

Giorgio Valentini - Curriculum vitae

Contents

1	Education and Academic position	2
2	Research	2
3	Participation and coordination of research projects and laboratories	4
4	Conference and workshop organization	5
5	Editorial activity	9
6	Teaching	10
7	Membership in national and international research associations and society.	11

1 Education and Academic position

I received the *degree in Biology* at the University of Genova in 1981. I have been teacher of biology and chemistry in Italian high school from 1982 to 1999. In 1999 I obtained the *degree in Computer Science* at the University of Genova and obtained a *fellowship at INFN* (Italian National Institute of Physical Matter). During 2001 I were visiting fellow at the Computer Science Department of the Oregon State University, USA. In 2003 I achieved the *PhD in Computer Science* at the University of Genova.

From 2003 to the end of 2004 I were a *post-doc researcher* and from January 2005 to October 2010 I were an *Assistant professor* at the Computer Science Department, University of Milano.

From November 2010 I am an *Associate professor* at the same Computer Science Department:

2 Research

My main research interests are in the areas of Bioinformatics and Machine Learning, and in particular in the design and application of machine learning methods to relevant problems in *Molecular Biology* and *Medicine*.

The algorithms I developed, even if mostly conceived in the framework of computational biology problems, are general enough to be applied in other contexts where inductive and transductive data analysis can play a key role.

The research activities are documented by about 40 papers published in international journals and about 70 publications in international conference proceedings and books (see Bibliography at the end of the document). All my papers are available from homes.di.unimi.it/valentini/pub.html.

A schematic taxonomy of my research activity in Bioinformatics and Machine Learning with the references to my main papers are summarized below. More information about my research activity in Bioinformatics is available from AnacletoLab: <http://anacletolab.di.unimi.it/>.

A. Bioinformatics

1. Automated Protein Function Prediction
 - Hierarchical ensembles of learning machines for gene/protein function prediction [4, 11, 14, 18]
 - Network-based protein function prediction [8, 7]
2. Drug ranking and repositioning and drug-target prediction [5]
3. Disease gene prioritization [2, 10]
4. Integration of multiple sources of bio-molecular data
 - Network combination algorithms for the massive integration of omics data [2, 51, 5]
 - Supervised ensemble methods for biomolecular data integration [17, 19]
5. Unsupervised search for patterns in biomolecular data
 - Stability-based methods for the assessment of the reliability of clusters discovered in complex bio-molecular data [28, 25, 31]
 - Ensemble clustering methods for the analysis of patterns in bio-molecular data [22]
6. Big data analysis in computational biology [1]

B. Machine Learning

1. Analysis and design of *Ensembles of Learning Machines* [58, 101]:
 - Hierarchical ensembles for multi-class, multi-label and multi-path classification problems [11, 14, 74]
 - Ensemble methods for data integration [17, 19]
 - Methods to analyze cluster reliability based on clustering ensembles [76, 28, 31]

- Ensemble methods based on the bias-variance decomposition of the error [33, 36, 98]
 - Supervised ensemble methods based on random projections [35]
 - Error Correcting Output Coding ensemble methods for multi-class classification [38, 39, 40]
2. Analysis and design of *graph-based* semi-supervised transductive methods
- Algorithms based on kernelized score functions to rank node labels in complex networks [5, 8, 10]
 - Cost-sensitive Hopfield networks for node label ranking and classification and network integration [7, 60]
 - Scalable vertex centric algorithms for secondary memory-based network learning [1]
3. Development of machine learning and bioinformatics software libraries [3, 26, 29, 32]

3 Participation and coordination of research projects and laboratories

Past projects:

- PASCAL2 network of excellence, within the 7th European Framework Programme funded by the European Union (2007-2013)
- European Network of Excellence *Pattern Analysis, Statistical Modelling and Computational Learning (PASCAL)* in the context of 6th European Framework Programme (2004-2006)
- NSF funded project *From Bits to Information: Statistical Learning Technologies for Digital Information Management Search* (2002)

- *Computational methods for bio-medical pattern analysis* funded by the University of Milan (2011-2013)
- MIUR COFIN-PRIN project *Machine Learning Techniques for Bioinformatics: Analysis and Modeling of Functional and Structural Data of Gene Expression* (2001-2002)
- MIUR COFIN-PRIN project *Formal Languages and automata: methods, models and applications* (2003-2005)
- Project *Diagnosis and bio-molecular therapy of acute myeloid leukaemia* with the Genetic Department of the University of Milan and the Niguarda Hospital of Milano

Current projects and coordination of research groups:

- MIUR PRIN project *Automata and formal languages: mathematical and applicative aspects*
- CNR-IEEIT project *Machine learning techniques for modelling and growing up*
- Coordinator of the University of Milan unit for the international Critical Assessment of Function Annotation (CAFA2)
- Coordinator of the Milan unit of *InfoLife*, the italian network for Bioinformatics of the National Interuniversity Consortium for Informatics (CINI)
- Supervisor of the post-doc project *Neural and graph-based algorithms for the analysis of biological networks* funded by the University of Milan
- Coordinator of the bioinformatics laboratory AnacletoLab at the Dept. of Computer Science of the University of Milan

4 Conference and workshop organization

I organized international conferences and workshops on bioinformatics and ensemble methods, and participated to several scientific committees

of international conferences. In particular, in 2007 I were *co-chair of CIBB 2007, Fourth International Conference on Computational Intelligence Methods for Bioinformatics and Biostatistics*, Portofino, Italy, and chair of *SUEMA 2007, International Workshop on Supervised and Unsupervised Ensemble Methods and Their Applications* (in conjunction with IbPRIA2007), Girona, Spain.

