.
My contribution consists in the development of machine learning methods specific to this problem.
The international group includes several research groups from Europe, North America and Australia, including Queen Mary University of London; Genomics England, UK; the Wellcome Trust Sanger Institute, Hinxton, UK; the Max Planck Institute for Molecular Genetics, Berlin; the Department of Biomedical Informatics and Intelligent Systems Program, University of Pittsburgh; the Division of Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA; the Department of Medical Informatics and Clinical Epidemiology,

Oregon Health & Science University, Portland, USA; the Kinghorn Center for Clinical Genomics, Garvan Institute of Medical Research, Darlinghurst, Australia; the Charité-Universitätsmedizin Berlin.

The computational methodology developed for the search of mutations associated with Mendelian genetic diseases in non-coding regions of the genome has been published in the American Journal of Human Genetics, a leading journal in this field. This method, called Genomiser, represents the state of the art at international level for the search for causative mutations of Mendelian genetic diseases. The Machine Learning method, developed specifically by me in collaboration with the Computational Biology group of the Charitè UniversitatzMedizin in Berlin, and which constitutes the core of Genomiser, was recently published in a journal of the Nature Publishing group. This method is general enough to be used in other contexts to search for genetic variants associated with genetic and tumor diseases, and in fact a collaboration is underway with the School of Medicine of the University of Utah to study the hereditary factors of various tumors.
(Period: 2014 to present)

D) Collaboration in a joint project with the Berlin Institute of Health (BIH, which brings together research groups from the von Humboldt and the Freie Universität in Berlin and from the Max Delbrück Center for Molecular Medicine) for the systematic study of genetic variations in regulatory regions of the human genome and of their impact on genetic diseases and cancer. This project requires the development of innovative Artificial Intelligence methodologies for the analysis of data generated by new biotechnologies, such as Massive Parallel Reporter Assay (MPRA) for the study of the functional in vivo effect of genetic variants, and for the analysis and massive integration of epigenomic data recently made available by the International Human Epigenome Consortium (IHEC). The goal is to decode the "genetic code" at the base of gene regulation in the non-coding regions of the genome, and to study its alterations to predict the pathological effect of genetic mutations underlying cancer and genetic diseases.
The project is currently funded by the German DAAD and the MIUR, but the complexity of the problem requires collaboration with other European and American scientific partners and relevant funding for the generation and bioinformatic processing of biotechnological data. For this reason we are working together with the BIH and other European and American partners for the preparation of a European research or a NIH funded project.
(Period: 2017 to present)

E) Collaboration with the Jackson Lab for Genomic Medicine (CT, USA) and with the Wellcome Sanger Institute - Hinxton (Cambridge, UK) for the search of splicing mutations in Exonic Sequence Enhancers (ESE, that promote exon splicing) in the human genome. These mutations, which are extremely difficult to detect because they are apparently "silent", are the basis of various tumors. The idea we are working on is to build predictors based on artificial intelligence techniques capable of identifying these mutations on the basis of a set of appropriate genomic and epigenomic features and based on ESE manually labeled through the analysis of the medical literature. After the publication of the first results (which appear very promising) we will prepare an application for a grant from the NIH (National Institute of Health of the United States) to finance this line of research.
(Period: 2018 to present)

F) Research collaboration with the Urology Department (URI) of the S. Raffaele Hospital (OSR) in Milan, regulated by a Confidentiality Agreement signed between UNIMI and OSR, for the project "Male infertility as a proxy of the progerois syndrome", of which I am the scientific director for UNIMI (Andrea Salonia, head of the URI unit is the scientific director for OSR). The purpose of this agreement is to start-up joint research activities for prparing a grant to the NIH.
(Period: 2019 to present)

G) In the a.y. 2015/16 I enjoyed a sabbatical year during which I carried out research at the European Center for Living Technologies of Venice (Prof. Pelillo's research group), at the Charitè, Faculty of Medicine of the von Humboldt University of Berlin (research group of Computational Biology of prof. Robinson), and the Artificial Intelligence Department of University of Granada (prof. Blanco).

Scientific responsibility (Principal Investigator - PI) of international and national research projects, eligible for funding on the basis of competitive calls for proposals involving peer review

- Scientific coordinator for UNIMI of the EU Collaborative Doctoral Partnership in Genomics and Bioinformatics, funded by the European Commission in collaboration with the EU Joint Research Center (2018-2022). Our University has been selected among the top 5 European Universities for the area of Genomics and Bioinformatics. The five-year collaboration contract is renewable at the end of the five-year period, subject to agreement between the parties, and is aimed at training researchers in the field of Genomics and Bioinformatics, able both to carry out innovative research in these scientific disciplines and to scientifically support the European Commission in making decesions about European policies for Medical Genomics and Data Analytics in Health-care.

- PI of the project "Developing machine learning methods for the prioritization of regulatory variants in human disease" in collaboration with the Berlin Institute of Health, funded by MIUR and DAAD (Germany) (2018-2019)

- PI of "HyperGeV: Detection of Deleterious Genetic Variation through Hyper-ensemble Methods" (2016-2018), funded by CINECA and the Lombardy Region

- PI of "HPC-SoMuC: Development of Innovative HPC Methods for the Detection of Somatic Mutations in Cancer" (2017-2018), funded by CINECA and the Lombardy Region.

- Head of the UNIMI unit of the project "A composite predictive model of response to Fingolimod: integration of clinics, neuroimaging and genomics", funded by the Italian Multiple Sclerosis Foundation. The project is leaded by the Genetics Laboratory of the Complex Neurological Diseases of the S. Raffaele Hospital in Milan. The genetic biomarkers associated with the response to the Fingolimod drug, identified by the data of S. Raffaele and the predictive model developed by the UNIMI unit are currently being patented (2016-2018).

- Head of the UNIMI unit (via subcontracting) of the Finding-MS project within the European project ERA-PerMed Joint Transnational Call (JTC) 2018 on "Research projects on Personalized Medicine - smart combinations for pre-clinical and clinical research with data and ICT solutions ". The project is in collaboration with the S.Raffaele Hospital of Milan, the CNR - Institute for Bio-medical Technologies, the Center Hospitalier Universitaire de Toulouse and geneXplain, a German bioinformatics company (2019-2022)

Participation in international and national research projects, admitted to the financing on the basis of competitive calls for peer review

- PRIN Project “Multicriteria Data Structures and Algorithms: from compressed to learned indexes, and beyond”, coordinated by the University of Pisa (2019-2021)
- "Discovering Patterns in Multi-Dimensional Data" project (2016-2017) funded by the University of Milan.
- Participation in the Milan unit of the Network of Excellence "Pattern Analysis, Statistical Modeling and Computational Learning 2 (PASCAL2)", 7th European Framework Program financed by the European Union (2007-2013)
- Participation in the Milan unit of the Network of excellence PASCAL, 6th European Framework Program (2004-2006)
- Project “Computational methods for bio-medical pattern analysis” funded by UNIMI (2011-2013)
- MIUR project COFIN-PRIN Automata and formal languages: mathematical and applicative aspects (2010-11);
- PUR 2009 Project: "Automatic methods for pattern analysis in the biomedical field" funded by the University of Milan;
- PUR 2008 Project: "Innovative computational models" funded by the University of Milan; IEIIT-CNR 2007 Project: Machine Learning Techniques for Modeling and Growing Up " (2007-2008);
- PUR 2006 Project: "Stochastic and quantum models for computational and di problems bioinformatics " funded by the University of Milan;

- PUR 2006 Project: "Identification of transcriptional profiles in Acute Myeloid Leukemia by microarray on hematopoietic stem fractions " funded by the University of Milan;
- MIUR Project COFIN-PRIN Formal Languages and automata: methods, models and applications (2003-2005)
- MIUR Project COFIN-PRIN Machine Learning Techniques for Bioinformatics: Analysis and Modeling of Functional and Structural Data of Gene Expression (2001-2002)
- From Bits to Information: Statistical Learning Technologies for Digital Information Management Search - USA, Project funded by the National Scientific Foundation (NSF) (2002).

Participation in editorial committees of international journals

- Member of the Editorial board of Scientific Reports, Nature (IF 2017: 4.122), indexed in Scopus and ISI Web of Science (from 1 June 2018).
- Member of the editorial board of Advances in Bioinformatics, Hindawi (IF 2017: 1.8), indexed in Scopus (from 2014)
- Guest Editor of Artificial Intelligence in Medicine, Elsevier (IF: 2.879) indexed in Scopus and ISI Web of Science for the Special issue "Computational Intelligence and Machine Learning in Bioinformatics" (2008-2009).
- I was also a member of the Editorial board of other minor journals in the Bioinformatics and Machine Learning area from 2008 to 2017.

Reviewer activity for international journals

I have worked as reviewer for the main international journals in Bioinformatics and Machine Learning, including:

- Journal of Machine Learning Research
- Machine Learning
- IEEE Transactions on Pattern Analysis and Machine Intelligence
- Pattern Recognition
- IEEE Transactions on Neural Networks and Learning Systems
- IEEE Transactions on Knowledge and Data Engineering
- Artificial Intelligence
- Computational Intelligence,
- Neurocomputing
- Neural Networks
- IEEE Transactions on Cybernetics,
- GigaScience
- Bioinformatics
- PLoS Computational Biology
- BMC Bioinformatics
- Briefings in Bioinformatics
- IEEE ACM Transactions on Computational Biology and Bioinformatics
- Artificial Intelligence in Medicine
- Journal of Bioinformatics and Computational Biology
- PLoS One

Conference and Workshop Organization

- Co-organizer of the workshop Soft computing methods for characterizing diseases from omics data – CIBB 2019, Bergamo 3-4 settembre 2019
- Chair of CIBB 2018 - Computational Intelligence methods for Bioinformatics and Biostatistics, Lisboa (Portogallo) (Caparica 6-8 Settembre 2018)

- Chair of the Workshop BigTargets, European Conference on Machine Learning (ECML), Porto (Portogallo), 2015;
- Chair of the Fourth Italian Workshop on Machine Learning and Data Mining (MLDM 2015) – XIV Conference of AIIA – Pisa, 2015;
- Chair of SUEMA 2010, Third International Workshop on Supervised and Unsupervised Ensemble Methods and Their Applications at ECML – European Conference on Machine Learning 2010, Barcelona, Spagna;
- Wokshop Learning from Multiple Sources with Applications to Robotics at NIPS 2009, Whistler, Canada, 2009;
- Chair of SUEMA 2008, European Conference on Artifical intelligence (ECAI) 2008 Patras, Grecia;
- Chair of CIBB 2007, Fourth International Conference on Computational Intelligence Methods for Bioinformatics and Biostatistics, Portofino, Italia
- Chair of SUEMA 2007, International Workshop on Supervised and Unsupervised Ensemble Methods and Their Application (nell'ambito di IbPRIA2007) a Girona (Spagna).

