

# Anne-Laure Boulesteix

## List of Publications by Year in descending order

Source: <https://exaly.com/author-pdf/3702281/publications.pdf>

Version: 2024-02-01

130  
papers

12,005  
citations

47006

47  
h-index

29157

104  
g-index

144  
all docs

144  
docs citations

144  
times ranked

17391  
citing authors

#	ARTICLE	IF	CITATIONS
1	Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics, 2007, 8, 25.	2.6	2,328
2	Conditional variable importance for random forests. BMC Bioinformatics, 2008, 9, 307.	2.6	2,129
3	Hyperparameters and tuning strategies for random forest. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2019, 9, e1301.	6.8	651
4	Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. Briefings in Bioinformatics, 2006, 8, 32-44.	6.5	611
5	Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2012, 2, 493-507.	6.8	516
6	Random forest versus logistic regression: a large-scale benchmark experiment. BMC Bioinformatics, 2018, 19, 270.	2.6	446
7	NetCoMi: network construction and comparison for microbiome data in R. Briefings in Bioinformatics, 2021, 22, .	6.5	222
8	Unbiased split selection for classification trees based on the Gini Index. Computational Statistics and Data Analysis, 2007, 52, 483-501.	1.2	201
9	An AUC-based permutation variable importance measure for random forests. BMC Bioinformatics, 2013, 14, 119.	2.6	179
10	Regularized estimation of large-scale gene association networks using graphical Gaussian models. BMC Bioinformatics, 2009, 10, 384.	2.6	169
11	PLS Dimension Reduction for Classification with Microarray Data. Statistical Applications in Genetics and Molecular Biology, 2004, 3, 1-30.	0.6	160
12	Stability and aggregation of ranked gene lists. Briefings in Bioinformatics, 2009, 10, 556-568.	6.5	138
13	Microarray-Based Prediction of Tumor Response to Neoadjuvant Radiochemotherapy of Patients With Locally Advanced Rectal Cancer. Clinical Gastroenterology and Hepatology, 2008, 6, 53-61.	4.4	128
14	Random forest for ordinal responses: Prediction and variable selection. Computational Statistics and Data Analysis, 2016, 96, 57-73.	1.2	125
15	AUC-RF: A New Strategy for Genomic Profiling with Random Forest. Human Heredity, 2011, 72, 121-132.	0.8	122
16	A computationally fast variable importance test for random forests for high-dimensional data. Advances in Data Analysis and Classification, 2018, 12, 885-915.	1.4	115
17	Essential guidelines for computational method benchmarking. Genome Biology, 2019, 20, 125.	8.8	114
18	What is the "normal" fetal heart rate?. PeerJ, 2013, 1, e82.	2.0	112

#	ARTICLE	IF	CITATIONS
19	Over-optimism in bioinformatics research. <i>Bioinformatics</i> , 2010, 26, 437-439.	4.1	110
20	Stability Investigations of Multivariable Regression Models Derived from Low- and High-Dimensional Data. <i>Journal of Biopharmaceutical Statistics</i> , 2011, 21, 1206-1231.	0.8	104
21	Predicting transcription factor activities from combined analysis of microarray and ChIP data: a partial least squares approach. <i>Theoretical Biology and Medical Modelling</i> , 2005, 2, 23.	2.1	98
22	Survival prediction using gene expression data: A review and comparison. <i>Computational Statistics and Data Analysis</i> , 2009, 53, 1590-1603.	1.2	98
23	Differential expression of the chemokines GRO-2, GRO-3, and interleukin-8 in colon cancer and their impact on metastatic disease and survival. <i>International Journal of Colorectal Disease</i> , 2010, 25, 573-581.	2.2	98
24	Random forest Gini importance favours SNPs with large minor allele frequency: impact, sources and recommendations. <i>Briefings in Bioinformatics</i> , 2012, 13, 292-304.	6.5	92
25	Over-optimism in bioinformatics: an illustration. <i>Bioinformatics</i> , 2010, 26, 1990-1998.	4.1	90
26	A Plea for Neutral Comparison Studies in Computational Sciences. <i>PLoS ONE</i> , 2013, 8, e61562.	2.5	86
27	CMA “a comprehensive Bioconductor package for supervised classification with high dimensional data. <i>BMC Bioinformatics</i> , 2008, 9, 439.	2.6	84
28	Time and rate of sinus formation in pilonidal sinus disease. <i>International Journal of Colorectal Disease</i> , 2008, 23, 359-364.	2.2	83
29	Extracorporeal Life Support in Cardiogenic Shock Complicating Acute Myocardial Infarction. <i>Journal of the American College of Cardiology</i> , 2019, 73, 2355-2357.	2.8	79
30	Molecular Differences between Chronic and Aggressive Periodontitis. <i>Journal of Dental Research</i> , 2013, 92, 1081-1088.	5.2	77
31	Microarray-based classification and clinical predictors: on combined classifiers and additional predictive value. <i>Bioinformatics</i> , 2008, 24, 1698-1706.	4.1	76
32	Cross-study validation for the assessment of prediction algorithms. <i>Bioinformatics</i> , 2014, 30, i105-i112.	4.1	75
33	A CART-based approach to discover emerging patterns in microarray data. <i>Bioinformatics</i> , 2003, 19, 2465-2472.	4.1	74
34	Evaluating Microarray-based Classifiers: An Overview. <i>Cancer Informatics</i> , 2008, 6, CIN.S408.	1.9	70
35	Subsampling Versus Bootstrapping in Resampling-Based Model Selection for Multivariable Regression. <i>Biometrics</i> , 2016, 72, 272-280.	1.4	70
36	On the necessity and design of studies comparing statistical methods. <i>Biometrical Journal</i> , 2018, 60, 216-218.	1.0	66

