



A Comprehensive Survey of Explainable Artificial Intelligence (XAI) Methods: Exploring Transparency and Interpretability

Ambreen Hanif^{1(✉)}, Amin Beheshti^{1(✉)}, Boualem Benatallah², Xuyun Zhang¹, Habiba¹, EuJin Foo¹, Nasrin Shabani¹, and Maryam Shahabikargar¹

¹ Macquarie University, Sydney, Australia
{ambreen.hanif,eujin.foo,nasrin.shabani,
maryam.shahabi-kargar}@hdr.mq.edu.au,
{amin.beheshti,xuyun.zhang}@mq.edu.au, habibah@gmail.com
² Dublin City University, Dublin, Ireland
boualem.benatallah@adaptcentre.ie

Abstract. Artificial Intelligence (AI) is undergoing a significant transformation. In recent years, the deployment of AI models, from Analytical to Cognitive and Generative AI, has become imminent; however, the widespread utilization of these models has prompted questions and concerns within the research and business communities regarding their transparency and interpretability. A primary challenge lies in comprehending the underlying reasoning mechanisms employed by AI-enabled systems. The absence of transparency and interpretability into the decision-making process of these systems indicates a deficiency that can have severe consequences, e.g., in domains such as medical diagnosis and financial decision-making, where valuable resources are at stake. This survey explores Explainable AI (XAI) techniques within the AI system pipeline based on existing literature. It covers tools and applications across various domains, assessing current methods and addressing challenges and opportunities, particularly in the context of Generative AI.

Keywords: Explainable Artificial Intelligence · Artificial Intelligence · Transparency · Interpretability

1 Introduction

The rapid development of Artificial Intelligence (AI) and the supporting domain including Big Data and high-performing computational infrastructure has triggered a tectonic shift. The substantial refinement of deep learning-based systems including foundation models(e.g. Transformer, GPT-4, Bard, DALL-E, RoBERTa, etc.) has equipped the AI-based systems to penetrate high-stake applications. These applications span across various critical domains, including

healthcare, finance, law enforcement, and agriculture [35,39]. This speedy penetration has sociotechnical, privacy, and safety implications associated with these intelligent systems and is categorized as an existential threat to the human race. One of the key contributions to the scepticism is the non-transparent nature of these models. The opaqueness of the ML algorithms restricts trust and resists deployment in vulnerable domains.

AI is defined as a set of approaches to mimic human behaviour in general whereas by and large, Machine Learning (ML) algorithms are predictive models using existing data features to build class mapping during the learning phase. This learning phase is based on the data retrieved from the usual user activities (e.g. online shopping, medical history, social interactions, customer profiles, etc.). This enormous amount of data is liable to contain human biases and predispositions. So the decision models also have the innate ability to have presumptions on the learning which can lead to wrong decisions. As black box models are extensively developed and tested on huge datasets, numerous stakeholders emphasize system transparency [23,24]. In general, people are restrained from using techniques that are not justifiable and transparent, which results in demanding ethical AI [38]. The increasing complexity of these opaque systems enhances performance, yet they still lack transparency. While designing and developing the ML model, keeping transparency as a design consideration drive impartial decision-making and can guarantee to use of important variables to generate the model predictions.

As a consequence, Explainable Artificial Intelligence (XAI), an emerging frontier of AI, is pertinent due to its ability to help answer the raised concerns and mitigate the associated risks. XAI gives a suite of ML techniques to generate an explainable model and develop a trustworthy human-understandable system. Various communities are using explanations for Model decisions. The objectives and perspectives of the developed various XAI system vary as per the need for explanations. The contribution of this paper is summarized below. This paper begins by providing an overview of the key requirements in the field and subsequently conducts a comprehensive review of explainable artificial intelligence (XAI) approaches for the machine learning (ML) pipeline. The review specifically focuses on different stages of the ML lifecycle to analyze and evaluate the effectiveness of XAI techniques. This is an extension of our previous survey on XAI techniques [16].

The paper is organized as follows. In Sect. 2, we discuss the terminology of the domain. Following that, Sect. 3 contributes to providing the overview of the approaches in XAI, and the evaluation and discussion on these approaches are presented in Sect. 4. Lastly, the conclusion is highlighted in Sect. 5.

2 Desiderata for Explainable Artificial Intelligence

As the domain is emerging and we don't have enough context, some research works [6,21], are using the terms interpretability and explainability interchangeably, preventing the creation of common grounds. There is communal agreement

Table 1. This table elucidates essential terminology within the XAI literature, serving as a reference for clarifying the concepts explored in this study.