Participation in the scientific committee of international conferences.

Member of the Scientific Committee Program of over 50 international conferences within the scope of the machine learning and computational biology from 2007 to present, including:

- European Conference in Computational Biology
- European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases
- International Joint Conference on Artificial Intelligence (IJCAI) - Machine Learning Track
- International Symposium on Foundations and Applications of Big Data Analytics
- Automatic Function Prediction - Critical Assessment of Functional Annotation experiment (nell'ambito di ISMB – Intelligent Systems for Molecular Biology)
- ICANN - International Conference on Artificial Neural Networks
- SIAM International Conference on Data Mining
- ACM SIGKDD International Conference on Knowledge Discovery and Data Mining
- S+SSPR IAPR Joint International Workshops on Statistical + Structural and Syntactic Pattern Recognition
- International Conference on Pattern Recognition (ICPR)
- International Conference on Multiple Classifier Systems (MCS)

Invited speaker at international and national congresses and conferences

- Invited talk at the Statistical Workshop "Statistical approaches and validation in clustering: mixture models and nonparametric methods ", organized by the Department of Statistics of the University of Caen (France), entitled "Stability-based methods for cluster validation".
  (29-06-2007)
- Invited talk at CIBB 2008, FIFTH INTERNATIONAL MEETING ON COMPUTATIONAL INTELLIGENCE METHODS FOR BIOINFORMATICS AND BIOSTATISTICS IIASS - Vietri sul Mare, Salerno (Italy) entitled: "Unsupervised stability-based ensembles to discover reliable structures in complex bio-molecular data ".
  (04-10-2008)
- Invited talk at the Third Italian Workshop on Machine Learning and Data Mining - XIII AI * IA Symposium on Artificial Intelligence, Pisa December 2014, "Analysis of bio-molecular networks through semi-supervised graph-based learning methods".
  (10-12-2014)
- Invited talk at Fifth Italian Workshop on Machine Learning and Data Mining - XV AI * IA Symposium on Artificial Intelligence, Genoa, November 2016, "A hyper-ensemble approach for the genome-wide prediction of disease and trait-associated genetic variants".
  (28-11-2016)

- Invited speaker at the Interdisciplinary Aspects of Biomolecular Modeling Workshop "Machine Learning approaches for Modeling Complex Biomolecular Systems" (26-6-2019)
- Invited talk at the Annual Workshop in Bioinformatics and Genomics of the Catalan Society of Biology and the Bioinformatics Barcelona association, Barcelona, Spain "Machine Learning for Computational Biology and Precision Medicine" (17-12-2019)

Invited speaker at Italian and European universities and research centers.

I held several seminars on Machine Learning and Bioinformatics topics in various Italian and foreign Univeristies and research centers since 2005, including:

- Department of Computer Science della Royal Holloway, University of London;
- Department of Computer Science, Aalto University, Helsinki;
- Department of Computer Science, Aristotle University of Thessaloniki;
- Charitè Universitatzmedizin Berlin
- European Center for Living Technologies, Venezia;
- Departamento de Sistemas Informáticos y Computación Universitat Politècnica de València;
- Computational Genomics Department del Centro de Investigacion Prencipe Felipe, Valencia;
- School in Bioinformatics, University of Brno (Czech Republic);
- Computer Science and Artificial Intelligence Department, University of Granada;
- Dipartimento di Informatica, Università di Pisa;
- Dipartimento di Matematica e Informatica dell' Università di Palermo;
- Dipartimento di Ingegneria ed Elettronica (DIEE) Universita' di Cagliari;
- IIASS Istituto Internazionale Alti Studi Scientifici "E. R. Caianiello di Vietri (Salerno);
- Dipartimento di Informatica dell' Università degli Studi di Salerno;
- IST - Istituto Nazionale per la Ricerca sul Cancro, Genova.
- INT – Istituto Nazionale Tumori, Milano
- Dipartimento di Biologia e Genetica per le Scienze Mediche, Università degli Studi di Milano

International awards.

- The paper M. Notaro, M. Schubach, P.N. Robinson, G. Valentini. Prediction of Human Phenotype BMT Bioinformatics, vol. 18 (1), 2017 has been awarded as one of the 5 best articles of the year by the International Medical Informatics Association (IMIA) for the "Knowledge Representation and Management" section.

Patents
A signature of 21 SNPs (Singe Nucleotide Polymorphism) for the prediction of the response of patients suffering from multiple sclerosis to the Fingolimod drug is being patented.
This signature has been carried out using an algorithm and a software I developed and applied to the analysis of genomic data provided by the Department of Neurology of the San Raffaele Hospital in Milan.

Evaluation activities in the context of national and international competitive selection procedures.

- Research expert for the European Commission (number: EX2014D182522) for the evaluation of proposals, projects and programs submitted for evaluation for European funding (from 2014)
- Auditor registered with REPRISE (register of scientific experts established at the MIUR) for the “basic research” section: Scientific-disciplinary sectors: Informatics (INF / 01) Sectors ERC: Computer Science and Informatics: Informatics and Information Systems, computer science, scientific computing and intelligent systems (PE6), Bioinformatics, biocomputing, and DNA and

molecular computation (PE6_13), Diagnostic tools (eg genetic, imaging) (LS7_2), Computational biology (LS2_11) (from 2018).

- Reviewer for proposals submitted to the Netherlands Organization for Scientific Research (NWO):
  - Reviewer of the project "Collaborative experimentation for data mining" (2011-2012)
  - Reviewer of the project "Context-aware protein function prediction" (2012/13)
- Reviewer for proposals submitted to the Research Foundation - Flanders (Fonds Wetenschappelijk Onderzoek - Vlaanderen, FWO):
  - Reviewer of the project "COCO-MULT: Constrained Constructive Machine Learning for Multiple Targets" (2018)
- Reviewer for SIR (Scientific Independence of young Researchers) projects funded by MIUR (2014/15)
- Reviewer for projects of the "Bloodwise" program for the study of hematological tumors, funded by 14M Genomics, a company for the diagnostic spin-out of the Wellcome Trust Sanger Institute of Hinxton, UK.
  (from 2012 to 2016)

Formal assignment of teaching or research fellowships from qualified foreign or supranational universities and research institutes

- Research fellowship - Computer Science Dept. - Oregon State University - USA: development of ensemble methods based on bias-variance analysis.
  (from 01-03-2001 to 04-04-2001)
- Visiting researcher - Principe Felipe Research Center, Valencia, Spain (2009)
  (from 09/14/2009 to 10/15/2009)
- Visiting professor at the Computer Science Department of the Aristotle University of Thessaloniki for common research activities in the area of machine learning methods for multi-target prediction problems characterized by high dimensionality both in the input space and in the output space.
  (from 29-09-2014 to 10-28-2014)
- Visiting professor at the Computer Science Department of Royal Holloway, University of London for common research activities (development of methods for the integration and analysis of complex biomolecular networks) and for holding seminars to support the teaching of the master and PhD in Computer Science at Royal Holloway.
  (March 2015)
- Fellowship for visiting professors at the Computer Science Department of Aalto Univeristy (Helsinki) as part of the "Machine Learning for Metagenomics" project, funded by the Aalto Science Foundation of Helsinki (AScI Visiting Fellow Program 2014-15).
  (from 25-05-2015 to 25-06-2015)
- Visiting professor at the European Center for Living Technologies (ECLT). During the visit we strengthened joint research activities in the field of game theory and semi-supervised network-based methods for the analysis of bio-molecular networks.
  (from 01-02-2016 to 02/28/2016)
- Fellowship for visiting professors at the Computational Biology group of the Charité - Universitätsmedizin Berlin within the Funding program “Research Stays for University Academics and Scientists” funded by the DAAD (German Academic Exchange Service) for the project "Hierarchical ensemble methods for structured predictions in Biological Ontologies ".
  (from 01-03-2016 to 06/01/2016)
- Visiting professor at the Departamento de Ciencias de la Computación and Inteligencia Artificial de la Universidad de Granada. During the visit, common research activities were developed for the development of graph-based algorithms and their application to ranking problems of genes associated with specific pathologies, and to "repositioning" problems (alternative therapeutic use of drugs).
  (from 07-06-2016 to 06-28-2016)

Responsibility for studies and scientific research entrusted by public or private institutions

- Research assignment for the development and modeling of machine learning methods with applications in computational biology, as a member of the CNR affiliated project "Machine Learning Techniques for Modeling and Growing Up ", entrusted by IEIIT-CNR (2007-2008)
- Consultancy contract with the company SIGEDA of Milan for the development of ensemble methods based on error correction codes for the multi-class classification of data coming from an artificial olfactory system (electronic nose)
  (2007)
- Responsibility of a research project for the prediction with machine learning methods of quantitative tumor phenotypes in model organisms and for the study of susceptibility to pulmonary and hepatic carcinogenesis, entrusted by the research unit of Genetic Epidemiology and Pharmacogenomics of the National Cancer Institute (INT) of Milan.
  (from 01-04-2015 to 1-10-2017)
- Commission of the Centro S.Raffaele Foundation of Milan for the development of a predictive model of response to treatment with II level drugs for patients with Multiple Sclerosis.
  (from June 2016 to June 2018)

Design and development of software libraries publicly available to the scientific community.