I also organized the Second edition of the SUEMA workshop within the ECAI 2008 (European Conference on Artificial Intelligence) in Patras (Greece), July 2008, with the support of the Pascal2 European network of excellence.

In 2009 I contributed to the organization of the *NIPS workshop Learning from Multiple Sources with Applications to Robotics* (Whistler, Canada).

In 2010, in the context of the European Conference in Machine Learning (ECML 2010, Barcelona), I organized and were chair of SUEMA 2010, Third International Workshop on Supervised and Unsupervised Ensemble Methods and Their Applications.

I am or I have been member of the scientific program committee of several bioinformatics and machine learning international workshops and conferences:

- ECML-PKDD 2014 - The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Nancy, France, 15-19 September, 2014
- AFP 2014 - Automatic Function Prediction - Critical Assessment of Functional Annotation experiment (colocated with ISMB 2014), Boston, MA, USA, 11-12 July 2014
- CIBB 2014 - Eleventh International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics. Cambridge, UK, 26-28 June, 2014.
- ICANN 2014 24th International Conference on Artificial Neural Networks Hamburg, Germany, 15-19 September 2014
- ELWA'14 - International Workshop on Ensemble Learning for Web Analytics, co-located with the 2014 IEEE/WIC/ACM International

Conference on Web Intelligence. Warsaw, Poland, 11-14 August 2014.

- SDM14 2014 SIAM International Conference on Data Mining Philadelphia, Pennsylvania, USA, April 24-26, 2014
- S+SSPR 2014 IAPR Joint International Workshops on Statistical + Structural and Syntactic Pattern Recognition Joensuu, Finland, 20-22 August 2014
- AIAB 2013 3rd Workshop on Artificial Intelligence Applications in Biomedicine Paphos, Cyprus, 30th September 2013
- ICANN 2013 23rd International Conference on Artificial Neural Networks Sofia, Bulgaria, 10-13 September 2013
- Multiclust 2013 4th Workshop on Multiple Clusterings, Multi-view Data, and Multi-source Knowledge-driven Clustering in conjunction with the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, Illinois, USA, August 11-14, 2013
- CIBB 2013 Tenth International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics Nice, France, 20-22 June 2013
- MCS 2013 11th International Conference on Multiple Classifier Systems Nanjing University (Xianlin Campus), China, May 15-17 2013
- ICPRAM 2013 2nd International Conference on Pattern Recognition Applications and Methods Barcelona, Spain, 2013
- CIBB 2012 9th International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics Houston Texas (USA), 2012
- SDM 2012 2012 SIAM International Conference on Data Mining Anaheim, California, USA, 2012

- ICPR 2012 21st International Conference on Pattern Recognition Tsukuba Science City, Japan, 2012
- ICPRAM 2012 International Conference on Pattern Recognition Applications and Methods Vilamoura, Algarve, Portugal, 2012
- BIODDD'11 2nd International Workshop on Biological Knowledge Discovery and Data Mining Toulouse, France, 2011
- ECML PKDD 2011 The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases Athens, 2011
- CIBB 2011 Eighth International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics Gargnano del Garda, Italy, 2011
- MCS 2011 10th International Workshop on Multiple Classifier Systems Napoli, 2011
- SEPLN 2010 Workshop on Language Technology applied to biomedical and health documents. Valencia, Spain, 2010
- BIODDD '10 2nd International Workshop on Biological Knowledge Discovery and Data mining Bilbao, Spain, 2010
- Ensemble Learning: Methods and Applications, IEA/AIE 2010 Special Session, Cordoba, Spain, 2010.
- MCS 2010 9th International Workshop on Multiple Classifier Systems Cairo, Egypt, 2010
- S+SSPR 2010 Joint IAPR International Workshops on Structural and Syntactic Pattern Recognition (SSPR 2010) and Statistical Techniques in Pattern Recognition (SPR 2010), Cesme, Izmir, Turkey
- Learning from Multiple Sources with Applications to Robotics, NIPS 2009 workshop, Whistler, BC, Canada, 2009

- CIBB 2009 Sixth International Conference on Bioinformatics and Biostatistics
- NETTAB 2008 Bioinformatics Methods for Biomedical Complex System Applications
- CIBB 2008 Fifth International Conference on Bioinformatics and Biostatistics
- Bioinformatics, Genomics & Proteomics, an Artificial Intelligence Approach (ECAI 2008 workshop)
- CIBB 2007 Fourth International Conference on Bioinformatics and Biostatistics
- ISSNIP 2007 - Intelligent Sensors Networks and Information Processing
- CI-BIO 2007, Post-IJCNN 2007 Workshop on Computational Intelligence Approaches for the Analysis of Bioinformatics Data

5 Editorial activity

I am member of the Editorial Board of the *Open Bioinformatics Journal*, *Open Medical Informatics Journal*, *Open Applied Informatics Journal*. From 2012 I am Associate editor of the *Universal Journal of Information Technology*, of the *Journal of Integrated OMICS*. From 2013 I am member of the Editorial Board of the *International Journal of Neural Networks and Advanced Applications*, of *Advances in Biology* and of the *Computational Biology and Bioinformatics* journal. From 2014 I am member of the Editorial Board of the *Journal of Proteomics and Computational Biology*.

I serve as reviewer for the main international bioinformatics journals (Bioinformatics, BMC Bioinformatics, Briefings in Bioinformatics, IEEE ACM Trans. on Comp. Biol. and Bioinformatics, PLoS Computational Biology, and others), and for several machine learning journals (Journal of Machine Learning Research, Machine Learning, IEEE Transactions on

Neural Networks, Artificial Intelligence, Neurocomputing and others). In 2009 I were associated editor of a special issue of *Artificial Intelligence in Medicine* dedicated to computational intelligence methods in bioinformatics, and co-editor of three Springer books (series: *Lectures Notes in Computational Intelligence*) on ensemble methods.

