algorithms

Editorial

Special Issue on Ensemble Learning and/or Explainability

Panagiotis Pintelas * and Ioannis E. Livieris *

Department of Mathematics, University of Patras, 265-00 GR Patras, Greece
* Correspondence: ppintelas@gmail.com (P.P.); livieris@upatras.gr (I.E.L.)

Citation: Pintelas, P.; Livieris, I.E. Special Issue on Ensemble Learning and/or Explainability. *Algorithms* **2023**, *16*, 49. https://doi.org/10.3390/a16010049

Received: 5 January 2023
Accepted: 10 January 2023
Published: 11 January 2023

1. Introduction

This article will summarize the works published in a Special Issue of *Algorithms*, entitled "*Ensemble Learning and/or Explainability*"(https://www.mdpi.com/journal/algorithms/special_issues/Ensemble_Algorithms_Explainability, accessed on 24 November 2020). The main aim of this Special Issue is to present the recent advances related to a wide range of ensemble learning algorithms and methodologies and investigate the impact of their application in a diversity of real-world problems. At the same time, the need to research the explainability issues involved in theory and practice is of paramount importance for all kinds of daily and industrial applications. A total of eight papers were accepted, after undergoing a rigorous peer-review process based on their scientific merit and other quality criteria. All accepted papers possess considerable elements of novelty and cover a diverse range of application domains, providing the reader with a glimpse of the state-of-the-art research in the machine learning (ML) area. This Special Issue can be considered the sequel (second volume) of "*Ensemble learning and their applications*" [1].

During the last few decades, in the area of machine learning and data mining, ensemble methods constitute a state-of-the-art option for the development of powerful and robust prediction models. These models exploit the individual predictions of a variety of constituent learning algorithms to obtain better prediction performance, which has been proved both theoretically and experimentally. Thus, many ensemble learning algorithms have been proposed in the literature and have found applications in various real-word problems ranging from face and emotion recognition through text classification and medical diagnosis to financial forecasting, among others.

Quite recently, the European Union General Data Protection Regulation (GDPR) demanded a "*right to explanation*" about decisions made by automated and artificial intelligent algorithmic systems. This demand, combined with the recognized need to Interpret or Explain and justify the decisions/predictions of a classifier or ensemble, led to the development of "*Interpretable/Explainable Machine Learning and Artificial Intelligence*", which has earned great attention from the scientific community.

Therefore, the needs of this data-driven era, as well as the challenging necessities of the industrial sector, resulted in the development of robust, accurate, and explainable prediction models based on novel ensemble-learning methods and explainability techniques.

2. Ensemble Learning and/or Explainability

The first paper in this Special Issue is entitled "*The Study of Multiple Classes Boosting Classification Method Based on Local Similarity*" and it is authored by Wang and Chen [2]. This paper presents a new approach to cross-modal retrieval based on a simple enhancement framework which utilizes local similarity and it is extended to multimodal, multiclass enhancement Boosting. To achieve mutual retrieval of text and images, local similarity is utilized for the development of weak learners and a Boosting technique is used to create a powerful ensemble prediction model. The authors theoretically proved that in their proposed framework, the training loss is exponentially minimized. Additionally, they conducted a series of experiments using two co-opened multimodal databases (NUS-WIDE and Wiki) to verify and evaluate the performance of their proposed algorithm. Their experimental results illustrated that their approach has considerable advantages for cross-mode retrieval compared to traditional methods.

The second paper is entitled "*Ensembling EfficientNets for the Classification and Interpretation of Histopathology Images*" and it is authored by Kallipolitis et al. [3]. In this work, the authors proposed an ensemble scheme for the classification of histopathology breast and colon cancer images. In more detail, they implemented an ensemble classifier based on the state-of-the-art EfficientNets networks, which was combined with a Grad-CAM explanation scheme used to highlight the visual patterns responsible for each class prediction. The presented numerical experiments provided empirical evidence that the proposed approach is able to implement a robust, interpretable, and explainable histopathology image classifier. Furthermore, the authors developed a standalone application (web service) following the principles of distributed computing for online evaluation, validation, and experimentation.

