algorithms

MDPI

Editorial

Special Issue "Algorithms in Data Classification"

Ioannis G. Tsoulos

Department of Informatics and Telecommunications, University of Ioannina, 45110 Ioannina, Greece; itsoulos@uoi.gr

Data classification is a well-known procedure, with many applications to real-world problems. This procedure can be divided into several subcategories, such as:

1. Binary classification.
2. Multi-class classification.
3. Multi-label classification.
4. Imbalanced classification.
5. Feature selection for classification.
6. Probabilistic models for classification.
7. Big data classification.
8. Text classification.
9. Multimedia classification.
10. Uncertain data classification.

Many methods have been used in data classification, such as Bayes methods [1], K-NN models [2], decision trees [3], Support Vector Machines (SVMs) [4], artificial neural networks [5], etc. The applications of data classification may include sentiment analysis [6], spam classification [7], document classification [8], image classification [9], etc. This Special Issue presents twelve papers that are analyzed in the following. The published papers in this Special Issue cover a wide range of applications from the field of data classification. The List of Contributions for this Special Issue is briefly described in the following.

Contribution 1 discusses the impact of class imbalance on the performance of Gaussian mixture models in classification tasks. The paper analyzes various methods to tackle this problem, such as cost-sensitive learning, threshold adjustments, and sampling-based techniques, and their effectiveness is thoroughly analyzed in a series of experiments.

Contribution 2 proposes a combination of various techniques, like deep learning techniques and custom CNN architectures, for the blood cell classification problem. The paper utilizes a series of methods, such as fine-tuning and ensemble strategies (CBAM and averaging ensembles), to achieve unprecedented accuracy and interpretability.

Contribution 3 discusses the problem of student dropout in online education. The authors collected data from three different LMS Moodle servers and utilized a deep learning algorithm to correctly classify the collected data.

Contribution 4 discusses the recent literature on the problem of constructing confidence ellipses in multidimensional scaling (MDS) solutions and proposes a new method based on a hierarchical cluster analysis (HCA) to tackle this problem.

Additionally, Contribution 5 deals with the energy consumption problem in industrial plants by proposing a mixture of regression models. The implemented algorithm combines autocorrelated observations with spline and polynomial regressions.

Contribution 6 proposes an intelligent system that utilizes data mining algorithms for the rehabilitation guidance of recreational runners with musculoskeletal discomfort. The system classifies recreational runners based on a questionnaire and proceeds to design specialized exercises for them to perform.

The authors of Contribution 7 propose the HyperGE method, which is a two-stage model for automatically tuning the hyperparameters of artificial neural networks. This

Citation: Tsoulos, I.G. Special Issue "Algorithms in Data Classification". *Algorithms* **2024**, *17*, 5. https://doi.org/10.3390/a17010005

Received: 19 December 2023
Accepted: 21 December 2023
Published: 22 December 2023

process is guided by the grammatical evolution [10] method. The proposed method was evaluated by conducting experiments on several datasets from the relevant literature.

Another problem that concerns the scientific community is the big data classification problem, handled in Contribution 8, where the authors suggest the use of fuzzy cognitive maps (FCMs) to efficiently handle this problem. The authors suggest using FCMs to select the most relevant features from the given dataset in order to improve the classification accuracy.

In Contribution 9, the authors propose a new cooperative multi-population differential evolution [11] to determine the optimal number of clusters in data. The proposed method was applied on series of datasets from the UCI machine learning repository and it was compared against other classification methods from the relevant literature.

Contribution 10 discusses the application of various techniques for the classification of different types of acute psychological stress (APS) and physical activity (PA), utilizing data collected via a wristband device. The original features are produced using random convolutional kernel transformation and, subsequently, feature selection techniques, such as principal component analysis and partial least squares–discriminant analysis (PLS-DA), among others, are used for data classification. Furthermore, a long short-term memory recurrent neural network model is utilized to estimate PA and APS from the data collected via wristband devices.

Contribution 11 presents specialized software for data classification. The software creates artificial features from the original ones using grammatical evolution and artificial neural networks. The produced features are subsequently used to train another machine learning model. The software is general enough to also be applied to data classification problems and regression problems. Furthermore, in order to speed up the process, the software utilizes modern parallel programming techniques.

Contribution 12 utilizes artificial intelligence models for learning the discriminant spectral information stored in the vibration signals, such as one-dimensional convolutional neural networks (1D-CNNs) and multi-layer perceptron (MLP) neural networks.

Finally, as the guest editor, it was my pleasure to work with the editorial staff of the *Algorithms* journal to prepare this Special Issue.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflicts of interest.