- HEMDAG: Hierarchical Ensemble Methods for Directed Acyclic Graphs https://cran.r-project.org/web/packages/HEMDAG
- HyperSMURF: Hyper-Ensemble Smote Undersampled Random Forests (software library for supervised prediction with highly imbalanced big data).
  https://cran.r-project.org/web/packages/hyperSMURF
- RANKS: Ranking of Nodes with Kernelized Score Functions (software library for node label learning in graphs).
  https://cran.r-project.org/web/packages/RANKS
- COSNET: Cost Sensitive Network for node label prediction on graphs with highly unbalanced labels.
  http://www.bioconductor.org/packages/devel/bioc/html/COSNet.html;
- HCGene: an R package to support the hierarchical classification of genes.
  https://homes.di.unimi.it/valentini/SW/hcgene;
- Clusterv: an R package for cluster validation.
  https://homes.di.unimi.it/valentini/SW/clusterv;
- Mosclust: an R package for the discovery of significant structures in bio-molecular data.
  https://homes.di.unimi.it/valentini/SW/mosclust;
- NEURObjects a set of C++ library classes for neural networks development.
  http://homes.di.unimi.it/valentini/SW/NEURObjects;
- PerfMeas: an R package implementing different performance measure for classification and ranking tasks.
  http://cran.r-project.org/web/packages/PerfMeas;
- NetPreProc an R package that implements preprocessing and normalization methods for network-structured data.
  http://cran.r-project.org/web/packages/NetPreProc;
- Bionetdata an R data package that includes several examples of chemical and biological data networks.
  http://cran.r-project.org/web/packages/bionetdata.

Participation in national and international scientific societies and associations

I am a member of ISCB - International Society of Computational Biology, BITS - Italian Society of Bioinformatics, INNS (International Neural Network Society), Data Mining and Big Data Analytics Technique Committee (DMTC) - IEEE Computational Intelligence Society (IEEE-CIS), of the CINI AIIS laboratory, and of the scientific committee of the working group on Machine Learning and Data Mining, within AI * IA, Italian Association for Artificial Intelligence.

Academic teaching

- My academic teaching activity is articulated both through the institutional courses, held at the University of Milan and through the doctorate courses, masters and courses held at the University of Milan and in other universities especially in the field of Bioinformatics and Machine Learning.
  In particular, I held courses for the three-year and master's degree in Computer Science, for the degree in Digital Communication, for the Master Degree in Molecular Biotechnology and Bioinformatics and for the degree in Industrial and Environmental Biotechnology, as well as basic Computer science courses for various three-year degrees in Medicine. I have also held and still maintain courses in English.
- I am currently also responsible for the General Computer Science course for the degree in Science and Technology for the study and conservation of cultural heritage and information media (the Computer Science Department is a department associated with this degree course).
- I am member of the teaching board of the PhD in Computer Science of the University of Milan since 2008. In this context I am responsible for the Machine Learning for Genomic Medicine course.
- From the academic year 2018/19 I am a member of the coordinating committee for the Department of Computer Science of the Master in Bioinformatics and Functional Genomics of the University of Milan, carried out in collaboration with the National Institute of Molecular Genetics (INGM) and the Policlinico - 'University of Milan. In the academic year. 2017/18 I already held some lectures (Machine Learning for Personalized and Precision Medicine) within the same master.
- I am the scientific director for UNIMI of the EU Collaborative Doctoral Partnership in Genomics and Bioinformatics, funded by the European Commission in collaboration with the Joint Research Center (JRC) of the EU. The five-year plan of the collaborative doctorate between UNIMI and JRC is aimed at training researchers in the field of Genomics and Bioinformatics able both to carry out innovative research in this field and to scientifically support the political decisions of the Commission for Medical Genomics and Data Analytics in Health-care.

The following table summarizes the teaching assignments carried out in UNIMI starting from the a.a. 2004/05 to date:

Academic year	course	degree	Liv.
18/19	Machine Learning for Genomic Medicine	Ph.D. course in Computer Science	D
18/19	Bioinformatica	Informatica	M
18/19	Informatica Generale	Scienze e tecnologie per lo studio e la conservazione dei beni culturali e dei supporti dell'informazione	T
18/19	Machine Learning for Personalized and Precision Medicine	master in Bioinformatics and Functional Genomics UNIMI	Master
17/18	Bioinformatica	Informatica	M
17/18	Informazione Multimediale	Comunicazione Digitale	T
17/18	Informatica Generale	Scienze e tecnologie per lo studio e la conservazione dei beni culturali e dei supporti dell'informazione	T
16/17	Bioinformatica	Informatica	M
16/17	Bioinformatics Methods (in inglese)	Molecular Biotechnologies and Bioinformatics	M
16/17	Bioinformatica e Biostatistica	Biotecnologie Industriali ed Ambientali	T
14/15	Bioinformatica	Informatica	M
14/15	Metodi Bioinformatici	Biotecnologie Biomolecolari e Bioinformatica	M
14/15	Informatica Avanzata	Biotecnologie Industriali ed Ambientali	T
13/14	Bioinformatica	Informatica	M
13/14	Metodi Bioinformatici	Biotecnologie Biomolecolari e Bioinformatica	M

13/14	Informatica Avanzata	Biotecnologie Industriali ed Ambientali	T
12/13	Bioinformatica	Informatica	M
12/13	Metodi Bioinformatici	Biotecnologie Biomolecolari e Bioinformatica	M
12/13	Informatica Avanzata	Biotecnologie Industriali ed Ambientali	T
11/12	Bioinformatica	Informatica	M
11/12	Metodi Bioinformatici	Biotecnologie Biomolecolari e Bioinformatica	M
11/12	Informatica Avanzata	Biotecnologie Industriali ed Ambientali	T
10/11	Bioinformatica	Informatica	M
10/11	Metodi Bioinformatici	Biotecnologie Biomolecolari e Bioinformatica	M
10/11	Informatica Avanzata	Biotecnologie Industriali ed Ambientali	T
09/10	Bioinformatica	Informatica	M
09/10	Metodi Bioinformatici	Biotecnologie Biomolecolari e Bioinformatica	M
08/09	Linguaggi di programmazione per la bioinformatica	Genomica Funzionale e Bioinformatica	M
08/09	Bioinformatica	Informatica	M
08/09	Informatica applicata ai processi biotecnologici	Biotecnologie Industriale e Ambientali	T
07/08	Linguaggi di programmazione per la bioinformatica	Genomica Funzionale e Bioinformatica	M
07/08	Bioinformatica	Informatica	M
07/08	Informatica	Fisioterapia e Dietistica	T
06/07	Linguaggi di programmazione per la bioinformatica	Genomica Funzionale e Bioinformatica	M
06/07	Bioinformatica	Informatica	M
06/07	Informatica	Podologia ed Igiene Dentale	T
05/06	Linguaggi di programmazione per la bioinformatica	Genomica Funzionale e Bioinformatica	M
05/06	Bioinformatica	Informatica	M
05/06	Informatica	Podologia ed Igiene Dentale	T
04/05	Linguaggi di programmazione per la bioinformatica	Genomica Funzionale e Bioinformatica	M
04/05	Algoritmi per la Bioinformatica	Genomica Funzionale e Bioinformatica	M
04/05	Laboratorio di Bioinformatica	Biologia	T

M : “laurea magistrale” degree; T : “laurea” degree; D : Ph.D. course

- Previously I held the Functional Bioinformatics 1 course for the degree in Physics of the University of Genoa in the academic year 2002/03 and 2004/04 and the Neural Networks 1 exercises and the System Programming Laboratory (a.a 2001/02) for the degree course in Computer Science of the University of Genoa.
- I obtaiend teaching assignments for doctorate courses in various universities, including the University of Palermo (PhD in Computer Science and Mathematics), the University of Salerno (PhD in Computer Science), the University of Brno (Czech Republic) (Summer Doctoral School in Bioinformatics), the Computer Science Department of Royal Holloway, University of London.

- I was a lecturer and coordinator of courses on the R language for the analysis of bio-medical data for Ph.D. students, research fellows and researchers in various research centers, including:

  - "An R course for Oncology Bioinformatics", CINECA - Casalecchio di Reno, June 2008

  - "Course on the R language for the analysis of DNA microarray data with machine learning methods", IST Genova, November 2009

  - "Analysis of biomolecular data with R", Parco Tecnologico di Lodi, 21-23 January 2009

I held several seminars and Ph.D. courses in various Italian and European universities, including

- "Discovering significant structures in bio-molecular data" Ph.D. course - 3rd International Summer School on Computational Biology, Mikulov, Czech Republic, 13 -15 August 2007
- “Ensemble methods in bioinformatics” Ph.D. course - il Dottorato in Matematica ed Informatica all'Università di Palermo (22-26 giugno 2008)
- "Supervised Gene Set Analysis for the molecular characterization of patients: a machine learning approach", seminar at Centro de Investigacion Principe Felipe di Valencia, 6-10-2009
- "True Path Rule hierarchical ensembles for genome-wide gene function prediction: a machine learning approach", seminar dor the Ph.D. course in Computer Science de l'Universitat Politecnica de Valencia, 13-10-2009
- "Hierachical ensemble methods for gene function prediction" seminar for the Ph.D. course in Information Engineering, Dipartimento di Ingegneria Elettrica ed Elettronica (DIEE) dell'Università di Cagliari (May 2010)
- “True Path Rule and H-Bayes hierarchical cost-sensitive ensembles for gene function prediction” held at ’Istituto Internazionale per gli Alti Studi Scientifici “E.R. Caianiello”, Vietri sul mare (Salerno), 27 May 2010
- “True Path Rule hierarchical ensemble methods for gene function inference” seminar for the Ph.D. course in Copmuter Science, Dipartimento di Informatica, Università di Salerno, 6 june 2011
- “Stability-based methods for the assessment of clusters discovered in bio-molecular data” lecture for the Ph.D. course in Matematica ed Informatica all'Università di Palermo, june 2012
- “Ensemble methods for multi-target classification and regression”, seminar at Informatics department Aristotle University of Thessaloniki, Grecia, 14 october 2014
- “Scalable methods for bio-molecular network analysis and for structured prediction in biological ontologies” seminar at the 'European Center for Living technologies - Univ. Ca' Foscari, Venezia, 27 febbraio 2015
- "Analysis of bio-molecular networks through semi-supervised graph-based learning methods" seminar for the Ph.D. course in Computer Science, Computer Science Dept. Royal Holloway, University of London, 9 March 2015
- "Multi-label hierarchical prediction methods and their application to the automatic function prediction of proteins" seminario per il dottorato in Computer Science, Computer Science Dept. Royal Holloway, University of London, 10 March 2015
- “Scalable methods for the analysis of complex biomolecular networks”, seminar at Dipartimento di Informatica, Università di Salerno il 5 Maggio 2015
- "Semi-supervised graph-based learning methods for the analysis of bio-molecular networks" seminar for the Ph.D. course in Computer Science, Aalto University, Helsinki, Finland, 13 June 2015