Terms	Definition
Domain	
Artificial Intelligence	It refers to a branch of computer sciences focusing on developing the machines capable of mimicking human intelligence.
Machine Learning	It is the subset of AI techniques focused on learning from Experience with respect to some classes of tasks and performance measures.
Explainable Artificial Intelligence	The capacity of the system to explain the results generated by the AI system [15]
Interpretable Machine Learning	The ability of the models to be fundamentally human understandable due to simpler nature [21].
Explainability	It refers to the capacity of the AI system to cater to the specific requirements of the audience in the explanations [28].
Explanation	This term is defined as the outcome of the model to explain or clarify its internal functionalities [5,20].
Interpretability	It defines as a tendency of the system to produce the human-intelligible understanding of the model decision making [21].
Ethical Terms	
Fairness (unbiased)	It refers to the ethical principle to treat individuals and groups equitable [27].
Model Evaluation Measures	
Transparency	The degree to which a model can explain the route to reach the decision [8].
Robustness	The capacity of the model to be persistent in the model decision with the minor input perturbations [2].
Trustworthy	It defines the confidence in a system’s performance when facing a given problem [35]
Completeness	The system is complete when the system behavior can be anticipated in more situations [13].

on the necessity for a formal definition of the nomenclature [1, 14]. In this section, we will outline the definitions that we adopt to comprehend the techniques within the domain. This understanding will allow us to assess the capabilities of the explanation system and evaluate its alignment with the responsible AI framework.

Explainability has varying definitions; for this study, we refer to the definition: “Explainability aims to bridge the gap between the algorithm and human interpretation of the decision-making process. It is capable of enhancing trust by answering the how and why of the system” [28].

Interpretability from a user-centric view, is the human-intelligible explanation of the model system output [7]. For different users understanding of the system varies. A point of caution is reliance on human evaluations can lead to persuasive systems rather than transparent systems. This limits the ability to define the appropriate scope of interpretation. The selected terminology and the definitions are listed in Table 1.

3 Explainable Artificial Intelligence Methods

These days ML tends deep learning systems are complex and comprise many more layers and huge training data and have achieved high accuracy [28]. Despite these successes, the associated risks are a reluctance to adopt these models. The challenge of the hour is to develop trustworthy systems with these complex systems with escalating performance to help the user to understand the why of the decision to mitigate the associated risks. Now the next challenge is building trustworthy systems. So now with this development, we have achieved high accuracy. There are numerous surveys on XAI [1, 5, 7, 13, 14, 21] and explainable deep learning [25, 41]. All of these surveys cover a large body of work in different dimensions. A standard ML pipeline consists of several phases as shown in Fig. 1. In this paper, We will analyze each phase and will discuss the XAI approaches in the literature to address the various problems of each stage. As various stakeholders are involved in the different stages of the pipeline. The appropriate explanation to answer the various questions associated with the different phases of the ML pipeline.

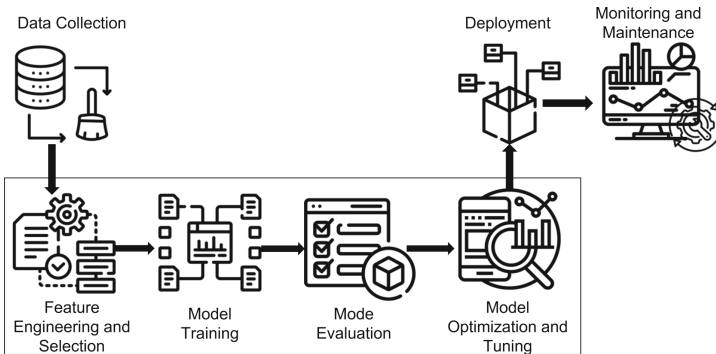


Fig. 1. A snapshot of a standard machine learning lifecycle.

3.1 Data Collection and Preparation

The foremost and most crucial stage of the ML lifecycle. It refers to the process of gathering and organizing data for utilization in machine-learning applications. This is a multi-step phase to ensure data relevance, reliability, and accuracy of the data using data cleaning, transformation, and integration techniques. The generation of explanations for data sources can address different aspects. We will classify these aspects into two categories: i) Detection of Data Bias, and iii) Annotation and Labeling of Data.

Detection of Data Bias. Counterfactual explanations often take the form of statements such as, “You were denied a loan due to an annual income of X. If your income had been X+Y, you would have been approved for the loan” [37].

The goal of a counterfactual statement is to identify the smallest modification in feature values that can yield the desired prediction.

Annotation and Labeling of Data. XAI can play a role in data annotation by providing transparency and interpretability to the annotation process. Forward Propagation-Based study [10] retrieved the feature importance with perturbation and identified the accountable mask for the results. Altering or blurring these salient features directly impacts the original classification outcome. These predictions can then be analyzed and interpreted using various XAI techniques to gain insights into the decision-making process of the model and understand the factors influencing the predictions.

3.2 Feature Engineering and Selection

Feature Interaction Analysis. Saliency Based Approaches highlight the importance of the region in the systems. These methods employ saliency maps to comprehend the contribution and significance of features in specific decisions. Visualization support aids in facilitating the comprehension of a diverse audience, allowing them to discern which feature influenced the decision. Prominent approaches proposed to calculate salience maps includes Integrated Gradients [34], SmoothGrad [32].

