

# Vipin Kumar

## List of Publications by Year in descending order

Source: <https://exaly.com/author-pdf/6721498/publications.pdf>

Version: 2024-02-01

90  
papers

7,997  
citations

201674

27  
h-index

149698

56  
g-index

95  
all docs

95  
docs citations

95  
times ranked

9099  
citing authors

#	ARTICLE	IF	CITATIONS
1	Top 10 algorithms in data mining. Knowledge and Information Systems, 2008, 14, 1-37.	3.2	4,153
2	Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data. IEEE Transactions on Knowledge and Data Engineering, 2017, 29, 2318-2331.	5.7	713
3	A daily global mesoscale ocean eddy dataset from satellite altimetry. Scientific Data, 2015, 2, 150028.	5.3	230
4	Process-Guided Deep Learning Predictions of Lake Water Temperature. Water Resources Research, 2019, 55, 9173-9190.	4.2	200
5	Discovery of Web Robot Sessions Based on their Navigational Patterns. Data Mining and Knowledge Discovery, 2002, 6, 9-35.	3.7	183
6	Enhancing data analysis with noise removal. IEEE Transactions on Knowledge and Data Engineering, 2006, 18, 304-319.	5.7	163
7	A Big Data Guide to Understanding Climate Change: The Case for Theory-Guided Data Science. Big Data, 2014, 2, 155-163.	3.4	160
8	Truncation and constitutive activation of the androgen receptor by diverse genomic rearrangements in prostate cancer. Nature Communications, 2016, 7, 13668.	12.8	134
9	Hyperclique pattern discovery. Data Mining and Knowledge Discovery, 2006, 13, 219-242.	3.7	132
10	Delay Within the 3-Hour Surviving Sepsis Campaign Guideline on Mortality for Patients With Severe Sepsis and Septic Shock*. Critical Care Medicine, 2018, 46, 500-505.	0.9	128
11	Parallel static and dynamic multi-constraint graph partitioning. Concurrency Computation Practice and Experience, 2002, 14, 219-240.	2.2	118
12	Physics Guided RNNs for Modeling Dynamical Systems: A Case Study in Simulating Lake Temperature Profiles. , 2019, , 558-566.		113
13	High spatiotemporal resolution of river planform dynamics from Landsat: The RivMAP toolbox and results from the Ucayali River. Earth and Space Science, 2017, 4, 46-75.	2.6	106
14	Document Categorization and Query Generation on the World Wide Web Using WebACE. Artificial Intelligence Review, 1999, 13, 365-391.	15.7	94
15	Mining strong affinity association patterns in data sets with skewed support distribution. , 0, , .		90
16	Associations between cortical thickness and general intelligence in children, adolescents and young adults. Intelligence, 2013, 41, 597-606.	3.0	90
17	Summarization " compressing data into an informative representation. Knowledge and Information Systems, 2007, 12, 355-378.	3.2	75
18	Parallel Formulations of Decision-Tree Classification Algorithms. Data Mining and Knowledge Discovery, 1999, 3, 237-261.	3.7	71

#	ARTICLE	IF	CITATIONS
19	Discovering genetic interactions bridging pathways in genome-wide association studies. Nature Communications, 2019, 10, 4274.	12.8	52
20	Predicting lake surface water phosphorus dynamics using process-guided machine learning. Ecological Modelling, 2020, 430, 109136.	2.5	50
21	Privacy Preserving Nearest Neighbor Search. , 2006, , .		46
22	Type 2 Diabetes Mellitus Trajectories and Associated Risks. Big Data, 2016, 4, 25-30.	3.4	46
23	Application of an Ontology for Characterizing Data Quality for a Secondary Use of EHR Data. Applied Clinical Informatics, 2016, 07, 69-88.	1.7	44
24	Predicting Water Temperature Dynamics of Unmonitored Lakes With Meta-Transfer Learning. Water Resources Research, 2021, 57, e2021WR029579.	4.2	42
25	Proteomic Profiles in Acute Respiratory Distress Syndrome Differentiates Survivors from Non-Survivors. PLoS ONE, 2014, 9, e109713.	2.5	40
26	Mapping Burned Areas in Tropical Forests Using a Novel Machine Learning Framework. Remote Sensing, 2018, 10, 69.	4.0	36
27	Physics-Guided Recurrent Graph Model for Predicting Flow and Temperature in River Networks. , 2021, , 612-620.		36
28	IDENTIFICATION OF FUNCTIONAL MODULES IN PROTEIN COMPLEXES VIA HYPERCLIQUE PATTERN DISCOVERY. , 2004, , .		36
29	Heterogeneous Metric Learning of Categorical Data with Hierarchical Couplings. IEEE Transactions on Knowledge and Data Engineering, 2018, 30, 1254-1267.	5.7	34
30	Mining Low-Support Discriminative Patterns from Dense and High-Dimensional Data. IEEE Transactions on Knowledge and Data Engineering, 2012, 24, 279-294.	5.7	33
31	Mining needle in a haystack. SIGMOD Record, 2001, 30, 91-102.	1.2	32
32	A graph-based approach to find teleconnections in climate data. Statistical Analysis and Data Mining, 2013, 6, 158-179.	2.8	29
33	Scalable parallel data mining for association rules. SIGMOD Record, 1997, 26, 277-288.	1.2	28
34	Can machine learning accelerate process understanding and decision-relevant predictions of river water quality?. Hydrological Processes, 2022, 36, .	2.6	26
35	A Data Quality Ontology for the Secondary Use of EHR Data. AMIA ... Annual Symposium proceedings, 2015, 2015, 1937-46.	0.2	25
36	TAPER: a two-step approach for all-strong-pairs correlation query in large databases. IEEE Transactions on Knowledge and Data Engineering, 2006, 18, 493-508.	5.7	23