6 Teaching

I currently hold the following courses:

- Bioinformatics (advanced degree in Computer Science, University of Milan)
- Bioinformatics Methods (advanced degree in Biomolecular Biotechnologies and Bioinformatics, University of Milan)
- Advanced Informatics (degree in Industrial and Environmental Biotechnologies, University of Milan)

In the years from 2003 to 2013 I held the following courses at the University of Milan:

- *Bioinformatics* (degree in Pharmacology, University of Milan),
- *Programming Languages for Bioinformatics* (advanced degree in Functional Genomics and Bioinformatics),
- *Computer Science applied to biotechnological processes* (degree Industrial and Environmental Biotechnologies),
- *Algorithms for Bioinformatics* (advanced degree in Functional Genomics and Bioinformatics),
- *Bioinformatics Laboratory* (degree in Biology) and I held also several basic Informatics courses for the Faculty of Medicine.

In the years from 2001 to 2003 I held the following courses at the University of Genova:

- *Functional Bioinformatics* (degree in Physics),
- *Laboratory of Neural Networks* (degree in Computer Science),
- *Laboratory of System programming* (degree in Computer Science).

I have been *advisor for several theses* (degree or advanced degree in Computer Science, in Biomolecular Biotechnologies and Bioinformatics and in Mathematics) and *advisor of 5 PhD theses* in Computer Science.

In 2007 I held several bioinformatics classes for the International PhD School in Bioinformatics at the University of Brno (Czech Republic) and in 2008 I held bioinformatics classes for the PhD School at the University of Palermo (Italy). I also held lectures at the International School on Neural Nets "E.R. Caianiello" in Vietri (Italy). In 2008 and 2009 I were coordinator of the course *R language for Bioinformatics* for the Italian Network for Oncology Bioinformatics. I held lectures and seminars in several academic and research centres in Italy and abroad.

7 Membership in national and international research associations and society.

I am member of ISCB (International Society of Computational Biology), BITS (Italian Bioinformatics Society), the INNS (International Neural Networks Society), and SIREN (Italian Society of Neural Networks), I am associated member of the CNR-IEEIT (Electronic Engineering Institute of Telecommunications of the National Research Council).

List of Publications

Journals

- [1] M. Mesiti, M. Re, and G. Valentini. Think globally and solve locally: secondary memory-based network learning for automated multi-species function prediction. *GigaScience*, 3:5, 2014.
- [2] G. Valentini, A. Paccanaro, H. Caniza, A. Romero, and M. Re. An extensive analysis of disease-gene associations using network integration and fast kernel-based gene prioritization methods. *Artificial Intelligence in Medicine*, 61(2):63–78, 2014.
- [3] H. Caniza, A. Romero, S. Heron, H. Yang, A. Devoto, M. Frasca, M. Mesiti, G. Valentini, and A. Paccanaro. GOssTo: a user-friendly stand-alone and web tool for calculating semantic similarities on the Gene Ontology. *Bioinformatics*, 2014. (in press).
- [4] G. Valentini. Hierarchical Ensemble Methods for Protein Function Prediction. *ISRN Bioinformatics*, 2014(Article ID 901419):34 pages, 2014.
- [5] 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):1359–1371, 2013.
- [6] I. Cattinelli, G. Valentini, E. Paulesu, and 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):1166–1173, 2013.
- [7] M. Frasca, A. Bertoni, M. Re, and G. Valentini. A neural network algorithm for semi-supervised node label learning from unbalanced data. *Neural Networks*, 43:84–98, 2013.
- [8] 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):1812–1818, 2012.
- [9] 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, Morra E., and R. Cairoli. Regeneration-associated Wnt signaling is activated in long-term reconstituting AC133bright acute myeloid leukemia cells. *Neoplasia*, 14(12):1236–1248, 2012.
- [10] M. Re and G. Valentini. Cancer module genes ranking using kernelized score functions. *BMC Bioinformatics*, 13(Suppl 14/S3), 2012.
- [11] N. Cesa-Bianchi, M. Re, and G. Valentini. Synergy of multi-label hierarchical ensembles, data fusion, and cost-sensitive methods for gene functional inference. *Machine Learning*, 88(1):209–241, 2012.
- [12] M. Re, M. Mesiti, and G. Valentini. Drug repositioning through pharmacological spaces integration based on networks projection. *EMBnet.journal*, 18(Supplement A):30–31, 2012.
- [13] M. Frasca, A. Bertoni, and G. Valentini. Regularized network-based algorithm for predicting gene functions with high-imbalanced data. *EMBnet.journal*, 18(Supplement A):41–42, 2012.
- [14] G. Valentini. True Path Rule hierarchical ensembles for genome-wide gene function prediction. *IEEE ACM Transactions on Computational Biology and Bioinformatics*, 8(3):832–847, 2011.
- [15] 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*, 8:1385–1392, 2011.
- [16] M. Re and G. Valentini. Noise tolerance of Multiple Classifier Systems in data integration-based gene function prediction. *Journal of Integrative Bioinformatics*, 7(3:139), 2010.
- [17] M. Re and 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, Machine Learning in Systems Biology*, 8:98–111, 2010.
- [18] N. Cesa-Bianchi and G. Valentini. Hierarchical cost-sensitive algorithms for genome-wide gene function prediction. *Journal of Machine Learning Research, W&C Proceedings, Machine Learning in Systems Biology*, 8:14–29, 2010.

