

Editorial

Editorial Topical Collection: “Explainable and Augmented Machine Learning for Biosignals and Biomedical Images”

Cosimo Ieracitano ^{1,*}, Mufti Mahmud ^{2,3,4}, Maryam Dobarjeh ⁵ and Aimé Lay-Ekuakille ⁶

¹ DICEAM Department, University Mediterranea of Reggio Calabria, Via Zehender, Feo di Vito, 89122 Reggio Calabria, Italy

² Department of Computer Science, Nottingham Trent University, Nottingham NG11 8NS, UK; mufti.mahmud@ntu.ac.uk

³ Computing and Informatics Research Centre, Nottingham Trent University, Nottingham NG11 8NS, UK

⁴ Medical Technologies Innovation Facility, Nottingham Trent University, Nottingham NG11 8NS, UK

⁵ Computer Science and Software Engineering, Auckland University of Technology, Auckland 1010, New Zealand; maryam.gholami.dobarjeh@aut.ac.nz

⁶ Department of Innovation Engineering, University of Salento, 73100 Lecce, Italy; aime.lay.ekuakille@unisalento.it

* Correspondence: cosimo.ieracitano@unirc.it

Machine learning (ML) is a well-known subfield of artificial intelligence (AI) that aims at developing algorithms and statistical models able to empower computer systems to automatically adapt to a specific task through experience or learning from data [1]. ML techniques have been demonstrating remarkable breakthroughs in the field of biomedical research, especially in predictive analytics and classification tasks [2,3].

However, the success of ML in this domain has been accompanied by a challenge, i.e., the inherent opaqueness of ML algorithms [4]. Indeed, despite their efficacy, ML algorithms lack transparency in their decision-making processes and are often seen as black boxes. This lack of transparency raises concerns, especially in critical domains, where understanding the rationale behind machine decisions is important for fostering trust in decision-making [5].

In this regard, the emergence of explainable artificial intelligence (xAI) techniques has become a pivotal focus within this field. In particular, xAI methods strive to unveil the internal mechanisms of the AI algorithms, aiming to shed light on the outcomes, predictions, decisions, and recommendations generated by such models [6,7]. The primary objective is to enhance the interpretability and transparency of machine decisions. This is of paramount importance in medical applications, where such enhanced comprehension could have a significant impact on clinicians’ final decision-making [8]. In this context, several xAI-based approaches have been emerging in clinical applications, for example, rehabilitation systems based on brain–computer interfaces [9], detection of neurological disorders [10], breast cancers [11,12], and medical imaging analysis [13].

Furthermore, the escalating availability of medical and clinical data, collected from an expanding network of interconnected biosensors within the Internet of Things (IoT) framework, provides a rich source for training and refining ML models [14,15]. In addition, recent advances in augmented techniques, i.e., generative adversarial networks (GANs), have enhanced the decision-making capabilities of ML algorithms. Indeed, generative models are able to produce synthetic samples, augmenting the training data, potentially addressing issues related to data scarcity, and improving the generalization of models [16,17].

In this context, this topical collection includes ten papers focused on the latest advancements in the field of explainable and augmented ML applied to biosignals and biomedical images. Each of the ten original contributions accepted for publication has undergone a rigorous review process by a minimum of two expert reviewers across at least two rounds



Citation: Ieracitano, C.; Mahmud, M.; Dobarjeh, M.; Lay-Ekuakille, A.

Editorial Topical Collection:

“Explainable and Augmented Machine Learning for Biosignals and Biomedical Images”. *Sensors* **2023**, *23*, 9722. <https://doi.org/10.3390/s23249722>

Received: 15 November 2023

Accepted: 28 November 2023

Published: 9 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

of revision. These studies published in the current topical collection are briefly summarized as follows:

In contribution 1, the authors developed a brain-inspired neural network to explore the effect of mindfulness training on the electroencephalographic (EEG) function. In particular, a spiking neural network (SNN) was employed to assess the neural patterns generated over both spatial and temporal features derived from EEG data, which captured the neural dynamics linked to event-related potentials (ERPs). Furthermore, the interpretability of the SNN model was also further investigated. Outcomes indicated that SNN models provide valuable insights in distinguishing between different brain states in response to specific tasks and stimuli, as well as tracking changes in brain states through psychological interventions.