#	ARTICLE	IF	CITATIONS
37	Generalizing the notion of confidence. Knowledge and Information Systems, 2007, 12, 279-299.	3.2	23
38	A Parameter-Free Spatio-Temporal Pattern Mining Model to Catalog Global Ocean Dynamics. , 2013, , .		23
39	EddyScan: A physically consistent ocean eddy monitoring application. , 2012, , .		21
40	Proteome Profiling in Lung Injury after Hematopoietic Stem Cell Transplantation. Biology of Blood and Marrow Transplantation, 2016, 22, 1383-1390.	2.0	18
41	Exploring the Predictability of 30-Day Extreme Precipitation Occurrence Using a Global SSTâ€“SLP Correlation Network. Journal of Climate, 2016, 29, 1013-1029.	3.2	18
42	Artificial Intelligence for Modeling Complex Systems: Taming the Complexity of Expert Models to Improve Decision Making. ACM Transactions on Interactive Intelligent Systems, 2021, 11, 1-49.	3.7	18
43	Quantifying the Effect of Data Quality on the Validity of an eMeasure. Applied Clinical Informatics, 2017, 08, 1012-1021.	1.7	17
44	Different Modes of Variability over the Tasman Sea: Implications for Regional Climate*. Journal of Climate, 2014, 27, 8466-8486.	3.2	16
45	Characterizing pattern preserving clustering. Knowledge and Information Systems, 2009, 19, 311-336.	3.2	15
46	Learning large-scale plantation mapping from imperfect annotators. , 2016, , .		15
47	A Teleconnection between the West Siberian Plain and the ENSO Region. Journal of Climate, 2017, 30, 301-315.	3.2	14
48	Bringing automated, remoteâ€“sensed, machine learning methods to monitoring crop landscapes at scale. Agricultural Economics (United Kingdom), 2019, 50, 41-50.	3.9	13
49	Toward Large-Scale Mapping of Tree Crops with High-Resolution Satellite Imagery and Deep Learning Algorithms: A Case Study of Olive Orchards in Morocco. Remote Sensing, 2021, 13, 1740.	4.0	12
50	Recurrent Generative Networks for Multi-Resolution Satellite Data: An Application in Cropland Monitoring. , 2019, , .		11
51	Post Classification Label Refinement Using Implicit Ordering Constraint Among Data Instances. , 2015, , .		10
52	Detection of Precursors to Aviation Safety Incidents Due to Human Factors. , 2013, , .		9
53	Predicting diabetes clinical outcomes using longitudinal risk factor trajectories. BMC Medical Informatics and Decision Making, 2020, 20, 6.	3.0	9
54	Localized prediction of continuous target variables using hierarchical clustering. , 0, , .		8