- [19] M. Re and G. Valentini. Integration of heterogeneous data sources for gene function prediction using Decision Templates and ensembles of learning machines. *Neurocomputing*, 73(7-9):1533–37, 2010.
- [20] 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*, S12:S7, 2009.
- [21] 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):318–330, 2009.
- [22] R. Avogadri and G. Valentini. Fuzzy ensemble clustering based on random projections for dna microarray data analysis. *Artificial Intelligence in Medicine*, 45(2):173–183, 2009.
- [23] G. Valentini, R. Tagliaferri, and F. Masulli. Computational Intelligence and Machine Learning in Bioinformatics. *Artificial Intelligence in Medicine*, 45(2):91–96, 2009.
- [24] G. Pavesi and G. Valentini. Classification of co-expressed genes from DNA regulatory regions. *Information Fusion*, 10(3):233–241, 2009.
- [25] A. Bertoni and G. Valentini. Discovering multi-level structures in bio-molecular data through the Bernstein inequality. *BMC Bioinformatics*, 9(S2), 2008.
- [26] G. Valentini and N. Cesa-Bianchi. Hcgene: a software tool to support the hierarchical classification of genes. *Bioinformatics*, 24(5):729–731, 2008.
- [27] F. Ruffino, M. Muselli, and G. Valentini. Gene expression modelling through positive boolean functions. *International Journal of Approximate Reasoning*, 47(1):97–108, 2008.
- [28] A. Bertoni and G. Valentini. Model order selection for bio-molecular data clustering. *BMC Bioinformatics*, 8(Suppl.3), 2007.
- [29] G. Valentini. Mosclust: a software library for discovering significant structures in bio-molecular data. *Bioinformatics*, 23(3):387–389, 2007.

- [30] G. Valentini and F. Ruffino. Characterization of lung tumor subtypes through gene expression cluster validity assessment. *RAIRO - Theoretical Informatics and Applications*, 40:163–176, 2006.
- [31] A. Bertoni and 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.
- [32] G. Valentini. Clusterv: a tool for assessing the reliability of clusters discovered in DNA microarray data. *Bioinformatics*, 22(3):369–370, 2006.
- [33] G. Valentini. An experimental bias-variance analysis of SVM ensembles based on resampling techniques. *IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics*, 35(6):1252–1271, 2005.
- [34] P. Campadelli, E. Casiraghi, and G. Valentini. Support vector machines for candidate nodules classification. *Neurocomputing*, 68:281–289, 2005.
- [35] A. Bertoni, R. Folgieri, and G. Valentini. Bio-molecular cancer prediction with random subspace ensembles of support vector machines. *Neurocomputing*, 63C:535–539, 2005.
- [36] G. Valentini and T.G. Dietterich. Bias–variance analysis of Support Vector Machines for the development of SVM-based ensemble methods. *Journal of Machine Learning Research*, 5:725–775, 2004.
- [37] G. Valentini, M. Muselli, and F. Ruffino. Cancer recognition with bagged ensembles of Support Vector Machines. *Neurocomputing*, 56:461–466, 2004.
- [38] F. Masulli and G. Valentini. An experimental analysis of the dependence among codeword bit errors in ecoc learning machines. *Neurocomputing*, 57:189–214, 2004.
- [39] F. Masulli and G. Valentini. Effectiveness of Error Correcting Output Coding decomposition schemes in ensemble and monolithic learning machines. *Pattern Analysis and Application*, 6:285–300, 2003.
- [40] G. Valentini. Gene expression data analysis of human lymphoma using support vector machines and output coding ensembles. *Artificial Intelligence in Medicine*, 26(3):283–306, 2002.

- [41] G. Valentini and F. Masulli. Neuroobjects: an object-oriented library for neural network development. *Neurocomputing*, 48(1–4):623–646, 2002.
- [42] G. Valentini. Gene expression-based prediction of malignancies. *AIIA Notizie*, XV(4):34–38, 2002.
- [43] M. Pardo, G. Sberveglieri, F. Masulli, and G. Valentini. Decompositive classification models for electronic noses. *Anal. Chimica Acta*, 446:223–232, 2001.

Books

- [44] O. Okun, G. Valentini, and M. Re. *Ensembles in Machine Learning Applications*, volume 373 of *Studies in Computational Intelligence*. Springer, Berlin, 2011.
- [45] O. Okun, M. Re, and G. Valentini. *Proceedings of the the Third Workshop on Supervised and Unsupervised Ensemble Methods and Their Applications (SUEMA), European Conference on Machine Learning*. Barcelona, Spain, 2010.
- [46] O. Okun and G. (eds.) Valentini. *Applications of Supervised and Unsupervised Ensemble Methods*, volume 245 of *Studies in Computational Intelligence*. Springer-Verlag, Berlin, 2009.
- [47] O. Okun and G. (eds.) Valentini. *Supervised and Unsupervised Ensemble Methods and their Applications*, volume 126 of *Studies in Computational Intelligence*. Springer-Verlag, Berlin, 2008.
- [48] O. Okun and G. (eds.) Valentini. *Proceedings of the the Second Workshop on Supervised and Unsupervised Ensemble Methods and Their Applications (SUEMA), European Conference on Artificial Intelligence*, volume 126. University of Patras, Greece, Patras, 2008. ISBN: 978-960-89282-2-0.

Proceedings of International Conferences and book chapters

- [49] M. Re, M. Mesiti, and G. Valentini. An automated pipeline for multi-species protein function prediction from the UniProt Knowledgebase. In *Automated Function Prediction SIG 2014 - ISMB 2014*, Boston, MA, USA, 2014.