- “Semi-supervised graph-kernel methods for disease gene prioritization”, seminar at Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada, Spagna, 15-6-2016
- “A Machine Learning and High Performance Computing tool to prioritize pathogenic variants in the human genome”, seminar for the Ph.D. course in Computer Science, Computer Science Dept. Royal Holloway, University of London, 26 Febbraio 2019

I was a member of various commissions for the final doctoral exam, and doctoral thesis reviewer for various Italian and European universities:

- Member of the final examination committee for the Doctorate in Computer Science of the University of Salerno
(April 2010)
- Member of the final examination committee for the PhD in Information Engineering of the Department of Electrical and Electronic Engineering (DIEE) of the University of Cagliari
(May 2010).
- Member of the final examination committee for the Doctorate in Computer Science of the University of Salerno
(February 2011)
- Member of the final examination committee for the doctorate in computer science of the DIBRIS University of Genoa
(May 2018)
- Member of the final exam for the PhD in Computer Science of the Freie Universitat Berlin (September 2018)
- Member of the final exam for the PhD in Computer Science of Royal Holloway - University of London (February 2019)
- Reviewer of the doctoral thesis of Isabel Segura Bedmar (Computer Science Department, University Carlos III of Madrid) "Application of Information Extraction techniques to pharmacological domain: Extracting drug-drug interactions" (2010)
- Reviewer of the doctoral thesis of Francesco Iorio (Doctorate in Computer Science - University of Salerno) "Automatic Discovery of Drug Mode of Action and Drug Repositioning from Gene Expression Data" (2010)
- Reviewer of the doctoral thesis of Luca Pinello (Department of Mathematics and Computer Science - University of Palermo) "Multi Layer Analysis" (2012)
- Reviewer of the doctoral thesis of Carmen Navarro (Departamento de Ciencias de la Computación and Inteligencia Artificial, Universidad de Granada, Espana) “Approach to Personalized Medicine through the development of new Artificial Intelligence methodologies "(2017).
- Reviewer of the doctoral thesis of Guido Zampieri (Ph.D. Course in Biosciences - Curriculum: Genetics, Genomics and Bioinformatics, University of Padua) "Prioritization of candidate disease genes via multi-omics data integration". (2017)
- Reviewer of the Master of Science thesis in Artificial Intelligence by Jeremy Borg (Faculty of ICT, University of Malta) "Improved Performance of Error Correcting Output Codes for Multiclass Classification" (2017)
- Reviewer of the doctoral thesis of Juan Caceres Silva (Department of Computer Science, Royal Holloway - University of London), "Charting disease gene propagation of disease phenotypic similarities". (2019)
- Reviewer of the doctoral thesis of Samuele Fiorini (Department of Computer Science, Bioengineering, Robotics and Systems Engineering University of Genoa) "Challenges in biomedical data science: data-driven solutions to clinical questions". (2018)
- Reviewer of Max Schubach's doctoral thesis (Department of Mathematics and Computer Science, Freie Universitat Berlin) "Learning the Non-Coding Genome" (2018)

- Reviewer of the doctoral thesis of Juan Caceres Silva (Department of Computer Science, Royal Holloway - University of London), "Network Medicine Characterisation of Genetic Disorders by Propagation of Disease Phenotypic Similarities". (2019)

Supplementary teaching and student support activities.

- Tutor, thesis supervisor or co-supervisor for the following PhD students for the Doctorate in Computer Science, University of Milan:

  1) Francesca Ruffino: “Supervised Learning Methods for the Analysis of Gene Expression Data”
  2) Raffaella Folgieri: “Ensembles based on Random Projection for gene expression data analysis”
  3) Roberto Avogadri: “Unsupervised clustering methods for high-dimensional data anslysis”
  4) Francesco Saccà (dottorato in Matematica e Statistica per le Scienza Computazionali): “Problemi di Clustering con vincoli: algoritmi e complessita'”
  5) Marco Frasca: “Graph-based approaches for imbalanced data in functional genomics”
  6) Rajab Ali Keshavarz Emami: ”Machine learning methods for the prediction of epileptic seizures”
  7) Marco Notaro: “Hierarchical Ensemble Methods for Ontology-based Predictions in Computational Biology”
  8) Alessandro Petrini: “Machine learning methods for the the prediction of pathogenic variants in non coding regions of the human genome”
- Supervisor of the following research fellows: 1) Matteo Re; 2) Marco Frasca
- Supervisor of the graduate student Jessica Gliozzo, winner of a scholarship abroad at the Dept. of Computer Science of Royal Holloway - University of London, for the purposes of preparing the Master's thesis a.a. 2015/2016, funded by the University of Milan.
- Speaker or co-supervisor of more than 40 degree theses (triennial or “magistrale” or of the old system), in Machine Learning and Bioinformatics, for the degrees in Computer Science, Information and Communication Technologies, Molecular Biotechnologies and Bioinformatics, Mathematics, from 2005 to date.
- I have worked as a tutor for dozens of students in the undergraduate and graduate degrees in Computer Science.

Institutional activities

- UNIMI contact person for Area 06 - Computer Science for the organization and census, in collaboration with UniMITT, of the research activities in the IT area with implications in the bio-medical and bio-technological fields.
  from 2005 to 2007).
- Associate Member of the CNR - Institute of Electronics and Information and Telecommunications Engineering for the project "Machine Learning Techniques for Modeling and Growing Up" (2007-2008)
- Member of the Doctoral College in Computer Science, University of Milan since 2008
- Member of the evaluation commission for the Faculty of Science of the University of Milan for the ERASMUS INTENSIVE PROGRAM (IP) a.a. 2009/2010 "Interdisciplinary approaches to microarray data analysis" organized by the University of Helsinki, Warwick, Naples (Federico II) and Milan.
- Member of the transfer commission and responsible for transfers to the degree course in Computer Science F1X of the University of Milan from other degree courses since the academic year 2010/11 to 2014/15.
- Member of the committee of the Computer Science department for educational and scientific exchange with the University of Beijing and for the technology transfer with the Chinese genomics

company Rose Genomics (Beijing) as part of an agreement between the UNIMI Foundation, UNIMI and Rose Genomics (from 2018)

- Member of the Data Mining and Big Data Analytics Technique Committee (DMTC) - IEEE Computational Intelligence Society (IEEE-CIS)
  (from 2015)
- Member of the scientific committee of the working group on Machine Learning and Data Mining, within AI * IA, Italian Association for Artificial Intelligence
  (from 2015)
- Member of the IEEE Computational Intelligence Society 2017 PhD Thesis Award committee (from 2018)

Publications

International journals

R52. M. Frasca, G. Grossi, J. Gliozzo, M. Mesiti, M. Notaro, P. Perlasca, A. Petrini and G. Valentini A GPU-based algorithm for fast node label learning in large and unbalanced biomolecular networks, *BMC Bioinformatics 19:Suppl 10 Oct. 15, 2018* *doi.org/10.1186/s12859-018-2301-4*

R51. S. Vascon, M. Frasca, R. Tripodi, G. Valentini, M. Pelillo Protein Function Prediction as a Graph-Transduction Game, *Pattern Recognition Letters (in press), 2018* doi.org/10.1016/j.patrec.2018.04.002

R50. M. Notaro, M. Schubach, P.N. Robinson, G. Valentini Prediction of Human Phenotype Ontology terms by means of hierarchical ensemble methods, *BMC Bioinformatics, vol. 18 (1), 2017* doi.org/10.1186/s12859-017-1854-y

R49. M. Schubach, M. Re, P.N. Robinson and G. Valentini Imbalance-Aware Machine Learning for Predicting Rare and Common Disease-Associated Non-Coding Variants, *Scientific Reports, Nature Publishing*, 7:2959, 2017. doi.org/10.1038/s41598-017-03011-5

R48. D. Smedley, M, Schubach, J. Jacobsen, S. Kohler, T. Zemojtel, M. Spielmann, M. Jager, H. Hochheiser, N. Washington, J. McMurry, M. Haendel, C. Mungall, S. Lewis, T. Groza, G. Valentini and P.N. Robinson A Whole-Genome Analysis Framework for Effective Identification of Pathogenic Regulatory Variants in Mendelian Disease, *The American Journal of Human Genetics*, 99:3, pp.595--606, September 2016. doi.org/10.1016/j.ajhg.2016.07.005

R47. Y. Jiang, P. Oron, ... G. Valentini, ... I. Friedberg and P. Radivojac An expanded evaluation of protein function prediction methods shows an improvement in accuracy, *Genome Biology*, 17:184 September 2016. doi.org/10.1186/s13059-016-1037-6

R46. G. Valentini, G. Armano, M. Frasca, J. Lin, M. Mesiti and M. Re RANKS: a flexible tool for node label ranking and classification in biological networks, *Bioinformatics*, 32(18), September 2016. doi:10.1093/bioinformatics/btw235

R45. M. Frasca, S.Bassis, G. Valentini Learning node labels with multi-category Hopfield networks, *Neural Computing and Applications*, 27(6), pp 1677-1692, 2016 doi:10.1007/s00521-015-1965-1

R44. M. Frasca, G. Valentini COSNet: an R package for label prediction in unbalanced biological networks, *Neurocompting*, 2016. doi:10.1016/j.neucom.2015.11.096

R43. M. Frasca, A. Bertoni, G. Valentini UNIPred: Unbalance-aware Network Integration and Prediction of protein functions, *Journal of Computational Biology*, 22(12): 1057-1074, 2015. doi:10.1089/cmb.2014.0110