In contribution 2, a novel explainable analysis of potential biomarkers denoting tumorigenesis in non-small cell lung cancer is proposed based on detailed mathematical formulation for mRNA, ncRNA, and mRNA–ncRNA regulators. Specifically, the authors developed a system involving coupled-reaction partial differential equations to model temporal gene expression profiles within a two-dimensional spatial domain, capturing the transition states before converging to the stationary state. Experimental results demonstrate that the mathematical gene-expression profile provides the most accurate fit for the population abundance of these oncogenes.

In contribution 3, Vargas-Lopez et al. introduced an explainable machine learning approach that employed statistical indexes and support vector machines (SVMs) to detect stress in automobile drivers based on electromyographic (EMG) signals. The authors investigated the efficacy of seventeen statistical time features and, based on the analysis of the results, concluded that combining variance and standard deviation with a support vector machine classifier utilizing a cubic kernel is an effective approach for detecting stress events, achieving an AUC of 0.9.

In contribution 4, the authors conducted an extensive analysis of the most effective methods for classifying the emotion of fear, encompassing a range of machine learning methods such as decision trees, k-Nearest Neighbors, support vector machines, and artificial networks. In addition, xAI was also explored by means of Local Interpretable Model-Agnostic Explanations in order to interpret and justify predictions in a human-understandable manner. Experimental results showed classification performance, achieving accuracy from 91.7% using to 93.5% using dimensionality reduction and SVM.

In contribution 5, Dobarjeh et al. introduced an innovative methodology aimed at enhancing the interpretability of a brain-inspired SNN for deep learning and knowledge extraction. Their methodology focused on the learning process from real-time spatiotemporal brain data in an incremental and online operational mode. The experimental results show that by selecting a specific group of EEG features, the accuracy of EEG classification could be enhanced to 92%, outperforming all-feature-based classification.

In contribution 6, a novel approach for assessing the degree of gait impairment in Parkinson's disease using a computer vision-based approach was proposed. In addition, the interpretability of the feature values could be used by clinicians to support their decision-making and provide insight into the model's objective UPDRS rating estimation.

In contribution 7, the authors explored several xAI techniques such as GradCAM, LIME, RISE, Squaregrid, and direct gradient approaches with the ultimate aim of further explaining COVID-19 CT-Scan classifiers. Experimental results reported that VGG16 was the most affected by biases related to misleading artifacts, whereas DenseNet was more robust against them. In addition, it was observed that even slight differences in validation accuracies could lead to significant alterations in the explanation heatmaps for DenseNet architectures.

In contribution 8, Usama et al. developed AI-based classifiers for classifying single-trial error-related potentials (ErrPs) produced by twenty-five subjects with stroke. Specifically, EEG recordings were partitioned into epochs (ErrPs and NonErrPs) and classified by means of multi-layer perceptron based on temporal features or the entire epoch. Moreover, feature

classification was also conducted using shrinkage LDA. The authors concluded that by employing physiological brain potentials (ErrP and NonErrP) as input to the classifiers, it may be possible to interpret the classifier outputs in the context of established physiological research within this domain.

In contribution 9, Seven et al. proposed a novel pipeline for xAI imaging based on radiomic features and Shapley values for explaining predictions achieved by complex models. In particular, the authors conducted a retrospective analysis of data from glioma patients and presented an explainable prediction model for identifying isocitrate dehydrogenase mutations using radiomics data. Such a model could serve as a valuable tool in clinical decision-making.

In contribution 10, the authors developed an interpretable diabetes detection system using an xAI. To this end, the Pima Indian diabetes dataset was employed, and six ML algorithms were implemented along with an ensemble classifier to diagnose the diabetes disease. Global and local explanations were performed by means of the Shapley additive explanations (SHAP). The results reported accuracy of 90% and an F1 score of 89% using a five-fold cross-validation.