- [50] M. Re, M. Mesiti, and G. Valentini. On the automated function prediction of big multi-species networks. In *Network Biology SIG 2014 ISMB 2014*, Boston, MA, USA, 2014.
- [51] M. Frasca, A. Bertoni, and G. Valentini. An unbalance-aware network integration method for gene function prediction. In *MLSB 2013 - Machine Learning for Systems Biology - Berlin*, Berlin, Germany, 2013.
- [52] G. Valentini, A. Paccanaro, H.C. Vierci, A.E. Romero, and M. Re. Network integration boosts disease gene prioritization. In *Network Biology SIG 2013 ISMB 2013*, Berlin, Germany, 2013.
- [53] M. Mesiti, M. Re, and G. Valentini. Scalable Network-based Learning Methods for Automated Function Prediction based on the Neo4j Graph-database. In *Automated Function Prediction SIG 2013 - ISMB 2013*, Berlin, Germany, 2013.
- [54] 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. In *Bio-Ontologies SIG 2013 - ISMB 2013*, Berlin, Germany, 2013.
- [55] M. Re, M. Mesiti, and G. Valentini. Comparison of early and late omics data integration for cancer modules gene ranking. In *NETTAB 2012 Workshop on Integrated Bio-Search*, Como, Italy, 2012.
- [56] M. Re and G. Valentini. Random walking on functional interaction networks to rank genes involved in cancer. In *2nd Artificial Intelligence Applications in Biomedicine Workshop, in: AIAI 2012 - Artificial Intelligence Applications and Innovations*, IFIP AICT Series, pages 66–75. Springer, 2012.
- [57] M. Re and G. Valentini. Large scale ranking and repositioning of drugs with respect to drugbank therapeutic categories. In *International Symposium on Bioinformatics Research and Applications (ISBRA 2012)*, volume 7292 of *Lecture Notes in Computer Science*, pages 225–236. Springer, 2012.
- [58] M. Re and G. Valentini. Ensemble methods: a review. In *Advances in Machine Learning and Data Mining for Astronomy*, Data Mining and Knowledge Discovery, pages 563–594. Chapman & Hall, 2012.

- [59] M. Re and G. Valentini. Genes prioritization with respect to Cancer Gene Modules using functional interaction network data. In *NETTAB 2011 Workshop on Clinical Bioinformatics*, Pavia, Italy, 2011.
- [60] A. Bertoni, M. Frasca, and G. Valentini. Cosnet: a cost sensitive neural network for semi-supervised learning in graphs. In *European Conference on Machine Learning, ECML PKDD 2011*, volume 6911 of *Lecture Notes on Artificial Intelligence*, pages 219–234. Springer, 2011.
- [61] 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*, volume 373 of *Studies in Computational Intelligence*, pages 151–167. Springer, 2011.
- [62] M. Frasca, A. Bertoni, and G. Valentini. A cost-sensitive neural algorithm to predict gene functions using large biological networks. In *Network Biology SIG: On the Analysis and Visualization of Networks in Biology, ISMB 2011*, Wien, 2011.
- [63] A. Bertoni, M. Re, F. Sacca, and G. Valentini. Identification of promoter regions in genomic sequences by 1-dimensional constraint clustering. In *Neural Nets WIRN11 - Proceedings*, volume 234 of *Frontiers in Artificial Intelligence and Applications*, pages 162–169, 2011.
- [64] A. Rozza, G. Lombardi, M. Re, E. Casiraghi, and G. Valentini. DDAG K-TIPCAC: an ensemble method for protein subcellular localization. In *Proc. of the Third Edition of SUEMA - ECML*, Barcelona, Spain, 2010.
- [65] N. Cesa-Bianchi, M. Re, and G. Valentini. Functional inference in FunCat through the combination of hierarchical ensembles with data fusion methods. In *ICML-MLD 2nd International Workshop on learning from Multi-Label Data*, pages 13–20, Haifa, Israel, 2010.
- [66] A. Bertoni, M. Frasca, G. Grossi, and G. Valentini. Learning functional linkage networks with a cost-sensitive approach. In *Neural Networks - WIRN 2010*, pages 52–61. IOS Press, 2010.
- [67] M. Re and G. Valentini. An experimental comparison of Hierarchical Bayes and True Path Rule ensembles for protein function prediction. In N. El-

- Gayar, Kittler, J., and F. Roli, editors, *Multiple Classifier Systems. Ninth International Workshop, MCS 2010, Cairo, Egypt*, volume 5997 of *Lecture Notes in Computer Science*, pages 294–303. Springer-Verlag, 2010.
- [68] M. Re and G. Valentini. Simple ensemble methods are competitive with state-of-the-art data integration methods for gene function prediction. In *Machine Learning in Systems Biology, Proceedings of the Third international workshop*, pages 95–104, Ljubljana, Slovenia, 2009.
- [69] N. Cesa-Bianchi and G. Valentini. Hierarchical cost-sensitive algorithms for genome-wide gene function prediction. In *Machine Learning in Systems Biology, Proceedings of the Third international workshop*, pages 25–34, Ljubljana, Slovenia, 2009.
- [70] G. Valentini and M. Re. Weighted True Path Rule: a multilabel hierarchical algorithm for gene function prediction. In *MLD-ECML 2009, 1st International Workshop on learning from Multi-Label Data*, pages 133–146, Bled, Slovenia, 2009.
- [71] M. Re and G. Valentini. Predicting gene expression from heterogeneous data. In *CIBB 2009, The Sixth International Conference on Bioinformatics and Biostatistics*, Genova, Italy, 2009.
- [72] M. Re and G. Valentini. Comparing early and late data fusion methods for gene function prediction. In *Neural Nets WIRN11 - Proceedings*, volume 204 of *Frontiers in Artificial Intelligence and Applications*, pages 197–207, 2009.
- [73] M. Re and G. Valentini. Ensemble based data fusion for gene function prediction. In J. Kittler, J. Benediktsson, and F. Roli, editors, *Multiple Classifier Systems. Eighth International Workshop, MCS 2009, Reykjavik, Iceland*, volume 5519 of *Lecture Notes in Computer Science*, pages 448–457. Springer, 2009.
- [74] G. Valentini. True path rule hierarchical ensembles. In J. Kittler, J. Benediktsson, and F. Roli, editors, *Multiple Classifier Systems. Eighth International Workshop, MCS 2009, Reykjavik, Iceland*, volume 5519 of *Lecture Notes in Computer Science*, pages 232–241. Springer, 2009.
- [75] O. Okun, G. Valentini, and H. Priisalu. Exploring the link between bolstered classification error and dataset complexity for gene expression based

