Guest Editorial

Jose Maria Alonso-Moral
University of Santiago de Compostela, SPAIN

Corrado Mencar
University of Bari Aldo Moro, ITALY

Hisao Ishibuchi
Southern University of Science and Technology, CHINA

Explainable and Trustworthy Artificial Intelligence

In the era of the Internet of Things and Big Data, data scientists are required to extract valuable knowledge from the given data. They first analyze, cure and pre-process data. Then, they apply Artificial Intelligence (AI) techniques to automatically extract knowledge from data. Actually, AI is identified as a strategic technology and it is already part of our everyday life. The European Commission states that "EU must therefore ensure that AI is developed and applied in an appropriate framework which promotes innovation and respects the Union's values and fundamental rights as well as ethical principles such as accountability and transparency". It emphasizes the importance of eXplainable AI (XAI in short), in order to develop an AI coherent with European values: "to further strengthen trust, people also need to understand how the technology works, hence the importance of research into the explainability of AI systems". Moreover, in addition to the European General Data Protection Regulation (GDPR), a new European regulation on AI is in progress and it remarks once again the need to push for a human-centric responsible, explainable and trustworthy AI that empowers citizens to make more informed and thus better decisions. In addition, as remarked in the XAI challenge stated by the US Defense Advanced Research Projects Agency (DARPA), "even though current AI systems offer many benefits in many applications, their effectiveness is limited by a lack of explanation ability when interacting with humans". Accordingly, humankind requires a new generation of XAI systems. They are expected to naturally interact with humans, thus providing comprehensible explanations of decisions automatically made.

XAI is an endeavor to evolve AI methodologies and technology by focusing on the development of intelligent agents capable of generating decisions that a human could understand in a given context, and explicitly explaining such decisions. This way, it is possible to scrutinize the underlying intelligent models and verify if automated decisions are made on the basis of accepted rules and principles, so that decisions can be trusted and their impact justified.

This Special Issue is supported by the IEEE CIS Task Force on Explainable Fuzzy Systems, with the aim of providing readers with a holistic view of fundamentals and current research trends in the XAI field, paying special attention to fuzzy-grounded knowledge representation and reasoning but also regarding ways to enhance human-machine interaction through multi-modal (e.g., graphical or textual modalities) effective explanations. Therefore, the scope of this special issue is not limited to the community of researchers in Fuzzy Logic but it is open to contributions by researchers, from both academy and industry, working in the multidisciplinary field of XAI. As a result, this special issue attracted 31 submissions which reported state-of-the-art contributions on the latest research and development, up-to-date issues, challenges, and applications in the field of XAI. Following a rigorous peer review process, five papers are finally accepted for publication.

The first paper, entitled "Towards Understanding Human Functional Brain Development with Explainable Artificial Intelligence: Challenges and Perspectives", by M. Kiani et al., depicts a big picture of non-explainable and explainable AI methods for tackling challenging problems in Cognitive Neuroscience, with particular emphasis on Developmental Cognitive Neuroscience (DCN). AI methods represent valid tools to manage the complexity of cognitive phenomena and can be profitably used to detect possible cognitive developmental delays and abnormalities; yet, current DCN research has not benefited much from cutting-edge methodologies (such as deep neural networks) because of the lack of insight offered from the designed models. On the other hand, the adoption of XAI-based methodologies enables the extraction of meaningful developmental patterns and sheds light for future medical applications. Interestingly, the paper highlights some limitations of current XAI methodologies, which are inspiring in paving the way to further research.

The second paper, entitled "PIP: Pictorial Interpretable Prototype Learning for Time Series Classification" by A. Ghods and D. J. Cook, is focused on time-series

Digital Object Identifier 10.1109/MCI.2021.3129953
Date of current version: 12 January 2022

classification, for which XAI methods are still not widespread. The problem tackled by the authors is to make time-series classes easier to discern by users. To this end, the authors propose to learn pictorial class prototypes together with the class of time-series, so as to explain classes by using visual clues. They accomplish this task with a combination of Convolutional Neural Networks and a Multilayer Perceptron that is used for encoding time-series, generating visual prototypes and making final classification. The experiments were conducted by involving end-users, divided into three groups (STEM, clinicians and others), which were employed to evaluate perceived interpretability, accuracy, response time, and trust.

The third paper, entitled "recoXplainer: A Library for Development and Offline Evaluation of Explainable Recommender Systems" by L. Coba et al., goes into a different direction and tries to fill a gap in explainable recommender systems (XRS) by proposing a library with the aim of boosting research reproducibility and software interoperability in this emerging field. The library enables the development of XRSs through a pipeline (pre-process, train, recommend, explain and evaluate) and includes several solutions for building explanations, including model-based and post-hoc approaches. The library (which is available as open source, under the MIT license) has been applied to develop XRSs for publicly available benchmark datasets. It shows the different performance of the implemented methods, in terms of explainability as measured by the available metrics.

The fourth paper is entitled "Collective eXplainable AI: Explaining Cooperative Strategies and Agent Contribution in Multiagent Reinforcement Learning with Shapley Values" by A. Heuillet et al. It tackles the issue of explainability in Reinforcement Learning (RL), a field that is very active in Machine Learning research, but received little attention in XAI so far. Nevertheless, the introduction of XAI in RL could provide important insights in many scientific fields, including Economy and Social Sciences, as well as in industrial applications. The authors focus on explainability in Cooperative RL, where agents must cooperate in order to achieve a common goal. Explainability is introduced by approximating Shapley values that evaluate the contribution of each agent in achieving the common goal.

The Special Issue is closed by the fifth paper, entitled "Bridging the Gap Between AI and Explainability in the GDPR: Towards Trustworthiness-by-Design in Automated Decision-Making" by R. Hamon et al. This work is of interdisciplinary nature as it studies the current state of the art of XAI solutions and the legal requirements for explainability, with special emphasis on the GDPR which sets out the fundamental rights of individuals when their personal data are subject to processing (e.g., in profiling and automated decision-making). The authors put forward a number of challenges that limit current AI approaches in fulfilling the legal requirements set by GDPR (namely, complexity, accuracy vs. explainability, and correlation vs. causation) and describe a number of proposals that shed light on future research on XAI towards the development of an AI that is not only technically effective but also socially and legally acceptable.

The guest editors of this special issue would like to thank Prof. Chuan-Kang Ting, the Editor-in-Chief of IEEE Computational Intelligence Magazine, for his great support in initiating and developing this special issue together. Many thanks to all members of the editorial team for their kind support during the editing process of this special issue. Last but not least, we would also like to thank the authors for submitting their valuable research outcomes as well as the reviewers who critically evaluated the papers. We sincerely hope and expect that readers will find this special issue useful.