R42. M. Mesiti, M. Re, G. Valentini Think globally and solve locally: secondary memory-based network learning for automated multi-species function prediction, *GigaScience*, 3:5, 2014

R41. G. Valentini, A. Paccanaro, H. Caniza, A. Romero, M. Re, An extensive analysis of disease-gene associations using network integration and fast kernel-based gene prioritization methods, *Artificial Intelligence in Medicine*, Volume 61, Issue 2, pages 63-78, June 2014

R40. H. Caniza, A. Romero, S. Heron, H. Yang, A. Devoto, M. Frasca, M. Mesiti, G. Valentini, A. Paccanaro, GOssTo: a user-friendly stand-alone and web tool for calculating semantic similarities on the Gene Ontology, *Bioinformatics*, Vol. 30 no. 15, pages 2235-2236, 2014

R39. G. Valentini, Hierarchical Ensemble Methods for Protein Function Prediction, *ISRN Bioinformatics*, vol. 2014, Article ID 901419, 34 pages, 2014

R38. M. Re, and G. Valentini, Network-based Drug Ranking and Repositioning with respect to DrugBank Therapeutic Categories,

IEEE ACM Transactions on Computational Biology and Bioinformatics 10(6), pp. 1359-1371, Nov-Dec 2013 IEEE link Supplemental Material

R37. I. Cattinelli, G. Valentini, E. Paulesu, A. Borghese A Novel Approach to the Problem of Non-uniqueness of the Solution in Hierarchical Clustering ,
IEEE Transactions on Neural Networks and Learning Systems 24(7) pp.1166-1173, July 2013

R36. M. Frasca, A. Bertoni, M. Re, and G. Valentini, A neural network algorithm for semi-supervised node label learning from unbalanced data,
Neural Networks 43, pp.84-98, July 2013

R35. M. Re, M. Mesiti and G. Valentini, A Fast Ranking Algorithm for Predicting Gene Functions in Biomolecular Networks,
IEEE ACM Transactions on Computational Biology and Bioinformatics 9(6) pp. 1812-1818, 2012.

R34. A. Beghini, F. Corlazzoli, L. Del Giacco, M. Re, F. Lazzaroni, M. Brioschi, G. Valentini, F. Ferrazzi, A. Ghilardi, M. Righi, M. Turrini, M. Mignardi, C. Cesana, V. Bronte, M. Nilsson, E. Morra and R. Cairoli, Regeneration-associated Wnt signaling is activated in long-term reconstituting AC133bright acute myeloid leukemia cells,
Neoplasia 14:12, pp. 1236-1248, 2012

R33. M. Re and G. Valentini Cancer module genes ranking using kernelized score functions
BMC Bioinformatics 13 (Suppl 14): S3, 2012.

R32. N. Cesa-Bianchi, M. Re, G. Valentini, Synergy of multi-label hierarchical ensembles, data fusion, and cost-sensitive methods for gene functional inference,
Machine Learning, vol.88(1), pp. 209-241, 2012.

R31. M. Re, M. Mesiti, G. Valentini, Drug repositioning through pharmacological spaces integration based on networks projection,
EMBnet.journal, vol 18, Supplement A, pp.30-31, 2012.

R30. M. Frasca, A. Bertoni, G. Valentini, Regularized Network-Based Algorithm for Predicting Gene Functions with High-Imbalanced Data,
EMBnet.journal, vol 18, Supplement A, pp.41-42, 2012.

R29. G. Valentini, True Path Rule hierarchical ensembles for genome-wide gene function prediction,
IEEE ACM Transactions on Computational Biology and Bioinformatics, vol.8 n.3 pp. 832-847, 2011.

R28. M. Muselli, A. Bertoni, M. Frasca, A. Beghini, F. Ruffino, and G. Valentini, A mathematical model for the validation of gene selection methods,
IEEE ACM Transactions on Computational Biology and Bioinformatics, vol.8 n.5 pp. 1385-1392, 2011.

R27. M. Re, G. Valentini, Noise tolerance of Multiple Classifier Systems in data integration-based gene function prediction,
Journal of Integrative Bioinformatics, 7(3):139, 2010

R26. M. Re, G. Valentini, Simple ensemble methods are competitive with state-of-the-art data integration methods for gene function prediction
Journal of Machine Learning Research, W&C Proceedings, vol.8: Machine Learning in Systems Biology, pp. 98-111, 2010.

R25. N. Cesa-Bianchi, G. Valentini, Hierarchical cost-sensitive algorithms for genome-wide gene function prediction,
Journal of Machine Learning Research, W&C Proceedings, vol.8: Machine Learning in Systems Biology, pp.14-29, 2010.

R24. M. Re, G. Valentini, Integration of heterogeneous data sources for gene function prediction using Decision Templates and ensembles of learning machines,
Neurocomputing, 73:7-9 pp. 1533-37, 2010 doi:10.1016/j.neucom.2009.12.012

R23. M. Mesiti, E. Jimenez-Ruiz, I. Sanz, R. Berlanga-Llavori, P. Perlasca, G. Valentini and D. Manset, XML-Based Approaches for the Integration of Heterogeneous Bio-Molecular Data
BMC Bioinformatics 10:(S12)S7, 2009

R22. R. Avogadri, M. Brioschi, F. Ferrazzi, M. Re, A. Beghini, and G. Valentini, A stability-based algorithm to validate hierarchical clusters of genes,
International Journal of Knowledge Engineering and Soft Data Paradigms, 1(4), pp. 318-330, 2009

R21. G.Valentini, R.Tagliaferri, F.Masulli, Computational Intelligence and Machine Learning in Bioinformatics
Artificial Intelligence in Medicine 45(2), pp. 91-96, 2009

R20. R. Avogadri, G.Valentini, Fuzzy ensemble clustering based on random projections for DNA microarray data analysis
Artificial Intelligence in Medicine 45(2), pp. 173-183, 2009

R19. G.Pavesi, G.Valentini, Classification of co-expressed genes from DNA regulatory regions,
Information Fusion 10(3), pp. 233-241, 2009

R18. A. Bertoni, G.Valentini, Discovering multi-level structures in bio-molecular data through the Bernstein inequality
BMC Bioinformatics **9**(Suppl 2)**:**S4, 2008

R17. G.Valentini, N. Cesa-Bianchi, HCGene: a software tool to support the hierarchical classification of genes,
Bioinformatics, 24(5), pp. 729-731, 2008.

R16. F. Ruffino, M. Muselli, G.Valentini, Gene expression modelling through positive Boolean functions,
International Journal of Approximate Reasoning, 47(1), pp. 97-108, 2008.

R15. A.Bertoni, G.Valentini, Model order selection for biomolecular data clustering,
BMC Bioinformatics, vol.8, Suppl.3, 2007.

R14. G.Valentini, Mosclust: a software library for discovering significant structures in bio-molecular data.
Bioinformatics 23(3):387-389, 2007.

R13. G. Valentini, F.Ruffino, Characterization of Lung tumor subtypes through gene expression cluster validity assessment,
RAIRO - Theoretical Informatics and Applications, 40:163-176, 2006.

R12. A.Bertoni, G. Valentini, Randomized maps for assessing the reliability of patients clusters in DNA microarray data analyses,
Artificial Intelligence in Medicine 37(2):85-109 *2006.*

R11. G.Valentini, Clusterv: a tool for assessing the reliability of clusters discovered in DNA microarray data,
Bioinformatics 22(3):369-370, 2006.

R10. G.Valentini, An experimental bias-variance analysis of SVM ensembles based on resampling techniques,
IEEE Transactions on Systems, Man and Cybernetics, Part B vol.35(6) pp. 1252-1271, 2005

R9. P. Campadelli, E. Casiraghi, G.Valentini, Support Vector Machines for candidate nodules classification, *Neurocomputing* vol.68 pp. 281-289, 2005 Science Direct access

R8. A. Bertoni, R. Folgieri, G. Valentini, Bio-molecular cancer prediction with random subspace ensembles of Support Vector Machines,
Neurocomputing vol. 63C pp. 535-539, 2005 Science Direct access

R7. G. Valentini, T. G. Dietterich, Bias-variance analysis of Support Vector Machines for the development of SVM-based ensemble methods,
Journal of Machine Learning Research, 5(Jul) pp. 725--775, 2004, MIT Press, JMLR link

R6. F. Masulli, G. Valentini, An experimental analysis of the dependence among codeword bit errors in ECOC learning machines.
Neurocomputing 57 pp. 189-214, 2004, science direct link

R5. G. Valentini, M. Muselli and F. Ruffino, Cancer recognition with bagged ensembles of Support Vector Machines,
Neurocomputing 56 pp. 461-466, 2004.

R4. F. Masulli, G. Valentini, Effectiveness of output coding decomposition schemes in ensemble and monolithic learning machines.
Pattern Analysis and Applications 6 pp. 285-300, 2003.

R3. G. Valentini, Gene expression data analysis of human lymphoma using Support Vector Machines and Output Coding ensembles.
Artificial Intelligence in Medicine 26(3) pp 283-306, 2002

R2. G. Valentini, F. Masulli, NEURObjects: an object-oriented library for neural network development, *Neurocomputing* 48(1-4) pp. 623-646 , 2002.

R1. M. Pardo, G. Sberveglieri, A.Taroni, F. Masulli, G. Valentini Decompositive classification models for electronic noses.
Anal. Chim. Acta (446) pp. 223-232, 2001.

Editor of books

E5. O. Okun, G. Valentini, M. Re (eds.), Ensembles in Machine Learning Applications, *Studies in Computational Intelligence*, vol. 373 Springer, ISBN: 978-3-642-22909-1, 2011.

E4. O. Okun, M. Re, G. Valentini (eds.), Proceedings of the the Third Workshop on Supervised and Unsupervised Ensemble Methods and Their Applications (SUEMA), European Conference on Machine Learning, Barcelona, Spain, 2010.

E3. O. Okun, G. Valentini (eds.), Applications of Supervised and Unsupervised Ensemble Methods, *Studies in Computational Intelligence*, vol. 245 Springer, ISBN: 978-3-642-03998-0, 2010.