#	ARTICLE	IF	CITATIONS
37	Aprotinin and Anaphylaxis: Analysis of 12,403 Exposures to Aprotinin in Cardiac Surgery. <i>Annals of Thoracic Surgery</i> , 2007, 84, 1144-1150.	1.3	65
38	Machine learning versus statistical modeling. <i>Biometrical Journal</i> , 2014, 56, 588-593.	1.0	65
39	IPF-LASSO: Integrative $\langle \text{mml:math xmlns:mml="http://www.w3.org/1998/Math/MathML" id="M1">\langle \text{mml:mrow}\langle \text{mml:msub}\langle \text{mml:mrow}\langle \text{mml:mi}L\langle \text{mml:mi}\rangle \langle \text{mml:mrow}\langle \text{mml:mn fontstyle="italic">1\langle \text{mml:mn}\rangle \langle \text{mml:mrow}\langle \text{mml:msub}\langle \text{mml:mrow}\rangle \langle \text{mml:math}\rangle \text{Penalized Regression with Penalty Factors for Prediction Based on Multi-Omics Data. Computational and Mathematical Methods in Medicine, 2017, 2017, 1-14.$	1.3	62
40	Methylene Blue halves the long-term recurrence rate in acute pilonidal sinus disease. <i>International Journal of Colorectal Disease</i> , 2008, 23, 181-187.	2.2	59
41	An eight-gene expression signature for the prediction of survival and time to treatment in chronic lymphocytic leukemia. <i>Leukemia</i> , 2011, 25, 1639-1645.	7.2	59
42	Optimal classifier selection and negative bias in error rate estimation: an empirical study on high-dimensional prediction. <i>BMC Medical Research Methodology</i> , 2009, 9, 85.	3.1	56
43	Penalized Partial Least Squares with applications to B-spline transformations and functional data. <i>Chemometrics and Intelligent Laboratory Systems</i> , 2008, 94, 60-69.	3.5	55
44	Large-scale benchmark study of survival prediction methods using multi-omics data. <i>Briefings in Bioinformatics</i> , 2021, 22, .	6.5	53
45	Critical review of reporting of the data analysis step in metabolomics. <i>Metabolomics</i> , 2018, 14, 7.	3.0	52
46	Inhibition of radiation induced migration of human head and neck squamous cell carcinoma cells by blocking of EGF receptor pathways. <i>BMC Cancer</i> , 2011, 11, 388.	2.6	50
47	Critical appraisal of artificial intelligence-based prediction models for cardiovascular disease. <i>European Heart Journal</i> , 2022, 43, 2921-2930.	2.2	50
48	Assessment of predictive performance in incomplete data by combining internal validation and multiple imputation. <i>BMC Medical Research Methodology</i> , 2016, 16, 144.	3.1	49
49	Influence of diatoms on copepod reproduction. II. Uncorrelated effects of diatom-derived $\hat{1}\pm, \hat{1}^2, \hat{1}^3, \hat{1}$ -unsaturated aldehydes and polyunsaturated fatty acids on <i>Calanus helgolandicus</i> in the field. <i>Progress in Oceanography</i> , 2008, 77, 30-44.	3.2	48
50	Letter to the Editor: On the term 'interaction' and related phrases in the literature on Random Forests. <i>Briefings in Bioinformatics</i> , 2015, 16, 338-345.	6.5	48
51	Added predictive value of high-throughput molecular data to clinical data and its validation. <i>Briefings in Bioinformatics</i> , 2011, 12, 215-229.	6.5	46
52	Ten Simple Rules for Reducing Overoptimistic Reporting in Methodological Computational Research. <i>PLoS Computational Biology</i> , 2015, 11, e1004191.	3.2	43
53	On stability issues in deriving multivariable regression models. <i>Biometrical Journal</i> , 2015, 57, 531-555.	1.0	43
54	Towards evidence-based computational statistics: lessons from clinical research on the role and design of real-data benchmark studies. <i>BMC Medical Research Methodology</i> , 2017, 17, 138.	3.1	43