- cancer classification. In *New Signal Processing Research*, pages 249–278. Nova Publishers, 2009.
- [76] A. Bertoni and 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*, volume 5488 of *Lecture Notes in Computer Science*, pages 25–43. Springer, 2009.
- [77] M. Re and G. Valentini. Prediction of gene function using ensembles of SVMs and heterogeneous data sources. In *Applications of supervised and unsupervised ensemble methods*, volume 245 of *Computational Intelligence Series*, pages 79–91. Springer, 2009.
- [78] M. Mesiti, E. Jimenez-Ruiz, I. Sanz, R. Berlanga-Llavori, G. Valentini, P. Peralasca, and 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*, pages 263–286. Information Science, 2009.
- [79] 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 *Knowledge-Based Intelligent Information and Engineering Systems, 12th International Conference, KES 2008*, volume 5179 of *Lecture Notes in Computer Science*, pages 764–770. Springer, 2008.
- [80] M. Mesiti, E. Jimenez-Ruiz, I. Sanz, R. Berlanga-Llavori, G. Valentini, P. Peralasca, and D. Manset. XML-based approaches for the integration of heterogeneous bio-molecular data. In *NETTAB 2008 workshop on: Bioinformatics Methods for Biomedical Complex System Applications*, Varenna, Italy, 2008.
- [81] O. Okun and G. Valentini. Dataset Complexity Can Help to Generate Accurate Ensembles of K-Nearest Neighbors. In *IEEE International Joint Conference on Neural Networks - IJCNN 2008*, pages 450–457, 2008.
- [82] R. Avogadri and G. Valentini. Ensemble clustering with a fuzzy approach. In *Supervised and Unsupervised Ensemble Methods and their Applications*, volume 126 of *Studies in Computational Intelligence*, pages 50–69. Springer, 2008.
- [83] 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. In *Workshop on Computational Intelligence approaches for the analysis of Bioinformatics data - IJCNN 2007*, Orlando, USA, 2007.
- [84] A. Bertoni and G. Valentini. Discovering Significant Structures in Clustered Bio-molecular Data Through the Bernstein Inequality. In *Knowledge-Based Intelligent Information and Engineering Systems, 11th International Conference, KES 2007*, volume 4964 of *Lecture Notes in Computer Science*, pages 886–891, Berlin, 2007.
- [85] R. Avogadri and G. Valentini. Fuzzy ensemble clustering for DNA microarray data analysis. In *CIBB 2007, The Fourth International Conference on Bioinformatics and Biostatistics*, volume 4578 of *Lecture Notes in Computer Science*, pages 537–543. Springer, 2007.
- [86] A. Bertoni and G. Valentini. Randomized embedding cluster ensembles for gene expression data analysis. In *SETIT 2007 - IEEE International Conf. on Sciences of Electronic, Technologies of Information and Telecommunications*, Hammamet, Tunisia, 2007.
- [87] F. Ruffino, M. Muselli, and G. Valentini. Modeling gene expression data via positive boolean functions. In *NETTAB 2006 workshop on Distributed Applications, Web Services, Tools and GRID Infrastructures for Bioinformatics*, S.Margherita di Pula, Italy, 2006.
- [88] A. Bertoni and G. Valentini. Model order selection for clustered bio-molecular data. In J. Rousu, S. Kaski, and E. Ukkonen, editors, *Probabilistic Modeling and Machine Learning in Structural and Systems Biology*, pages 85–90, Tuusula, Finland, 2006. Helsinki University Printing House.
- [89] A. Bertoni and G. Valentini. Ensembles based on random projections to improve the accuracy of clustering algorithms. In *Neural Nets, WIRN 2005*, volume 3931 of *Lecture Notes in Computer Science*, pages 31–37. Springer, 2006.
- [90] B. Apolloni, G. Valentini, and A. Brega. BICA and Random Subspace ensembles for DNA microarray-based diagnosis. In *Proc. of 7th International FLINS Conference on Applied Artificial Intelligence*, pages 623–631. World Scientific, 2006.