In summary, this topical collection has tackled numerous significant challenges in xAI and has presented innovative computational methods with potential deployment in clinical contexts. We would like to express our deepest gratitude to *Sensors* journal's Managing Team for their continuous support throughout the preparation of this collection. We greatly thank all the contributing authors and the anonymous expert reviewers whose invaluable efforts helped to select the submissions with the utmost quality.

Author Contributions: Conceptualization and supervision C.I., writing—original draft preparation, C.I., M.M., M.D. and A.L.-E.; writing—review and editing, C.I., M.M., M.D. and A.L.-E. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by Programma Operativo Nazionale (PON) “Ricerca e Innovazione” 2014–2020 CCI2014IT16M2OP005 (CUP C35F21001220009 code: I05).

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

List of Contributions:

1. Doborjeh, Z.; Doborjeh, M.; Crook-Rumsey, M.; Taylor, T.; Wang, G.Y.; Moreau, D.; Krägeloh, C.; Wrapson, W.; Siegert, R.J.; Kasabov, N.; et al. Interpretability of Spatiotemporal Dynamics of the Brain Processes Followed by Mindfulness Intervention in a Brain-Inspired Spiking Neural Network Architecture. *Sensors* **2020**, *20*, 7354.
2. Farouq, M.W.; Boulila, W.; Hussain, Z.; Rashid, A.; Shah, M.; Hussain, S.; Ng, N.; Ng, D.; Hanif, H.; Shaikh, M.G.; et al. A Novel Coupled Reaction-Diffusion System for Explainable Gene Expression Profiling. *Sensors* **2021**, *21*, 2190.
3. Vargas-Lopez, O.; Perez-Ramirez, C.A.; Valtierra-Rodriguez, M.; Yanez-Borjas, J.J.; Amezcua-Sanchez, J.P. An Explainable Machine Learning Approach Based on Statistical Indexes and SVM for Stress Detection in Automobile Drivers Using Electromyographic Signals. *Sensors* **2021**, *21*, 3155.
4. Petrescu, L.; Petrescu, C.; Oprea, A.; Mitruț, O.; Moise, G.; Moldoveanu, A.; Moldoveanu, F. Machine Learning Methods for Fear Classification Based on Physiological Features. *Sensors* **2021**, *21*, 4519.
5. Doborjeh, M.; Doborjeh, Z.; Kasabov, N.; Barati, M.; Wang, G.Y. Deep Learning of Explainable EEG Patterns as Dynamic Spatiotemporal Clusters and Rules in a Brain-Inspired Spiking Neural Network. *Sensors* **2021**, *21*, 4900.
6. Ruppel, S.; Morinan, G.; Peng, Y.; Foltyniec, T.; Sibley, K.; Weil, R.S.; Leyland, L.-A.; Baig, F.; Morgante, F.; Gilron, R.; et al. A Clinically Interpretable Computer-Vision Based Method for Quantifying Gait in Parkinson's Disease. *Sensors* **2021**, *21*, 5437.
7. Palatnik de Sousa, I.; Vellasco, M.M.B.R.; Costa da Silva, E. Explainable Artificial Intelligence for Bias Detection in COVID CT-Scan Classifiers. *Sensors* **2021**, *21*, 5657.
8. Usama, N.; Niazi, I.K.; Dremstrup, K.; Jochumsen, M. Detection of error-related potentials in stroke patients from EEG using an artificial neural network. *Sensors* **2021**, *21*, 6274.

9. Severn, C.; Suresh, K.; Görg, C.; Choi, Y.S.; Jain, R.; Ghosh, D. A Pipeline for the implementation and visualization of explainable machine learning for medical imaging using radiomics features. *Sensors* **2022**, *22*, 5205.
10. Kibria, H.B.; Nahiduzzaman, M.; Goni, M.O.F.; Ahsan, M.; Haider, J. An ensemble approach for the prediction of diabetes mellitus using a soft voting classifier with an explainable AI. *Sensors* **2022**, *22*, 7268.