#	ARTICLE	IF	CITATIONS
55	Combining location-and-scale batch effect adjustment with data cleaning by latent factor adjustment. BMC Bioinformatics, 2016, 17, 27.	2.6	37
56	Improved prediction of complex diseases by common genetic markers: state of the art and further perspectives. Human Genetics, 2016, 135, 259-272.	3.8	37
57	Independent validation of a new reirradiation risk score (RRRS) for glioma patients predicting post-recurrence survival: A multicenter DTK/ROG analysis. Radiotherapy and Oncology, 2018, 127, 121-127.	0.6	37
58	Investigating the prediction ability of survival models based on both clinical and omics data: two case studies. Statistics in Medicine, 2014, 33, 5310-5329.	1.6	36
59	Live music reduces stress levels in very low birthweight infants. Acta Paediatrica, International Journal of Paediatrics, 2015, 104, 360-367.	1.5	36
60	Methylation of Tumor-Related Genes in Neoadjuvant-Treated Gastric Cancer: Relation to Therapy Response and Clinicopathologic and Molecular Features. Clinical Cancer Research, 2007, 13, 5095-5102.	7.0	35
61	Priority-Lasso: a simple hierarchical approach to the prediction of clinical outcome using multi-omics data. BMC Bioinformatics, 2018, 19, 322.	2.6	33
62	Testing the additional predictive value of high-dimensional molecular data. BMC Bioinformatics, 2010, 11, 78.	2.6	31
63	Effects of Aprotinin Dosage on Renal Function. Anesthesiology, 2008, 108, 189-198.	2.5	31
64	A Statistical Framework for Hypothesis Testing in Real Data Comparison Studies. American Statistician, 2015, 69, 201-212.	1.6	29
65	The multiplicity of analysis strategies jeopardizes replicability: lessons learned across disciplines. Royal Society Open Science, 2021, 8, 201925.	2.4	29
66	Validation of cluster analysis results on validation data: A systematic framework. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2022, 12, .	6.8	29
67	Maximally Selected Chi-square Statistics for Ordinal Variables. Biometrical Journal, 2006, 48, 451-462.	1.0	25
68	Outcome of patients treated with extracorporeal life support in cardiogenic shock complicating acute myocardial infarction: 1-year result from the ECLS-Shock study. Clinical Research in Cardiology, 2021, 110, 1412-1420.	3.3	24
69	Introduction to statistical simulations in health research. BMJ Open, 2020, 10, e039921.	1.9	24
70	On representative and illustrative comparisons with real data in bioinformatics: response to the letter to the editor by Smith et al. Bioinformatics, 2013, 29, 2664-2666.	4.1	22
71	A Replication Crisis in Methodological Research?. Significance, 2020, 17, 18-21.	0.4	21
72	Maximally Selected Chi-Square Statistics and Binary Splits of Nominal Variables. Biometrical Journal, 2006, 48, 838-848.	1.0	20