The third paper is entitled "*A Rule Extraction Technique Applied to Ensembles of Neural Networks, Random Forests, and Gradient-Boosted Trees*". Bologna [4] highlighted the considerable disadvantage of neural network ensemble models in explaining their responses, namely their lack of interpretability or transparency with regard to how the input data are processed to model's outputs. In this work, a new rule extraction method is presented which is able to locate the discriminating hyperplanes that constitute the antecedents of the decision rules. The proposed rule extraction method was applied to decision tree and multilayer perceptron ensembles, which were evaluated based on eight classification benchmarks from UCI repository. The presented experimental analysis showed that the proposed ensembles reported competitive performance with that of "Skope-Rules" [5] and other rule extraction techniques with respect to the characteristics of the generated rules.

The fourth paper is entitled "*Precision-Based Weighted Blending Distributed Ensemble Model for Emotion Classification*" and it is authored by Soman et al. [6]. In this study, the authors conduct emotion classification by exploring the capabilities of a distributed ensemble-based model utilizing precision-based weighted blending. The proposed model works in a distributed manner using the SparkML pipeline [7], which enhances the model's performance. The considerable advantages of this approach are the effectiveness of iterative computations and the model's quick in-memory processing capabilities. The experimental analysis included three well-known emotional benchmarks (the FER-2013 image dataset, the extended Cohn–Kanade facial expression dataset, and the FERG-DB dataset). Based on their numerical experiments, the authors stated that the proposed ensemble model network outperforms other state-of-the-art models, as it exploits the probabilities of each class along with the precision value of each classifier.

The fifth paper is entitled "*Prediction of Injuries in CrossFit Training: A Machine Learning Perspective*". It was authored by Moustakidis et al. [8]. The main scope of this paper was the identification of risk factors, as well as the development of ML models using ensemble-learning techniques to predict CrossFit injuries. To this end, the authors conducted a survey-based epidemiological study in Greece to collect data on musculoskeletal injuries in CrossFit practitioners. They also provided a detailed discussion on the nature of the selected features to increase the reader's understanding of their contribution to the task. The experimental analysis was based on an evaluation of several ML and ensemble models, while the best overall performance was reported by the Adaboost model for the group of six selected risk factors.

The sixth paper is entitled "*A Seed-Guided Latent Dirichlet Allocation Approach to Predict the Personality of Online Users Using the PEN Model*" and it is authored by Sagadevan et al. [9]. In this research, the authors proposed a new unsupervised model called SLDA (Seed-guided Latent Dirichlet Allocation) to classify social network (SN) messages according to personality (Psychoticism, Extraversion, and Neuroticism—PEN) traits. The proposed SLDA was compared against other topic models, such as LDA and Latent Semantic Analysis, which displayed superior performance. Additionally, the authors performed a comparative analysis using several classifiers to classify the ground-truth topics generated by the SLDA algorithm, which revealed that most of the classifiers are able to successfully predict the traits classes generated by SLDA.

The seventh paper is authored by Zou and Gao [10] and it is entitled "*Extreme Learning Machine Enhanced Gradient Boosting for Credit Scoring*". The authors aimed to enhance the diversity of base learners by incorporating the advantages of the Bagging training strategy, as well as a Boosting ensemble optimization pattern for credit scoring. They also proposed a new supervised efficient neural network-based augmented Gradient Boosting Decision Tree (GBDT), called AugBoost-ELM, which inherits the Boosting training pattern from the GBDT framework. Additionally, the authors stated that a significant advantage of the proposed approach is that the training strategy avoids the problem of falling into a local minimum, leading to the efficient generation of augmented features. AugBoost-ELM was evaluated on three well-known and widely utilized credit benchmarks (Australian, German, Japanese, Taiwan) from UCI repository against a variety of state-of-the-art single and ensemble models reporting the best overall performance.