- [91] F. Ruffino, M. Muselli, and G. Valentini. Biological specifications for a synthetic gene expression data generation model. In I. Bloch, A. Petrosino, and A. Tettamanzi, editors, *WILF 2005*, volume 3849 of *Lecture Notes in Artificial Intelligence*, pages 277–283, Berlin, 2006.
- [92] P. Campadelli, E. Casiraghi, and G. Valentini. Lung nodules detection and classification. In *ICIP 05, The IEEE International Conference on Image Processing*, Genova, Italy, 2005.
- [93] A. Bertoni and G. Valentini. Random projections for assessing gene expression cluster stability. In *IJCNN 2005, The IEEE-INNS International Joint Conference on Neural Networks*, volume 1, pages 149–154, Montreal, 2005.
- [94] F. Ruffino, M. Muselli, and G. Valentini. Biological specifications for a synthetic gene expression data generation model. In I. Bloch, A. Petrosino, and A. Tettamanzi, editors, *WILF 2005*, volume 3849 of *Lecture Notes in Artificial Intelligence*, pages 277–283, Berlin, 2005.
- [95] A. Bertoni, R. Folgieri, and G. Valentini. Feature selection combined with random subspace ensemble for gene expression based diagnosis of malignancies. In B. Apolloni, M. Marinaro, and R. Tagliaferri, editors, *Biological and Artificial Intelligence Environments*, pages 29–35. Springer, 2005.
- [96] A. Bertoni, R. Folgieri, and G. Valentini. Random subspace ensembles for the bio-molecular diagnosis of tumors. In *Models and Metaphors from Biology to Bioinformatics Tools, NETTAB 2004*, Camerino, Italy, 2004.
- [97] G. Valentini. Random aggregated and bagged ensembles of SVMs: an empirical bias-variance analysis. In *Multiple Classifier Systems. Fifth International Workshop, MCS2004, Cagliari, Italy*, volume 3077 of *Lecture Notes in Computer Science*, pages 263–272. Springer, 2004.
- [98] G. Valentini and T.G. Dietterich. Low Bias Bagged Support Vector Machines. In T. Fawcett and N. Mishra, editors, *Machine Learning, Proceedings of the Twentieth International Conference (ICML 2003)*, pages 752–759, Washington D.C., USA, 2003. AAAI Press.
- [99] G. Valentini. An application of low bias bagged svms to the classification of heterogeneous malignant tissues. In *Neural Nets*, volume 2859 of *Lecture Notes in Computer Science*, pages 316–321. Springer-Verlag, 2003.

- [100] G. Valentini, M. Muselli, and F. Ruffino. Bagged Ensembles of SVMs for Gene Expression Data Analysis. In *IJCNN2003, The IEEE-INNS-ENNS International Joint Conference on Neural Networks*, pages 1844–49, Portland, USA, 2003. IEEE.
- [101] G. Valentini and F. Masulli. Ensembles of learning machines. In *Neural Nets WIRN-02*, volume 2486 of *Lecture Notes in Computer Science*, pages 3–19. Springer, 2002.
- [102] G. Valentini and T.G. Dietterich. Bias–variance analysis and ensembles of SVM. In *Multiple Classifier Systems. Third International Workshop, MCS2002, Cagliari, Italy*, volume 2364 of *Lecture Notes in Computer Science*, pages 222–231. Springer-Verlag, 2002.
- [103] F. Masulli, M. Pardo, G Sberveglieri, and G. Valentini. Boosting and classification of electronic nose data. In *Multiple Classifier Systems. Third International Workshop, MCS2002, Cagliari, Italy*, volume 2364 of *Lecture Notes in Computer Science*, pages 262–271. Springer-Verlag, 2002.
- [104] G. Valentini. Supervised gene expression data analysis using Support Vector Machines and Multi-Layer perceptrons. In *Proc. of KES'2002, the Sixth International Conference on Knowledge-Based Intelligent Information & Engineering Systems, special session Machine Learning in Bioinformatics*, Amsterdam, the Netherlands, 2002. IOS Press.
- [105] M. Pardo, G. Sberveglieri, G. Valentini, and F. Masulli. Classification with Boosted Multilayer Perceptrons. In *Eighth International Conference on Chemometrics in Analytical Chemistry*, Seattle, Washington, USA, 2002.
- [106] F. Ruffino, M. Muselli, and G. Valentini. Feature selection and bagging improve malignancy prediction based on gene expression data. In *Understanding the Genome: Scientific Progress and Microarray Technology*, Genova, Italy, 2002.
- [107] G. Valentini. "identifying different types of human lymphomas by svm and ensembles of learning machines using dna microarray data". In *9th International Conference on Intelligent Systems and Molecular Biology*, Copenhagen, Denmark, 2001.

- [108] G. Valentini. Classification of human malignancies by machine learning methods using DNA microarray gene expression data. In G.M. Papadourakis, editor, *Fourth International Conference Neural Networks and Expert Systems in Medicine and HealthCare*, pages 399–408, Milos island, Greece, 2001. Technological Educational Institute of Crete.
- [109] M. Pardo, G. Sberveglieri, G. Valentini, D. Della Casa, and F. Masulli. Boosting applied to electronic nose data. In *LFTNC-SC 2001 - 2001 NATO ARW on Limits and Future Trends of Neural Computing*, Siena, Italy, 2001.
- [110] F. Masulli, G. Valentini, M. Pardo, and G. Sberveglieri. Classification of sensor array data by Output Coding decomposition methods. In *International Workshop MATCHEMS 2001*, pages 169–172, Brescia, Italy, 2001.
- [111] F. Masulli and G. Valentini. Quantitative Evaluation of Dependence among Outputs in ECOC Classifiers Using Mutual Information Based Measures. In K. Marko and P. Webos, editors, *Proceedings of the International Joint Conference on Neural Networks IJCNN'01*, volume 2, pages 784–789, Piscataway, NJ, USA, 2001. IEEE.
- [112] F. Masulli and G. Valentini. Dependence among Codeword Bits Errors in ECOC Learning Machines: an Experimental Analysis. In *Lecture Notes in Computer Science*, volume 2096, pages 158–167, Berlin, 2001. Springer-Verlag.
- [113] M. Pardo, G. Sberveglieri, D. Della Casa, F. Masulli, and G. Valentini. Multiple classifiers for electronic nose data. In *International Symposium on Olfaction and Electronic Noses, ISOEN 2001*, Washington D.C., USA, 2001.
- [114] F. Masulli and G. Valentini. Comparing decomposition methods for classification. In R.J. Howlett and L.C. Jain, editors, *KES'2000, Fourth International Conference on Knowledge-Based Intelligent Engineering Systems & Allied Technologies*, pages 788–791, Piscataway, NJ, 2000. IEEE.
- [115] F. Masulli and G. Valentini. Parallel Non linear Dichotomizers. In *IJCNN2000, The IEEE-INNS-ENNS International Joint Conference on Neural Networks*, volume 2, pages 29–33, Como, Italy, 2000.