#	ARTICLE	IF	CITATIONS
73	On the choice and influence of the number of boosting steps for high-dimensional linear Cox-models. Computational Statistics, 2018, 33, 1195-1215.	1.5	20
74	On the optimistic performance evaluation of newly introduced bioinformatic methods. Genome Biology, 2021, 22, 152.	8.8	20
75	Interleukin-22 is elevated in lavage from patients with lung cancer and other pulmonary diseases. BMC Cancer, 2016, 16, 409.	2.6	19
76	WilcoxCV: an R package for fast variable selection in cross-validation. Bioinformatics, 2007, 23, 1702-1704.	4.1	18
77	Mid-term outcomes of patients with PCI prior to CABG in comparison to patients with primary CABG. Vascular Health and Risk Management, 2010, 6, 495.	2.3	18
78	Correcting the Optimal Resampling-Based Error Rate by Estimating the Error Rate of Wrapper Algorithms. Biometrics, 2013, 69, 693-702.	1.4	18
79	Improving cross-study prediction through add-on batch effect adjustment or add-on normalization. Bioinformatics, 2017, 33, 397-404.	4.1	18
80	Examining the robustness of observational associations to model, measurement and sampling uncertainty with the vibration of effects framework. International Journal of Epidemiology, 2021, 50, 266-278.	1.9	18
81	Pitfalls of hypothesis tests and model selection on bootstrap samples: Causes and consequences in biometrical applications. Biometrical Journal, 2016, 58, 447-473.	1.0	17
82	Multiple Testing for SNP-SNP Interactions. Statistical Applications in Genetics and Molecular Biology, 2007, 6, Article37.	0.6	15
83	Publication Bias in Methodological Computational Research. Cancer Informatics, 2015, 14s5, CIN.S30747.	1.9	15
84	A measure of the impact of CV incompleteness on prediction error estimation with application to PCA and normalization. BMC Medical Research Methodology, 2015, 15, 95.	3.1	15
85	Statistical learning approaches in the genetic epidemiology of complex diseases. Human Genetics, 2020, 139, 73-84.	3.8	14
86	A computationally fast variable importance test for random forests for high-dimensional data. Advances in Data Analysis and Classification, 2018, 12, 885.	1.4	14
87	Breath isoprene concentrations in persons undergoing general anesthesia and in healthy volunteers. Journal of Breath Research, 2012, 6, 046004.	3.0	13
88	Gram-negative and -positive bacteria differentiation in blood culture samples by headspace volatile compound analysis. Journal of Biological Research, 2016, 23, 3.	2.1	13
89	Making complex prediction rules applicable for readers: Current practice in random forest literature and recommendations. Biometrical Journal, 2019, 61, 1314-1328.	1.0	13
90	Maximally selected Chi-squared statistics and non-monotonic associations: An exact approach based on two cutpoints. Computational Statistics and Data Analysis, 2007, 51, 6295-6306.	1.2	12

#	ARTICLE	IF	CITATIONS
91	Centrosome abnormalities in head and neck squamous cell carcinoma (HNSCC). Acta Oto-Laryngologica, 2009, 129, 205-213.	0.9	12
92	Sampling uncertainty versus method uncertainty: A general framework with applications to omics biomarker selection. Biometrical Journal, 2020, 62, 670-687.	1.0	12
93	La genómica funcional: herramientas para mejorar la sanidad y el bienestar del ganado. OIE Revue Scientifique Et Technique, 2005, 24, 355-377.	1.2	12
94	Identification of interaction patterns and classification with applications to microarray data. Computational Statistics and Data Analysis, 2006, 50, 783-802.	1.2	11
95	Combining clinical and molecular data in regression prediction models: insights from a simulation study. Briefings in Bioinformatics, 2020, 21, 1904-1919.	6.5	11
96	Categorical variables with many categories are preferentially selected in bootstrap-based model selection procedures for multivariable regression models. Biometrical Journal, 2016, 58, 652-673.	1.0	10
97	A plea for taking all available clinical information into account when assessing the predictive value of omics data. BMC Medical Research Methodology, 2019, 19, 162.	3.1	10
98	Reader's Reaction to "Dimension Reduction for Classification with Gene Expression Microarray Data" by Dai et al (2006). Statistical Applications in Genetics and Molecular Biology, 2006, 5, Article16.	0.6	9
99	Mediation analysis reveals common mechanisms of RUNX1 point mutations and RUNX1/RUNX1T1 fusions influencing survival of patients with acute myeloid leukemia. Scientific Reports, 2018, 8, 11293.	3.3	9
100	Exhaled breath volatile organic and inorganic compound composition in end-stage renal disease. Clinical Nephrology, 2016, 86, 132-140.	0.7	9
101	Overoptimism in benchmark studies and the multiplicity of design and analysis options when interpreting their results. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2022, 12, .	6.8	9
102	Complexity Selection with Cross-validation for Lasso and Sparse Partial Least Squares Using High-Dimensional Data. Studies in Classification, Data Analysis, and Knowledge Organization, 2013, , 261-268.	0.2	7
103	Interaction forests: Identifying and exploiting interpretable quantitative and qualitative interaction effects. Computational Statistics and Data Analysis, 2022, 171, 107460.	1.2	7
104	Added predictive value of omics data: specific issues related to validation illustrated by two case studies. BMC Medical Research Methodology, 2014, 14, 117.	3.1	6
105	Effects of the Cardio First Angel™ on chest compression performance. Technology and Health Care, 2018, 26, 69-80.	1.2	6
106	Immune function testing in sepsis patients receiving sodium selenite. Journal of Critical Care, 2019, 52, 208-212.	2.2	5
107	Improved Outcome Prediction Across Data Sources Through Robust Parameter Tuning. Journal of Classification, 2021, 38, 212-231.	2.2	5
108	Homografts in aortic position: does blood group incompatibility have an impact on patient outcomes? Interactive Cardiovascular and Thoracic Surgery, 2013, 16, 619-624.	1.1	4