Finally, the eighth paper is entitled "*Ensembles of random SHAPs*" and it is authored by Utkin and Konstantinov [11]. In this work, the authors proposed three ensemble-based modifications to the SHapley Additive exPlanations (SHAP) method for the local explanation of a black-box model, called ER-SHAP, ERW-SHAP, and ER-SHAP-RF. The rationale behind their approach was to approximate the SHAP values using an ensemble of SHAPs with a relative smaller number of features to reduce the computational cost. In ER-SHAP modification, a random subset of features from the feature set is selected many times while the Shapley value for each feature is computed by means of small SHAPs. In ERW-SHAP modification, a variety of points are generated around the explained instance, and the results of their explanation are combined, considering their weight. Finally, ER-SHAP-RF modification utilizes the random forest model to provide a preliminary explanation of the instances and determine a feature probability distribution, which is then applied to the selection of the features in ER-SHAP. Additionally, the authors provided a series of numerical experiments, which presented the the efficiency and properties of all modifications for a local explanation.

3. Conclusions and Future Approaches

The motivation and the rationale behind this Special Issue was to provide a contribution to the existing literature about *ensemble learning* and *explainability techniques*. It is anticipated that the interesting approaches presented in this Special Issue will prove constructive and will be deeply appreciated by the scientific community. Finally, a considerable expectation of this collection is that the presented research works will further inspire research on innovative ensemble learning methods and explainability techniques in various challenging domains used to develop powerful as well as explainable prediction models.

Author Contributions: Conceptualization, P.P. and I.E.L.; methodology, P.P. and I.E.L.; software, P.P. and I.E.L.; validation, P.P. and I.E.L.; formal analysis, P.P. and I.E.L.; investigation, P.P. and I.E.L.; resources, P.P. and I.E.L.; data curation, P.P. and I.E.L.; writing—original draft preparation, P.P. and I.E.L.; writing—review and editing, P.P. and I.E.L.; visualization, P.P. and I.E.L.; supervision, P.P. and I.E.L.; project administration, P.P. and I.E.L. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: The Guest Editors wish to express their appreciation and deep gratitude to all authors and reviewers who contributed to this Special Issue.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pintelas, P.; Livieris, I.E. Special Issue "Ensemble Algorithms and Their Applications". Available online: https://www.mdpi.com/journal/algorithms/special_issues/Ensemble_Algorithms (accessed on 24 December 2018).
2. Wang, S.; Chen, Q. The Study of Multiple Classes Boosting Classification Method Based on Local Similarity. *Algorithms* **2021**, *14*, 37. [CrossRef]
3. Kallipolitis, A.; Revelos, K.; Maglogiannis, I. Ensembling EfficientNets for the Classification and Interpretation of Histopathology Images. *Algorithms* **2021**, *14*, 278. [CrossRef]
4. Bologna, G. A Rule Extraction Technique Applied to Ensembles of Neural Networks, Random Forests, and Gradient-Boosted Trees. *Algorithms* **2021**, *14*, 339. [CrossRef]
5. Arrieta, A.B.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; García, S.; Gil-López, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [CrossRef]
6. Soman, G.; Vivek, M.; Judy, M.; Papageorgiou, E.; Gerogiannis, V.C. Precision-Based Weighted Blending Distributed Ensemble Model for Emotion Classification. *Algorithms* **2022**, *15*, 55. [CrossRef]
7. Meng, X.; Bradley, J.; Yavuz, B.; Sparks, E.; Venkataraman, S.; Liu, D.; Freeman, J.; Tsai, D.; Amde, M.; Owen, S.; et al. Mllib: Machine learning in apache spark. *J. Mach. Learn. Res.* **2016**, *17*, 1235–1241.
8. Moustakidis, S.; Siouras, A.; Vassis, K.; Misiris, I.; Papageorgiou, E.; Tsaopoulos, D. Prediction of Injuries in CrossFit Training: A Machine Learning Perspective. *Algorithms* **2022**, *15*, 77. [CrossRef]
9. Sagadevan, S.; Malim, N.H.A.H.; Husin, M.H. A Seed-Guided Latent Dirichlet Allocation Approach to Predict the Personality of Online Users Using the PEN Model. *Algorithms* **2022**, *15*, 87. [CrossRef]
10. Zou, Y.; Gao, C. Extreme Learning Machine Enhanced Gradient Boosting for Credit Scoring. *Algorithms* **2022**, *15*, 149. [CrossRef]
11. Utkin, L.V.; Konstantinov, A.V. Ensembles of random SHAPs. *Algorithms* **2022**, *15*, 431. [CrossRef]