References

1. Mohri, M.; Rostamizadeh, A.; Talwalkar, A. *Foundations of Machine Learning*; MIT Press: Cambridge, MA, USA, 2018.
2. Rajkomar, A.; Dean, J.; Kohane, I. Machine learning in medicine. *N. Engl. J. Med.* **2019**, *380*, 1347–1358. [[CrossRef](#)] [[PubMed](#)]
3. Ravi, D.; Wong, C.; Deligianni, F.; Berthelot, M.; Andreu-Perez, J.; Lo, B.; Yang, G.Z. Deep learning for health informatics. *IEEE J. Biomed. Health Inform.* **2016**, *21*, 4–21. [[CrossRef](#)] [[PubMed](#)]
4. Burrell, J. How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data Soc.* **2016**, *3*, 2053951715622512. [[CrossRef](#)]
5. Rasheed, K.; Qayyum, A.; Ghaly, M.; Al-Fuqaha, A.; Razi, A.; Qadir, J. Explainable, trustworthy, and ethical machine learning for healthcare: A survey. *Comput. Biol. Med.* **2022**, *149*, 106043. [[CrossRef](#)] [[PubMed](#)]
6. 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](#)]
7. Vilone, G.; Longo, L. Notions of explainability and evaluation approaches for explainable artificial intelligence. *Inf. Fusion* **2021**, *76*, 89–106. [[CrossRef](#)]
8. Tjoa, E.; Guan, C. A survey on explainable artificial intelligence (xai): Toward medical xai. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 4793–4813. [[CrossRef](#)] [[PubMed](#)]
9. Kim, S.; Choo, S.; Park, D.; Park, H.; Nam, C.S.; Jung, J.Y.; Lee, S. Designing an XAI interface for BCI experts: A contextual design for pragmatic explanation interface based on domain knowledge in a specific context. *Int. J. Hum.-Comput. Stud.* **2023**, *174*, 103009. [[CrossRef](#)]
10. Morabito, F.C.; Ieracitano, C.; Mammone, N. An explainable Artificial Intelligence approach to study MCI to AD conversion via HD-EEG processing. *Clin. EEG Neurosci.* **2023**, *54*, 51–60. [[CrossRef](#)] [[PubMed](#)]
11. Gulum, M.A.; Trombley, C.M.; Kantardzic, M. A review of explainable deep learning cancer detection models in medical imaging. *Appl. Sci.* **2021**, *11*, 4573. [[CrossRef](#)]
12. Lamy, J.B.; Sekar, B.; Guezennec, G.; Bouaud, J.; Séroussi, B. Explainable artificial intelligence for breast cancer: A visual case-based reasoning approach. *Artif. Intell. Med.* **2019**, *94*, 42–53. [[CrossRef](#)]
13. Chen, S.; Ren, S.; Wang, G.; Huang, M.; Xue, C. Interpretable CNN-Multilevel Attention Transformer for Rapid Recognition of Pneumonia from Chest X-Ray Images. *IEEE J. Biomed. Health Inf.* **2023**. [[CrossRef](#)] [[PubMed](#)]
14. Yu, Y.; Li, M.; Liu, L.; Li, Y.; Wang, J. Clinical big data and deep learning: Applications, challenges, and future outlooks. *Big Data Min. Anal.* **2019**, *2*, 288–305. [[CrossRef](#)]
15. Obermeyer, Z.; Emanuel, E.J. Predicting the future—Big data, machine learning, and clinical medicine. *N. Engl. J. Med.* **2016**, *375*, 1216. [[CrossRef](#)] [[PubMed](#)]
16. Lan, L.; You, L.; Zhang, Z.; Fan, Z.; Zhao, W.; Zeng, N.; Chen, Y.; Zhou, X. Generative adversarial networks and its applications in biomedical informatics. *Front. Public Health* **2020**, *8*, 164. [[CrossRef](#)] [[PubMed](#)]
17. Yi, X.; Walia, E.; Babyn, P. Generative adversarial network in medical imaging: A review. *Med. Image Anal.* **2019**, *58*, 101552. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.