

Table 1: The 50 most cited classification papers using the KEEL dataset. Only 16 out of 50 (32%) papers mentioned that data was scaled. Only 12 informed which scaling technique was used. Min-max was said to be used on 10 of the 12 papers that informed the technique.

#	Paper title	DOI	Mentioned that data was scaled?	Scaling technique(s)
1	A generalized mean distance-based k-nearest neighbor classifier	https://doi.org/10.1016/j.eswa.2018.08.021	No	
2	Under-sampling class imbalanced datasets by combining clustering analysis and instance selection	https://doi.org/10.1016/j.ins.2018.10.029	No	
3	Cost-sensitive and hybrid-attribute measure multi-decision tree over imbalanced data sets	https://doi.org/10.1016/j.ins.2017.09.013	No	
4	Dynamic selection of normalization techniques using data complexity measures	https://doi.org/10.1016/j.eswa.2018.04.008	Yes	Min-max, z-score
5	Class-specific extreme learning machine for handling binary class imbalance problem	https://doi.org/10.1016/j.neunet.2018.05.011	Yes	Min-max [-1,1]
6	A Local Mean Representation-based K-Nearest Neighbor Classifier	https://dl.acm.org/doi/10.1145/3319532	No	
7	General twin support vector machine with pinball loss function	https://doi.org/10.1016/j.ins.2019.04.032	Yes	Not informed
8	Amended fused TOPSIS-VIKOR for classification (ATOVIC) applied to some UCI data sets	https://doi.org/10.1016/j.eswa.2018.01.025	Yes	Not informed
9	IMCStacking: Cost-sensitive stacking learning with feature inverse mapping for imbalanced problems	https://dl.acm.org/doi/abs/10.1016/j.knosys.2018.02.031	No	
10	Class imbalance learning using UnderBagging based kernelized extreme learning machine	https://doi.org/10.1016/j.neucom.2018.10.056	Yes	Min-max [-1,1]
11	UnderBagging based reduced Kernelized weighted extreme learning machine for class imbalance learning	https://doi.org/10.1016/j.engappai.2018.07.002		
12	Class-specific kernelized extreme learning machine for binary class imbalance learning	https://doi.org/10.1016/j.asoc.2018.10.011	Yes	Min-max [-1,1]
13	A feature selection technique based on rough set and improvised PSO algorithm (PSORS-FS) for permission based detection of Android malwares	https://doi.org/10.1007/s13042-018-0838-1	No	
14	Improving the accuracy of k-nearest neighbor using local mean based and distance weight	http://dx.doi.org/10.1088/1742-6596/978/1/012047	No	
15	Hybrid Incremental Ensemble Learning for Noisy Real-World Data Classification	https://doi.org/10.1109/TCYB.2017.2774266	No	
16	Sparse Supervised Representation-Based Classifier for Uncontrolled and Imbalanced Classification	https://doi.org/10.1109/TNNLS.2018.2884444	Yes	unit l2-norm
17	Robust adaptive learning approach to self-organizing maps	https://doi.org/10.1016/j.knosys.2019.01.011	Yes	Min-max
18	Cluster-based zero-shot learning for multivariate data	https://doi.org/10.1007/s12652-020-02268-5	Yes	Min-max
19	Imbalanced learning based on data-partition and SMOTE	https://doi.org/10.3390/info9090238	No	
20	A new locally adaptive k-nearest neighbor algorithm based on discrimination class	https://doi.org/10.1016/j.knosys.2020.106185	No	
21	Class-specific cost-sensitive boosting weighted ELM for class imbalance learning	https://doi.org/10.1007/s12293-018-0267-4	Yes	Min-max [-1,1]
22	Integration of an improved dynamic ensemble selection approach to enhance one-vs-one scheme	https://doi.org/10.1016/j.engappai.2018.06.002	No	
23	Attribute Weighting Based K-Nearest Neighbor Using Gain Ratio	https://doi.org/10.1088/1742-6596/1007/1/012007	No	
24	Common and special knowledge-driven TSK fuzzy system and its modeling and application for epileptic EEG signals recognition	https://doi.org/10.1109/ACCESS.2019.2937657	No	
25	A novel dynamic ensemble selection classifier for an imbalanced data set: An application for credit risk assessment	http://dx.doi.org/10.1016/j.knosys.2020.106462	No	