List of Contributions

1. Scrucca, L. On the Influence of Data Imbalance on Supervised Gaussian Mixture Models. *Algorithms* **2023**, *16*, 563. https://doi.org/10.3390/a16120563.
2. Abou Ali, M.; Dornaika, F.; Arganda-Carreras, I. Blood Cell Revolution: Unveiling 11 Distinct Types with 'Naturalize' Augmentation. *Algorithms* **2023**, *16*, 562. https://doi.org/10.3390/a16120562.
3. Porras, J.M.; Lara, J.A.; Romero, C.; Ventura, S. A Case-Study Comparison of Machine Learning Approaches for Predicting Students Dropout from Multiple Online Educational Entities. *Algorithms* **2023**, *16*, 554. https://doi.org/10.3390/a16120554.
4. Nikitas, P.; Nikita, E. Assessing Algorithms Used for Constructing Confidence Ellipses in Multidimensional Scaling Solutions. *Algorithms* **2023**, *16*, 535. https://doi.org/10.3390/a16120535.
5. Pacella, M.; Mangini, M.; Papadia, G. Utilizing Mixture Regression Models for Clustering Time-Series Energy Consumption of a Plastic Injection Molding Process. *Algorithms* **2023**, *16*, 524. https://doi.org/10.3390/a16110524.
6. Tzelepis, T.; Matlis, G.; Dimokas, N.; Karvelis, P.; Malliou, P.; Beneka, A. An Intelligent Injury Rehabilitation Guidance System for Recreational Runners Using Data Mining Algorithms. *Algorithms* **2023**, *16*, 523. https://doi.org/10.3390/a16110523.

7. Vaidya, G.; Kshirsagar, M.; Ryan, C. Grammatical Evolution-Driven Algorithm for Efficient and Automatic Hyperparameter Optimisation of Neural Networks. *Algorithms* **2023**, *16*, 319. https://doi.org/10.3390/a16070319.
8. Haritha, K.; Judy, M.V.; Papageorgiou, K.; Georgiannis, V.C.; Papageorgiou, E. Distributed Fuzzy Cognitive Maps for Feature Selection in Big Data Classification. *Algorithms* **2022**, *15*, 383. https://doi.org/10.3390/a15100383.
9. Banerjee, A.; Abu-Mahfouz, I. A Novel Adaptive FCM with Cooperative Multi-Population Differential Evolution Optimization. *Algorithms* **2022**, *15*, 380. https://doi.org/10.3390/a15100380.
10. Askari, M.R.; Abdel-Latif, M.; Rashid, M.; Sevil, M.; Cinar, A. Detection and Classification of Unannounced Physical Activities and Acute Psychological Stress Events for Interventions in Diabetes Treatment. *Algorithms* **2022**, *15*, 352. https://doi.org/10.3390/a15100352.
11. Tsoulos, I.G. QFC: A Parallel Software Tool for Feature Construction, Based on Grammatical Evolution. *Algorithms* **2022**, *15*, 295. https://doi.org/10.3390/a15080295.
12. Akpudo, U.E.; Hur, J.W. Exploring the Efficiencies of Spectral Isolation for Intelligent Wear Monitoring of Micro Drill Bit Automatic Regrinding In-Line Systems. *Algorithms* **2022**, *15*, 194. https://doi.org/10.3390/a15060194.

References

1. Kim, S.B.; Han, K.S.; Rim, H.C.; Myaeng, S.H. Some Effective Techniques for Naive Bayes Text Classification. *IEEE Trans. Knowl. Data Eng.* **2006**, *18*, 1457–1466. [CrossRef]
2. Guo, G.; Wang, H.; Bell, D.; Bi, Y.; Greer, K. KNN Model-Based Approach in Classification. In *Proceedings of the on the Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*; Meersman, R., Tari, Z., Schmidt, D.C., Eds.; Springer: Berlin/Heidelberg, Germany, 2003; pp. 986–996.
3. Quinlan, J. Simplifying decision trees. *Int. J. Man-Mach. Stud.* **1987**, *27*, 221–234. [CrossRef]
4. Pisner, D.A.; Schnyer, D.M. Chapter 6—Support vector machine. In *Machine Learning*; Mechelli, A., Vieira, S., Eds.; Academic Press: Cambridge, MA, USA, 2020; pp. 101–121. [CrossRef]
5. Basheer, I.; Hajmeer, M. Artificial neural networks: Fundamentals, computing, design, and application. *J. Microbiol. Methods* **2000**, *43*, 3–31. [CrossRef] [PubMed]
6. Ouyang, X.; Zhou, P.; Li, C.H.; Liu, L. Sentiment Analysis Using Convolutional Neural Network. In Proceedings of the 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, Liverpool, UK, 26–28 October 2015; pp. 2359–2364. [CrossRef]
7. Yu, B.; Ben Xu, Z. A comparative study for content-based dynamic spam classification using four machine learning algorithms. *Knowl.-Based Syst.* **2008**, *21*, 355–362. [CrossRef]
8. Manevitz, L.M.; Yousef, M. One-class SVMs for document classification. *J. Mach. Learn. Res.* **2001**, *2*, 139–154.
9. Lu, D.; Weng, Q. A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* **2007**, *28*, 823–870. [CrossRef]
10. O'Neill, M.; Ryan, C. Grammatical evolution. *IEEE Trans. Evol. Comput.* **2001**, *5*, 349–358. [CrossRef]
11. Bilal; Pant, M.; Zaheer, H.; Garcia-Hernandez, L.; Abraham, A. Differential Evolution: A review of more than two decades of research. *Eng. Appl. Artif. Intell.* **2020**, *90*, 103479. [CrossRef]