E2. O. Okun, G. Valentini (eds.), Proceedings of the the Second Workshop on Supervised and Unsupervised Ensemble Methods and Their Applications (SUEMA), European Conference on Artificial Intelligence, University of Patras, Greece, ISBN: 978-960-89282-2-0, 2008.

E1. O. Okun, G. Valentini (eds.), Supervised and Unsupervised Ensemble Methods and their Applications, *Studies in ComputationalIntelligence*, vol. 126 Springer, ISBN: 978-3-540-78980-2, 2008.

Proceedings of international conferences and book chapters

C84. A. Cuzzocrea, L. Cappelletti, G. Valentini A neural model for the prediction of pathogenic genomic variants in Mendelian diseases, 1st International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI), Barcelona, 2019

C83. M. Notaro, M. Schubach, M.Frasca, M. Mesiti, P.N. Robinson, G. Valentini Ensembling Descendant Term Classifiers to Improve Gene - Abnormal Phenotype Predictions, *Lecture Notes in Bioinformatics*, vol. 10834, pp. 70-80, 2019

C82. Marco Frasca, Jean Fred Fontaine, Giorgio Valentini, Marco Mesiti, Marco Notaro, Dario Malchiodi et al. Disease-Genes Must Guide Data Source Integration in the Gene Prioritization Process *Lecture Notes in Bioinformatics*, vol. 10834, pp. 60-69, 2019

C81. C. T. Ba, E. Casiraghi, M. Frasca, J. Gliozzo, G. Grossi, M. Mesiti, M. Notaro, P. Perlasca, A. Petrini, M. Re and G. Valentini, A Graphical Tool for the Exploration and Visual Analysis of Biomolecular Networks, *CIBB 2018 - Computational Intelligence methods for Bioinformatics and Biostatistics*, Lisboa (Portogallo) (accepted)

C80. C. Cano, M. Re, M. Verbeni, M. Notaro, G. Valentini, A. Blanco, Characterization of coding and non-coding RNA interactions through topological descriptors, *CIBB 2018 - Computational Intelligence methods for Bioinformatics and Biostatistics*, Lisboa (Portogallo) (accepted)

C79. M. Notaro, M. Schubach, P.N. Robinson, G. Valentini Predicting new relationships between genes and Human Phenotype Ontology terms, presented at the *26th International Conference on Intelligent Systems for Molecular Biology (ISMB)*, Chicago, 2018.

C78. A. Petrini, M. Schubach, M. Re, M. Frasca, M. Mesiti, G. Grossi, T. Castrignano', P.N. Robinson, G. Valentini Parameters tuning boosts hyperSMURF predictions of rare deleterious non-coding genetic variants,
PeerJ Preprints 5:e3185v1, 2017 presented at Methods, tools & platforms for Personalized Medicine in the Big Data Era - NETTAB 2017, Palermo, Italy

C77. M. Schubach, M. Re, P.N. Robinson, G. Valentini Variant relevance prediction in extremely imbalanced training sets,

F1000Research 2017, 6(ISCB Comm J):1392 (poster) (doi: 10.7490/f1000research.1114637.1), presented at the 25th International Conference on Intelligent Systems for Molecular Biology (ISMB), Prague 2017

C76. M. Notaro, M. Schubach, P.N. Robinson, G. Valentini Ensembling Descendant Term Classifiers to Improve Gene - Abnormal Phenotype Predictions,
CIBB 2017, The 14th International Conference on Bioinformatics and Biostatistics, Cagliari, Italy, 2017.

C75. M. Frasca, J.F. Fontaine, G. Valentini, M. Mesiti, M. Notaro, D. Malchiodi and M.A. Andrade-Navarro Disease Genes must Guide Data Source Integration in the Gene Prioritization Process ,
CIBB 2017, The 14th International Conference on Bioinformatics and Biostatistics, Cagliari, Italy, 2017.

C74. J. Lin, M. Mesiti, M. Re and G. Valentini Within network learning on big graphs using secondary memory-based random walk kernels,
Complex Networks & Their Applications V: Proceedings of the 5th International Workshop on Complex Networks and their Applications (COMPLEX NETWORKS 2016), Studies in Computational Intelligence, Springer, pp. 235-245, 2017, doi.org/10.1007/978-3-319-50901-3_19

C73. P. Perlasca, G. Valentini, M. Frasca, M. Mesiti Multi-species Protein Function Prediction: Towards Web-based Visual Analytics,
Proceedings of the 18th International Conference on Information Integration and Web-based Applications & Services, Singapore, ACM, New York, USA pp. 1-5, 2016.
doi.org/10.1145/3011141.3011222

C72. H. Su, G. Valentini, S. Szedmak and J. Rousu Transport Protein Classification through Structured Prediction and Multiple Kernel Learning ,
NIPS Workshop on Machine Learning in Computational Biology (MLCB) & Machine Learning in Systems Biology (MLSB) 2015 - Montreal, Canada, December 2015

C71. P.N. Robinson, M.Frasca, S. Kohler, M. Notaro, M. Re, G. Valentini, A hierarchical ensemble method for DAG-structured taxonomies ,
Multiple Classifier Systems - MCS 2015 - Gunzburg, Germany *Lecture Notes in Computer Science*, vol. 9132, pp. 15-36, Springer, 2015

C70. G. Valentini, S. Kohler, M. Re, M. Notaro, P.N. Robinson, Prediction of human gene - phenotype associations by exploiting the hierarchical structure of the Human Phenotype Ontology,
3rd International Work-Conference on Bioinformatics and Biomedical Engineering - IWBBIO 2015, Granada, Spain *Lecture Notes in Bioinformatics*, vol. 9043, pp. 66-77, Springer, 2015

C69. M. Re, M.Mesiti, G. Valentini, An automated pipeline for multi-species protein function prediction from the UniProt Knowledgebase,
Automated Function Prediction SIG 2014 - ISMB 2014, Boston, USA

C68. M. Re, M.Mesiti, G. Valentini, On the Automated Function Prediction of Big Multi-Species Networks,
Network Biology SIG 2014 - ISMB 2014, Boston, USA

C67. M.Frasca, A. Bertoni, G. Valentini An unbalance-aware network integration method for gene function prediction,
MLSB 2013 - Machine Learning for Systems Biology, Berlin, 2013

C66. G. Valentini, A. Paccanaro, H. C. Vierci, A. E. Romero, M. Re, Network integration boosts disease gene prioritization,
Network Biology SIG 2013 - ISMB 2013, Berlin

C65. M.Mesiti, M. Re, G. Valentini Scalable Network-based Learning Methods for Automated Function Prediction based on the Neo4j Graph-database,
Automated Function Prediction SIG 2013 - ISMB 2013, Berlin

C64. H. C. Vierci, A. E. Romero, S. Heron, H. Yang, M. Frasca, M. Mesiti, G. Valentini and A. Paccanaro GOssTo & GOssToWeb: user-friendly tools for calculating semantic similarities on the Gene Ontology,
Bio-Ontologies SIG 2013 - ISMB 2013, Berlin

C63. M. Re, M.Mesiti, G. Valentini Comparison of early and late omics data integration for cancer modules gene ranking ,
NETTAB 2012 Workshop on Integrated Bio-Search, Como 14-16 November, 2012.

C62. M. Re and G. Valentini Random walking on functional interaction networks to rank genes involved in cancer
2nd Artificial Intelligence Applications in Biomedicine Workshop, in: AIAI 2012 - Artificial Intelligence Applications and Innovations, pp. 66-75, *IFIP AICT Series*, Springer, 2012

C61. M. Re, G. Valentini Large Scale Ranking and Repositioning of Drugs with Respect to DrugBank Therapeutic Categories, slides
In: L. Bleris et al. (Eds.): International Symposium on Bioinformatics Research and Applications (ISBRA 2012), Dallas, USA, *Lecture Notes in Bioinformatics* vol.7292, pp. 225-236, Springer, 2012.

C60. M. Re, G. Valentini, Ensemble methods: a review,
In: Advances in Machine Learning and Data Mining for Astronomy, Chapman & Hall Data Mining and Knowledge Discovery Series, Chap. 26, pp. 563-594, 2012.

C59. M. Re, G. Valentini Genes prioritization with respect to Cancer Gene Modules using functional interaction network data , *NETTAB 2011* Workshop on Clinical Bioinformatics, Pavia 12-14 October, 2011.

C58. A. Bertoni, M. Frasca, G. Valentini COSNet: a Cost Sensitive Neural Network for Semi-supervised Learning in Graphs.,
In: "Machine Learning and Knowledge Discovery in Databases". European Conference, ECML PKDD 2011, Athens, Greece, Proceedings, Part I, *Lecture Notes in Artificial Intelligence*, vol. 6911, pp.219-234, Springer, 2011.

C57. A. Rozza, G. Lombardi, M. Re, E. Casiraghi, G. Valentini and P. Campadelli A Novel Ensemble Technique for Protein Subcellular Location Prediction ,
In: "Ensembles in Machine Learning Applications", *Studies in Computational Intelligence* vol. 373, pp. 151-167, Springer, 2011

C56. M. Frasca, A. Bertoni, G. Valentini A cost-sensitive neural algorithm to predict gene functions using large biological networks.,
Network Biology SIG: On the Analysis and Visualization of Networks in Biology, ISMB 2011, Wien

C55. A. Bertoni, M. Re, F. Sacca, G. Valentini Identification of promoter regions in genomic sequences by 1-dimensional constraint clustering,
Frontiers in Artificial Intelligence and Applications, vol. 234, Neural Nets WIRN11 - Proceedings, pp. 162-169, 2011.

C54. A. Rozza, G. Lombardi, M. Re, E. Casiraghi, and G. Valentini, DDAG K-TIPCAC: an ensemble method for protein subcellular localization,
Proc. of the Third Edition of SUEMA, pp. 75-84 , ECML, Barcelona, Spain, 2010.