Table 1 continued from previous page

#	Paper title	DOI	Mentioned that data was scaled?	Scaling technique(s)
26	NBWELM: Naive Bayesian based weighted extreme learning machine	https://doi.org/10.1007/s13042-014-0318-1	No	
27	Medical diagnosis of chronic diseases based on a novel computational intelligence algorithm	https://doi.org/10.3217/jucs-024-06-0775	No	
28	An adaptive fuzzy K-nearest neighbor approach for MR brain tumor image classification using parameter free bat optimization algorithm	https://doi.org/10.1007/s11042-019-7498-3	Yes	Not informed
29	A clustering based ensemble of weighted kernelized extreme learning machine for class imbalance learning	https://doi.org/10.1016/j.eswa.2020.114041	Yes	Min-max [-1,1]
30	A Novel Diversity Measure and Classifier Selection Approach for Generating Ensemble Classifiers	https://doi.org/10.1109/ACCESS.2019.2949059	Yes	z-score
31	Semi-supervised anomaly detection algorithms: A comparative summary and future research directions	https://doi.org/10.1016/j.knosys.2021.106878	No	
32	Cosine K-nearest neighbor in milkfish eye classification	https://doi.org/10.22266/ijies2020.0630.02	No	
33	Classifying imbalanced data using ensemble of reduced kernelized weighted extreme learning machine	https://doi.org/10.1007/s13042-019-01001-9	Yes	Min-max [-1,1]
34	Experimental study on generalization capability of extended naive bayesian classifier	https://doi.org/10.1007/s13042-014-0311-8	No	
35	An intelligent feature selection approach based on moth flame optimization for medical diagnosis	https://doi.org/10.1007/s00521-020-05483-5	No	
36	Cost-sensitive learning classification strategy for predicting product failures	https://doi.org/10.1016/j.eswa.2020.113653	No	
37	EHSO: Evolutionary Hybrid Sampling in overlapping scenarios for imbalanced learning	https://doi.org/10.1016/j.neucom.2020.08.060	No	
38	Parameter-Free Extreme Learning Machine for Imbalanced Classification	https://doi.org/10.1007/s11063-020-10282-z	Yes	Not informed
39	An Improved Oversampling Method for imbalanced Data-SMOTE Based on Canopy and K-means	https://doi.org/10.1109/CAC48633.2019.8997367	No	
40	Health stages diagnostics of underwater thruster using sound features with imbalanced dataset	https://doi.org/10.1007/s00521-018-3407-3	No	
41	Multi-view learning with fisher kernel and bi-bagging for imbalanced problem	https://doi.org/10.1007/s10489-019-01428-1	No	
42	Feature Selection for Classification through Population Random Search with Memory	https://doi.org/10.1134/S0005117919020103	No	
43	Centroid Neural Network with Pairwise Constraints for Semi-supervised Learning	https://doi.org/10.1007/s11063-018-9794-8	No	
44	Ensemble Strategy for Hard Classifying Samples in Class-Imbalanced Data Set	https://doi.org/10.1109/BigComp.2018.00033	No	
45	Least squares KNN-based weighted multiclass twin SVM	https://doi.org/10.1016/j.neucom.2020.02.132	No	
46	A multiclass classification using one-versus-all approach with the differential partition sampling ensemble	https://doi.org/10.1016/j.engappai.2020.104034	No	
47	A Hybrid Approach for Heart Disease Diagnosis and Prediction Using Machine Learning Techniques	https://doi.org/10.1007/978-3-030-51517-1_26	No	
48	Adaptive geometric median prototype selection method for k-nearest neighbors classification	https://doi.org/10.3233/IDA-184190	No	
49	Functional expansions based multilayer perceptron neural network for classification task	https://doi.org/10.13053/cys-22-4-2602	Yes	Min-max[0.2, 0.8]
50	Ensemble and fuzzy techniques applied to imbalanced traffic congestion datasets: A comparative study	https://doi.org/10.1007/978-3-319-91641-5_16	No	
TOTALS			16 (32%)	