#	ARTICLE	IF	CITATIONS
109	The Residual-Based Predictiveness Curve: A Visual Tool to Assess the Performance of Prediction Models. <i>Biometrics</i> , 2016, 72, 392-401.	1.4	4
110	Detection of influential points as a byproduct of resampling-based variable selection procedures. <i>Computational Statistics and Data Analysis</i> , 2017, 116, 19-31.	1.2	4
111	On Fishing for Significance and Statistician's Degree of Freedom in the Era of Big Molecular Data. , 2017, , 155-170.		4
112	Over-optimistic evaluation and reporting of novel cluster algorithms: an illustrative study. <i>Advances in Data Analysis and Classification</i> , 2023, 17, 211-238.	1.4	4
113	Comments on: Augmenting the bootstrap to analyze high dimensional genomic data. <i>Test</i> , 2008, 17, 31-35.	1.1	3
114	Use of pretransformation to cope with extreme values in important candidate features. <i>Biometrical Journal</i> , 2011, 53, 673-688.	1.0	3
115	Editorial. <i>Briefings in Bioinformatics</i> , 2011, 12, 187-188.	6.5	3
116	Application of Microarray Analysis on Computer Cluster and Cloud Platforms. <i>Methods of Information in Medicine</i> , 2013, 52, 65-71.	1.2	3
117	Single-center versus multi-center data sets for molecular prognostic modeling: a simulation study. <i>Radiation Oncology</i> , 2020, 15, 109.	2.7	3
118	On the asymptotic behaviour of the variance estimator of a $U</math>-statistic. Journal of Statistical Planning and Inference, 2020, 209, 101-111.$	0.6	3
119	Benchmarking for Clustering Methods Based on Real Data: A Statistical View. <i>Studies in Classification, Data Analysis, and Knowledge Organization</i> , 2017, , 73-82.	0.2	3
120	COMPANION: development of a patient-centred complexity and casemix classification for adult palliative care patients based on needs and resource use – a protocol for a cross-sectional multi-centre study. <i>BMC Palliative Care</i> , 2022, 21, 18.	1.8	3
121	The Normal Fetal Heart Rate Study: Analysis Plan. <i>Nature Precedings</i> , 2007, , .	0.1	2
122	Gender mismatch in allograft aortic valve surgery. <i>Interactive Cardiovascular and Thoracic Surgery</i> , 2015, 21, 329-335.	1.1	2
123	The Normal Fetal Heart Rate Study: Analysis Plan. <i>Nature Precedings</i> , 2007, , .	0.1	1
124	Which Resampling-Based Error Estimator for Benchmark Studies? A Power Analysis with Application to PLS-LDA. <i>Springer Proceedings in Mathematics and Statistics</i> , 2016, , 45-57.	0.2	1
125	Iterative Reconstruction of High-Dimensional Gaussian Graphical Models Based on a New Method to Estimate Partial Correlations under Constraints. <i>PLoS ONE</i> , 2013, 8, e60536.	2.5	1
126	Use of Resampling Procedures to Investigate Issues of Model Building and Its Stability. , 2020, , 1-24.		1



#	ARTICLE	IF	CITATIONS
127	Letter to the Editor: On Reviews and Papers on New Methods. Briefings in Bioinformatics, 2015, 16, 365-366.	6.5	0
128	Guest Editorial“Special Collection Topic: Statistical Systems Theory in Cancer Modeling, Diagnosis, and Therapy. Cancer Informatics, 2018, 17, 117693511876094.	1.9	0
129	OC-0588: Validation of the reRT risk score (RRRS) in glioma patients: A multicenter DTK/ROG analysis. Radiotherapy and Oncology, 2018, 127, S307-S308.	0.6	0
130	Modelling Individual Response to Treatment and Its Uncertainty:A Review of Statistical Methods and Challenges for Future Research. Boston Studies in the Philosophy and History of Science, 2020, , 319-344.	0.9	0