C53. N. Cesa-Bianchi, M. Re, G. Valentini, Functional Inference in FunCat through the Combination of Hierarchical Ensembles with Data Fusion Methods,
ICML Workshop on learning from Multi-Label Data MLD'10 , Haifa, Israel, pp.13-20, 2010

C52. A. Bertoni, M. Frasca, G. Grossi, G. Valentini, Learning functional linkage networks with a cost-sensitive approach ,
Neural Networks - WIRN 2010, IOS Press, pp. 52-61, 2010

C51. M. Re, G. Valentini, An experimental comparison of Hierarchical Bayes and True Path Rule ensembles for protein function prediction,
In: (N. El Gayar, J. Kittler and F. Roli, Eds) Nineth International Workshop on Multiple Classifier Systems MCS 2010, *Lecture Notes in Computer Science*, vol. 5997, pp. 294-303, Springer, 2010.

C50. N. Cesa-Bianchi, G. Valentini, Hierarchical cost-sensitive algorithms for genome-wide gene function prediction,
Machine Learning in Systems Biology, Proceedings of the Third international workshop, Ljubljana, Slovenia, pp. 25-34, 2009.

C49 M. Re, G. Valentini, Simple ensemble methods are competitive with state-of-the-art data integration methods for gene function prediction,
Machine Learning in Systems Biology, Proceedings of the Third international workshop, Ljubljana, Slovenia, pp. 95-104, 2009.

C48 G. Valentini, M. Re, Weighted True Path Rule: a multilabel hierarchical algorithm for gene function prediction,

MLD-ECML 2009, 1st International Workshop on learning from Multi-Label Data, Bled, Slovenia, pp. 133-146, 2009.

C47. M. Re, G. Valentini, Predicting gene expression from heterogeneous data,
CIBB 2009, The Sixth International Conference on Bioinformatics and Biostatistics, Genova, Italy, 2009.

C46. M. Re, G. Valentini, Comparing early and late data fusion methods for gene function prediction, Neural Nets WIRN09 - Proceedings of the 19th Italian Workshop on Neural Nets, Vietri sul Mare, Salerno, Italy, 2009, *Frontiers in Artificial Intelligence and Applications* vol. 204, pp. 197-207, IOS Press, 2009.

C45. M. Re, G. Valentini, Ensemble based Data Fusion for Gene Function Prediction,
In: (J. Kittler, J. Benediktsson, F. Roli, Eds.) Eighth International Workshop on Multiple Classifier Systems MCS 2009, *Lecture Notes in Computer Science,* vol.5519 pp.448-457, Springer 2009.

C44. G. Valentini, True Path Rule Hierarchical Ensembles,
In: (J. Kittler, J. Benediktsson, F. Roli, Eds.) Eighth International Workshop on Multiple Classifier Systems MCS 2009, *Lecture Notes in Computer Science*, vol.5519 pp.232-241, Springer 2009.

C43. O. Okun, G. Valentini, H. Priisalu, Exploring the link between bolstered classification error and dataset complexity for gene expression based cancer classification,
In T. Maeda, ed., *New Signal Processing Research*, Nova Publishers, pp. 249-278, 2009.

C42. A. Bertoni, G. Valentini, Unsupervised stability-based ensembles to discover reliable structures in complex bio-molecular data,
in: Proc. CIBB 2008, The Fifth International Conference on Bioinformatics and Biostatistics, *Lecture Notes in Computer Science*, vol. 5488 pp. 25-43, Springer, 2009.

C41. M. Re, G. Valentini, Prediction of gene function using ensembles of SVMs and heterogeneous data sources,
in: Applications of supervised and unsupervised ensemble methods, *Computational Intelligence Series*, vol.245, pp. 79-91, Springer, 2010.

C40. M. Mesiti, E. J. Ruiz, I. Sanz, R. Berlanga, G. Valentini, P Perlasca, D. Manset, Data Integration and Opportunities in Biological XML Data Management,
in: E. Pardede (editor): Open and Novel Issues in XML Database Applications: Future Directions and Advanced Technologies, Information Science, pp. 263-286, 2009.

C39. R. Avogadri, M. Brioschi, F. Ruffino, F. Ferrazzi, A. Beghini and G. Valentini
An algorithm to assess the reliability of hierarchical clusters in gene expression data,
in: I. Lovrek, R. J. Howlett, L. C. Jain (Eds.): Knowledge-Based Intelligent Information and Engineering Systems, 12th International Conference, KES 2008, Zagreb, Croatia, September 3-5, 2008, Proceedings, Part III. *Lecture Notes in Computer Science*, vol.5179 pp. 764-770, Springer 2008.

C38. M. Mesiti, E. J. Ruiz, I. Sanz, R. Berlanga, G. Valentini, P Perlasca, D. Manset
XML-based approaches for the integration of heterogeneous bio-molecular data,
NETTAB 2008 workshop on: "Bioinformatics Methods for Biomedical Complex System Applications", 2008.

C37. O. Okun, G.Valentini, Dataset Complexity Can Help to Generate Accurate Ensembles of K-Nearest Neighbors,
IEEE International Joint Conference on Neural Networks - IJCNN 2008 (IEEE World Congress on Computational Intelligence), pp. 450-457, 2008.

C36. R. Avogadri, G.Valentini, Ensemble Clustering with a Fuzzy Approach,
in: "Supervised and Unsupervised Ensemble Methods and their Applications", *Studies in Computational Intelligence*, vol. 126, Springer, 2008.

C35. R. Tagliaferri, A. Bertoni, F. Iorio, G. Miele, F. Napolitano, G. Raiconi and G. Valentini A Review on clustering and visualization methodologies for Genomic data analysis (extended abstract)
Workshop on Computational Intelligence approaches for the analysis of Bioinformatics data, IJCNN 2007, Orlando, USA, 2007.

C34. A. Bertoni, G.Valentini, Discovering Significant Structures in Clustered Bio-molecular Data Through the Bernstein Inequality

Knowledge-Based Intelligent Information and Engineering Systems, 11th International Conference, KES 2007, *Lecture Notes in Computer Science*, vol. 4694 pp. 886-891, 2007.

C33. R. Avogadri, G.Valentini, Fuzzy ensemble clustering for DNA microarray data analysis,
CIBB 2007, The Fourth International Conference on Bioinformatics and Biostatistics, *Lecture Notes in Computer Science*, vol. 4578, pp.537-543, 2007

C32. R. Avogadri, G.Valentini, An unsupervised fuzzy ensemble algorithmic scheme for gene expression data analysis
NETTAB 2007 workshop on a Semantic Web for Bioinformatics, Pisa, Italy, 2007.

C31. A.Bertoni, G.Valentini, Randomized Embedding Cluster Ensembles for gene expression data analysis, *SETIT 2007* - IEEE International Conf. on Sciences of Electronic, Technologies of Information and Telecommunications, Hammamet, Tunisia, 2007.

C30. F. Ruffino, M. Muselli, G. Valentini, Modeling gene expression data via positive Boolean functions, *NETTAB 2006 workshop* on Distributed Applications, Web Services, Tools and GRID Infrastructures for Bioinformatics, S.Margherita di Pula 10-13 July, Italy, 2006.

C29. A.Bertoni, G. Valentini, Model order selection for clustered bio-molecular data,
In: *Probabilistic Modeling and Machine Learning in Structural and Systems Biology*, J. Rousu, S. Kaski and E. Ukkonen (Eds.), Tuusula, Finland, 17-18 June, pp. 85-90, Helsinki University Printing House, 2006, slides

C28. A.Bertoni, G. Valentini, Ensembles Based on Random Projections to Improve the Accuracy of Clustering Algorithms,
Neural Nets, WIRN 2005, *Lecture Notes in Computer Science*, vol. 3931, pp. 31-37, 2006.

C27. B. Apolloni, G. Valentini, A.Brega, BICA and Random Subspace ensembles for DNA microarray-based diagnosis,
CIBB 2006 - International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics In Proc. of 7th International FLINS Conference on Applied Artificial Intelligence pp. 623-631, World Scientific, 2006.

C26. F.Ruffino, M. Muselli, G.Valentini Biological specifications for a synthetic gene expression data generation model,
In: I.Bloch, A. Petrosino, A.Tettamanzi (Eds.) WILF 2005, *Lecture Notes in Artificial Intelligence* vol. 3849, pp. 277-283, 2006.

C25. P. Campadelli, E. Casiraghi, G.Valentini, Lung nodules detection and classification,
ICIP 05, The IEEE International Conference on Image Processing, Genova, Italy, 2005.

C24. A. Bertoni, G. Valentini, Random projections for assessing gene expression cluster stability,
IJCNN '05. Proceedings IEEE International Joint Conference on Neural Networks, vol. 1 pp. 149-154, 2005.

C23. A. Bertoni, R. Folgieri, G. Valentini, Feature selection combined with random subspace ensemble for gene expression based diagnosis of malignancies,
In: (B.Apolloni, M.Marinaro and R. Tagliaferri, eds) *Biological and Artificial Intelligence Environments*, pp. 29-36, Springer, 2005.

C22. A. Bertoni, R. Folgieri, G. Valentini, Random subspace ensembles for the bio-molecular diagnosis of tumors,
Models and Metaphors from Biology to Bioinformatics Tools, *NETTAB* 2004.

C21. G. Valentini, Random aggregated and bagged ensembles of SVMs: an empirical bias-variance analysis,
In: (F. Roli, J. Kittler , T. Windeatt Eds.) Fifth International Workshop on Multiple Classifier Systems, *Lecture Notes in Computer Science*, vol. 3077, pp. 263-272, 2004, Powerpoint slides

C20. G. Valentini, T.G. Dietterich, Low Bias Bagged Support Vector Machines,
The Twentieth International Conference on Machine Learning, ICML 2003, Washington D.C. USA, pp. 752-759, AAAI Press, 2003.

C19. G. Valentini, An application of Low Bias Bagged SVMs to the classification of heterogeneous malignant tissues,
Pre-WIRN workshop on Bioinformatics and Biostatistic, *Lecture Notes in Computer Science*, vol. 2859, pp.316-321, 2003.

C18. G. Valentini, M. Muselli and F. Ruffino, Bagged Ensembles of SVMs for Gene Expression Data Analysis,
IJCNN2003, Proc. of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, Portland, USA, pp. 1844-1849, IEEE, 2003.