#	ARTICLE	IF	CITATIONS
55	Association of BMI, comorbidities and all-cause mortality by using a baseline mortality risk model. PLoS ONE, 2021, 16, e0253696.	2.5	8
56	Prediction of HLA Genes from SNP Data and HLA Haplotype Frequencies. , 2012, , .		6
57	Forensic Style Analysis with Survival Trajectories. , 2015, , .		6
58	A Computational Method for Learning Disease Trajectories From Partially Observable EHR Data. IEEE Journal of Biomedical and Health Informatics, 2021, 25, 2476-2486.	6.3	6
59	OUP accepted manuscript. Journal of the American Medical Informatics Association: JAMIA, 2021, 29, 72-79.	4.4	6
60	Association Analysis Techniques for Discovering Functional Modules from Microarray Data. Nature Precedings, 0, , .	0.1	5
61	SubPatCNV: approximate subspace pattern mining for mapping copy-number variations. BMC Bioinformatics, 2015, 16, 16.	2.6	5
62	GRAPH PARTITIONING FOR DYNAMIC, ADAPTIVE AND MULTI-PHASE SCIENTIFIC SIMULATIONS. , 2002, , .		5
63	A Data Mining Approach to Determine Sepsis Guideline Impact on Inpatient Mortality and Complications. AMIA Summits on Translational Science Proceedings, 2016, 2016, 194-202.	0.4	5
64	Privacy leakage in multi-relational databases: a semi-supervised learning perspective. VLDB Journal, 2006, 15, 388-402.	4.1	4
65	Efficient Algorithms for Computing With Protein-Based Volumetric Memory Processors. IEEE Nanotechnology Magazine, 2011, 10, 881-890.	2.0	4
66	Importance of vegetation type in forest cover estimation. , 2012, , .		4
67	A reference based analysis framework for understanding anomaly detection techniques for symbolic sequences. Data Mining and Knowledge Discovery, 2014, 28, 702-735.	3.7	4
68	Adaptive Heterogeneous Ensemble Learning Using the Context of Test Instances. , 2015, , .		4
69	Global River Monitoring Using Semantic Fusion Networks. Water (Switzerland), 2020, 12, 2258.	2.7	4
70	Online Change Detection Algorithm for Noisy Time-Series: An Application Tonear-Real Time Burned Area Mapping. , 2015, , .		3
71	Automated Plantation Mapping in Southeast Asia Using MODIS Data and Imperfect Visual Annotations. Remote Sensing, 2020, 12, 636.	4.0	3
72	Learning with Small Data. , 2020, , .		3

#	ARTICLE	IF	CITATIONS
73	Evaluating the Impact of Data Representation on EHR-Based Analytic Tasks. Studies in Health Technology and Informatics, 2019, 264, 288-292.	0.3	3
74	Toward Real-Time Motion Planning. Machine Intelligence and Pattern Recognition, 1994, 15, 163-175.	0.2	3
75	Predicting the Performance of Randomized Parallel Search: An Application to Robot Motion Planning. Journal of Intelligent and Robotic Systems: Theory and Applications, 2003, 38, 31-53.	3.4	2
76	Identifying dynamic changes with noisy labels in spatial-temporal data: A study on large-scale water monitoring application. , 2016, , .		2
77	Estimating Disease Onset Time by Modeling Lab Result Trajectories via Bayes Networks. , 2017, 2017, 374-379.		2
78	Plantation Mapping in Southeast Asia. Frontiers in Big Data, 2019, 2, 46.	2.9	2
79	Association Analysis Techniques for Discovering Functional Modules from Microarray Data. Nature Precedings, 0, , .	0.1	2
80	A Framework for Visualizing Data Quality for Predictive Models and Clinical Quality Measures. AMIA Summits on Translational Science Proceedings, 2019, 2019, 630-638.	0.4	2
81	Summary Report of SDM 2008 Panel: Perspectives on Research Directions and Trends for the Data Mining Research Community. Statistical Analysis and Data Mining, 2008, 1, 211-213.	2.8	1
82	Systematic Evaluation of Scaling Methods for Gene Expression Data. , 2008, , .		1
83	Association analysis techniques for analyzing complex biological data sets. , 2009, , .		1
84	Spatially penalized regression for dependence analysis of rare events: A study in precipitation extremes. , 2012, , .		1
85	Evaluation of Label Dependency for the Prediction of HLA Genes. , 2013, , .		1
86	Identification of co-occurring insertions in cancer genomes using association analysis. , 2010, , .		0
87	Approximate Search on Massive Spatiotemporal Datasets. , 2012, , .		0
88	Building Predictive Models for Noisy and Heterogeneous Data: An Application in Global Monitoring of Inland Water Dynamics. , 2015, , .		0
89	Joint sparse auto-encoder: A semi-supervised spatio-temporal approach in mapping large-scale croplands. , 2017, , .		0
90	Report from the workshop on distributed and parallel knowledge discovery, ACM SIGKDD-2000. SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) on Knowledge Discovery & Data Mining, 2000, 2, 108-109.	4.0	0