- [116] M. Pardo, G. Sberveglieri, G. Valentini, and F. Masulli. Decompositive classification models for electronic noses. In *7th International Symposium on Chemometrics in Analytical Chemistry (CAC)*, Antwerp, 2000.
- [117] F. Masulli and G. Valentini. Effectiveness of error correcting output codes in multiclass learning problems. In *Lecture Notes in Computer Science*, volume 1857, pages 107–116. Springer-Verlag, Berlin, Heidelberg, 2000.
- [118] G. Valentini and F. Masulli. NEUROObjects, a set of library classes for neural networks development. In *Proceedings of the third International ICSC Symposia on Intelligent Industrial Automation (IIA'99) and Soft Computing (SOCO'99)*, pages 184–190, Millet, Canada, 1999. ICSC Academic Press.

Proceedings of National Conferences

- [119] M. Dugo, M. Callari, P. Miodini, V. Cappelletti, M.L. Carcangiu, R. Orlandi, G. Valentini, and MG Daidone. "performance of single sample predictors in defining breast cancer molecular subtypes". In *53rd Annual Meeting of the Italian Cancer Society*, Torino, Italy, 2011.
- [120] M. Bertoni, A. Frasca and G. Valentini. An efficient supervised method to integrate multiple biological networks. In *BITS 2011, Bioinformatics Italian Society Meeting*, Pisa Italy, 2011.
- [121] A. Rozza, G. Lombardi, M. Re, E. Casiraghi, G. Valentini, and P. Campadelli. "a novel ensemble approach for the subcellular localization of proteins". In *BITS 2011, Bioinformatics Italian Society Meeting*, Pisa Italy, 2011.
- [122] D. Malchiodi, M. Re, and G. Valentini. "uso di mathematica per la classificazione di dati di qualita' variabile". In *Mathematica Italia User Group Meeting - Atti del Convegno 2010*. Adalta, 2010.
- [123] M. Re and G. Valentini. "data fusion based gene function prediction using ensemble methods". In *BITS 2009, Bioinformatics Italian Society Meeting*, Genova Italy, 2009.
- [124] N. Cesa-Bianchi and G. Valentini. "genome-wide hierarchical classification of gene function". In *BITS 2009, Bioinformatics Italian Society Meeting*, Genova Italy, 2009.

- [125] R. Avogadri, A. Bertoni, and G. Valentini. "an integrated algorithmic procedure for the assessment and discovery of clusters in dna microarray data". In *BITS 2009, Bioinformatics Italian Society Meeting*, Genova Italy, 2009.
- [126] G. Valentini. "statistical methods for the assessment of clusters discovered in bio-molecular data". In *Proc. of the 6th SIB National Congress, Statistics in Life and Environment Sciences*, Pisa Italy, 2007.
- [127] A. Bertoni and G. Valentini. "a statistical test based on the bernstein inequality to discover multi-level structures in bio-molecular data". In *BITS 2007, Bioinformatics Italian Society Meeting*, Napoli Italy, 2007.
- [128] G. Pavesi and G. Valentini. "classification of co-expressed genes from dna regulatory regions". In *BITS 2007, Bioinformatics Italian Society Meeting*, Napoli Italy, 2007.
- [129] G. Pavesi, G. Valentini, G. Mauri, and G. Pesole. "motif based classification of coregulated genes". In *BITS 2006, Bioinformatics Italian Society Meeting*, Bologna Italy, 2006.
- [130] A. Bertoni and G. Valentini. Discovering significant structures in clustered data through bernstein inequality. In *CISI '06, Conferenza Italiana Sistemi Intelligenti*, Ancona, Italy, 2006.
- [131] A. Bertoni, R. Folgieri, F. Ruffino, and G. Valentini. "assessment of clusters reliability for high dimensional genomic data". In *BITS 2005, Bioinformatics Italian Society Meeting*, Milano Italy, 2005.
- [132] F. Ruffino, G. Valentini, and M. Muselli. "evaluation of gene selection methods through artificial and real-world data concerning dna microarray experiments". In *BITS 2005, Bioinformatics Italian Society Meeting*, Milano Italy, 2005.
- [133] M. Muselli, F. Ruffino, and G. Valentini. "an artificial model for validating gene selection methods". In *BITS 2004, Bioinformatics Italian Society Meeting*, Padova Italy, 2004.
- [134] F. Ruffino, G. Valentini, and M. Muselli. "metodi di bagging e di selezione delle variabili per l'analisi dei dati di dna microarray". In *SIS 2003*, Genova, Italy, 2003.

- [135] G. Valentini. "metodi di apprendimento automatico supervisionato per il riconoscimento di linfomi tramite dna microarray". In *Atti III Convegno Federazione Italiana Scienze della Vita - FISV 2001*, Riva del Garda, Italy, 2001.