C17. G. Valentini, F. Masulli, Ensembles of learning machines.
In R. Tagliaferri and M. Marinaro, editors, Neural Nets WIRN Vietri-2002, *Lecture Notes in Computer Sciences*, vol. 2486, pp. 3-19, 2002.

C16. G. Valentini, T.G. Dietterich, Bias-Variance Analysis and Ensembles of SVM.
In J. Kittler and F. Roli (Eds) Third International Workshop on Multiple Classifier Systems, *Lecture Notes in Computer Science* vol. 2364, pp. 222-231, 2002.

C15. F. Masulli, M. Pardo, G. Sberveglieri, G. Valentini, Boosting and Classification of Electronic Nose Data,
Third International Workshop on Multiple Classifier Systems, *Lecture Notes in Computer Science* vol. 2364, pp. 262-271, 2002.

C14. G. Valentini, Supervised gene expression data analysis using Support Vector Machines and Multi-Layer Perceptrons,
In: Knowledge-Based Intelligent Information Engineering Systems and Allied technologies - Sixth International Conference on Knowledge-Based Intelligent Information & Engineering Systems *KES'2002* , special session Machine Learning in Bioinformatics, pp. 482-487, 2002.

C13. F. Ruffino , M. Muselli and G. Valentini, Feature Selection and Bagging Improve Malignancy Prediction based on Gene Expression Data.
Understanding the Genome: Scientific Progress and Microarray Technology, Genova, Italy, 2002.

C12. G. Valentini, Identifying different types of human lymphomas by SVM and ensembles of learning machines using DNA microarray data,
ISMB 2001 9th International Conference on Intelligent Systems and Molecular Biology (Poster section), Copenaghen, Denmark, 2001.

C11. G. Valentini, Classification of human malignancies by machine learning methods using DNA microarray gene expression data,
Proceedings of the Fourth International Conference "*Neural Networks and Expert Systems in Medicine and HealthCare*", Milos island, Greece, pp. 399-408, 2001.

C10. M. Pardo, G. Sberveglieri, G. Valentini, D. Della Casa, F.Masulli, Boosting applied to electronic nose data,
LFTNC-SC 2001 - *2001 NATO ARW on Limits and Future Trends of Neural Computing*, 2001.

C9. F. Masulli, G. Valentini, M. Pardo, G. Sberveglieri Classification of sensor array data by Output Coding decomposition methods.
Proc of the International Workshop *MATCHEMS 2001*, pp. 169-172, Brescia, Italy, 2001

C8. F. Masulli, G. Valentini, Quantitative evaluation of dependence among outputs in ECOC classifiers using mutual information based measures,
Proceedings of the International Joint Conference on Neural Networks *IJCNN'01*, K. Marko and P. Webos (eds.), vol.2, IEEE, Piscataway, NJ, USA, pp. 784-789, 2001.

C7. F. Masulli and G. Valentini, Dependence among Codeword Bit Errors in ECOC Learning Machines: an Experimental Analysis,
In: J.Kittler and F.Roli (eds.) Proceedings of the Second International Workshop Multiple Classifier Systems MCS 2001, Cambridge, UK, *Lecture Notes in Computer Science* vol. 2096, pp. 158-167, 2001

C6. M. Pardo, G. Sberveglieri, D. Della Casa, F.Masulli, G. Valentini, Multiple classifiers for electronic nose data,
8th *International Symposium on Olfaction and Electronic Noses*, Washington, 2001

C5. F. Masulli, G. Valentini, Comparing Decomposition Methods for Classification,
KES'2000, Fourth International Conference on Knowledge-Based Intelligent Engineering Systems & Allied Technologies, Brighton, UK, IEEE, Piscataway, NJ, USA, pp. 788-791, 2000.

C4. F. Masulli, G. Valentini, Parallel Non Linear Dichotomizers,
IJCNN2000, The IEEE-INNS-ENNS International Joint Conference on Neural Networks, Como, Italy, vol.2, pp. 29-33, 2000.

C3. M. Pardo, G. Sberveglieri, G. Valentini, F. Masulli, Decompositive classification models for electronic noses.
7th *International Symposium on Chemometrics in Analytical Chemistry* (CAC), Antwerp, 2000.

C2. F. Masulli, G. Valentini, Effectiveness of error correcting output codes in multiclass learning problems,
In: J.Kittler and F.Roli (eds.) Proceedings of the First International Workshop Multiple Classifier Systems MCS 2000, Cagliari, Italy, *Lecture Notes in Computer Science* vol.1857, pp.107-116, 2000.

C1. G. Valentini, F. Masulli, NEURObjects, a set of library classes for neural networks development, Proceedings of the third International ICSC Symposia on Intelligent Industrial Automation (IIA'99) and Soft Computing (*SOCO'99*), ICSC Academic Press, Millet, Canada, 1999, pp. 184-190.

Proceedings of National Conferences

N21. J. Gliozzo, M. Notaro, A. Petrini, P. Perlasca, M. Mesiti, E. Casiraghi, M.Frasca, G. Grossi, M. Re, A. Paccanaro, G. Valentini Modeling biomolecular profiles in a graph-structured sample space for clinical outcome prediction with melanoma and ovarian cancer patients ,
BITS 2017, Bioinformatics Italian Society Meeting, Cagliari, Italy, 2017.

N20. A. Petrini, M. Notaro, J. Gliozzo, G. Valentini, G. Grossi, M. Frasca Speeding up node label learning in unbalanced biomolecular networks through a parallel and sparse GPU- based Hopfield model
BITS 2017, Bioinformatics Italian Society Meeting, Cagliari, Italy, 2017.

N19. P. Perlasca, M. Mesiti, M. Notaro, A. Petrini, J. Gliozzo, G. Valentini, M. Frasca A Web Graphical Tool for the Integration of Unbalanced Biomolecular Networks ,
BITS 2017, Bioinformatics Italian Society Meeting, Cagliari, Italy, 2017.

N18. M. Re, M. Mesiti, M. Frasca, J. Lin, G. Valentini Analysis of bio-molecular networks through semi-supervised graph-based learning methods ,
*Third Italian Workshop on Machine Learning and Data Mining - XIII AI*IA Symposium on Artificial Intelligence* (invited talk), Pisa December 2014.

N17. M. Dugo, M. Callari, P. Miodini, V. Cappelletti, M.L. Carcangiu, R. Orlandi, G. Valentini, MG Daidone, Performance of single sample predictors in defining breast cancer molecular subtypes ,
53rd Annual Meeting of the Italian Cancer Society , Torino, October 2011.

N16. A. Bertoni, M. Frasca, G.Valentini, An efficient supervised method to integrate multiple biological networks ,
BITS 2011, Bioinformatics Italian Society Meeting, Pisa, Italy, 2011.

N15. A. Rozza , G. Lombardi, M. Re, E. Casiraghi, G. Valentini, P. Campadelli, A Novel Ensemble Approach for the Subcellular Localization of Proteins ,
BITS 2011, Bioinformatics Italian Society Meeting, Pisa, Italy, 2011.

N14. D. Malchiodi, M. Re and G. Valentini, Uso di Mathematica per la classificazione di dati di qualitÃ variabile ,
Mathematica Italia User Group Meeting - Atti del Convegno 2010, Adalta (ISBN 978-88-96810-00-2), 2010.

N13. M. Re, G.Valentini, Data fusion based gene function prediction using ensemble methods,
BITS 2009, Bioinformatics Italian Society Meeting, Genova, Italy, 2009.

N12. N. Cesa-Bianchi, G. Valentini, Genome-Âwide hierarchical classification of gene function,
BITS 2009, Bioinformatics Italian Society Meeting, Genova, Italy, 2009.

N11. R. Avogadri, A. Bertoni, G. Valentini, An integrated algorithmic procedure for the assessment and discovery of clusters in DNA microarray data,
BITS 2009, Bioinformatics Italian Society Meeting, Genova, Italy, 2009.

N10. G.Valentini, Statistical methods for the assessment of clusters discovered in bio-molecular data,
Proc. of the 6th SIB National Congress, Statistics in Life and Environment Sciences, Pisa, Italy, 2007.

N9. A.Bertoni, G.Valentini, A statistical test based on the Bernstein inequality to discover multi-level structures in bio-molecular data
BITS 2007, Bioinformatics Italian Society Meeting, Napoli, Italy, 2007.

N8. G.Pavesi, G.Valentini, Classification of co-expressed genes from DNA regulatory regions
BITS 2007, Bioinformatics Italian Society Meeting, Napoli, Italy, 2007.

N7. G. Pavesi , G. Valentini, G. Mauri, G. Pesole, Motif Based Classification of Coregulated Genes,
BITS 2006, Bioinformatics Italian Society Meeting, Bologna Italy, 2006.

N6. A. Bertoni, R. Folgieri, F. Ruffino, G. Valentini, Assessment of clusters reliability for high dimensional genomic data
BITS 2005, Bioinformatics Italian Society Meeting, Milano Italy, 2005

N5. F. Ruffino, G. Valentini, M.Muselli, Evaluation of gene selection methods through artificial and real-world data concerning DNA microarray experiments,
BITS 2005, Bioinformatics Italian Society Meeting, Milano Italy, 2005

N4. M. Muselli, F. Ruffino, and G. Valentini, An Artificial Model for Validating Gene Selection Methods,
BITS 2004, Bioinformatics Italian Society Meeeting, Padova, Italy, 2004

N3. F. Ruffino, G. Valentini, and M. Muselli, Metodi di Bagging e di selezione delle variabili per l' analisi dei dati di DNA microarray, SIS 2003.

N2. G. Valentini, Metodi di apprendimento automatico supervisionato per il riconoscimento di linfomi tramite DNA microarray, Atti III Convegno Federazione Italiana Scienze della Vita - *FISV 2001*", Riva del Garda (TN), 2001.

N1. M. Pardo, G. Benussi, G. Sberveglieri, G. Valentini, F. Masulli and M. Riani, Application of parallel non-linear dichotomizers to electronic noses, *INFMeeting 2000*, Genova, 2000.