Outlier Detection. Automatic rule extraction Decompositional approaches work on the neurons to mimic the rules from the network architecture. Studies on transforming the neural network to the fuzzy rule are also available, the main work is on the extraction of approximation from the neurons [43]. Rule extraction techniques are valuable for identifying behavioural patterns, despite not being completely faithful to the models. As a result, there's a requirement for further research on explainability to address this limitation. Adversarial Examples will provide the interpretable model understanding. Most approaches suggest reducing the gap between the antagonistic example and the instance to be controlled while adjusting the prognosis to the desired outcome of the system. This method allows for diagnosing the outlier in the data [40].

3.3 Model Training

Propagation Based Approaches supports the identification of important regions. The output of the model is feedback to the system. The robustness of the system helps to understand the stability of the decision. *Backpropagation-based* methods take derivatives of the output w.r.t. input and the system gradients. These approaches are intrinsic as they are based on the important regions and rely on the model structure to understand the important region. DeepLift [31], an example approach assigns the positive and negative contributions to the features based on the actual and retrieved output difference.

3.4 Model Evaluation

One of the accepted convenient ways to explain networks is to develop proxy models that can approximate the behavior of these networks, such methods are referred to as 'model-agnostic' approaches. Ribeiro et al. have exemplified the proxy model by developing LIME(Local Interpretable Model-agnostic Explanations) [26]. Whereas SHAP(SHapely Additive exPlanations) method [18] supports interpretability by assigning feature scores to each attribute. The underlying mechanism of these models is to use the input to generate the linear proxy model to predict the behaviour by probing and perturbation. These models can be evaluated for their faithfulness to the original models. Decision Trees, another accepted type of proxy model, generate insights of Neural networks in equivalent Decision Trees [4]. The tree equivalence holds for fully connected, convolution, recurrent, and activation layers to satisfy the faithfulness. Although the generated trees are faithful these computations are computationally expensive and take substantial computational resources and time to develop.

Attention Mechanisms are neural networks that learn to assign weights to inputs or internal features, enabling them to focus on relevant information. These approaches have demonstrated incredible success in various complex tasks [36]. While attention units are not explicitly trained to generate human-readable explanations, they inherently provide a map of the information flow within the network, which can be interpreted as a form of explanation.

Knowledge Infused Explanations are general knowledge and knowledge-based methods. *General Knowledge* is generally referred to as a body of information or facts acquired through intellectual processes and diversity is the key trait. In this section, we will analyze the XAI techniques employing general knowledge for enrichment. Kim et al. [17] utilize the Concept Activation Vector(CAV) and analyze the importance of a concept in the task.

Knowledge-base (KB) Methods are deployed to enrich the model with human knowledge using available corpora explanations, which are suitable for specific situations. The KBs are generally represented as Knowledge Graphs(KG). The knowledge graph can be employed in the model design to enrich the feature entities and system rules to improve the model performance and explain the decisions. This is a known strategy for the recommender systems [19] to enrich the relation to identifying the similarity. A knowledge-based system can be employed after modeling to enhance reasoning, potentially through abductive reasoning, by utilizing its knowledge base to provide richer explanations [12, 29].

3.5 Model Optimization

Gradient-based explanation helps to understand the vector representation at the system and the intermediate layer level. These layer-based vector representations support the 'transfer learning' mechanism for the other problems to learn from the underlying vector representation. These are not the real explanation of the system, these are intermediate vectors for the system to understand and

resolve similar problem understanding and learning patterns. CAM [42], and GradCAM [30] are to name a few.

Explanation Generation Systems are designed to generate their own explanation units. The primary working principle is to generate the ‘because’ part with the model decision-making. These explanations are not directly interpretable and need proper evaluation to have faith in the model-generated answers. TEXTVQA [22], an extension of VQA a system that generates the multimodal explanation for the image captioning system can be visual as well as textual to help to understand why a certain caption is generated for the visual.

3.6 Deployment

The explanation without the training data can be categorized into human collaboration and policy abstraction-based binary categories.

Human Alliance: The studies [3, 9] proposed methods to automatically build the explanation corpus for the system agents to guide humans. The network learns to translate the action to the natural language generation. These beginning steps to experiential studies and exploring the machine learning pattern leads to a formal evaluation of the explanations to provide the information related to the events based on the experience gained throughout the processes.

Policy Abstraction highlights the policy information from the experience of the player. The generated summary is capable to enrich the context to understand the explanation of a specific action in the circumstances. A few relevant studies are [3, 33] the former study proposes the framework for abstraction and the latter support various abstraction levels for the same to be used in the system for the following action plan generation.

3.7 Monitoring and Maintenance

XAI can help detect and understand data drift by supplying explanations for model predictions, identifying changes in feature importance, and monitoring shifts in underlying patterns, enhancing the model’s adaptability. *incremental Permutation Feature Importance (iPFI)* on the interpretation of the complex feature is proposed by [11].

4 Discussion

The utilization of XAI techniques is crucial for algorithm transparency and interpreting model decisions, leading to improvements in the machine learning lifecycle. These techniques aid in understanding the decision-making process of models at different stages of the pipeline and enable achieving transparency objectives based on system requirements. We will summarize the approaches at the various stages of the ML pipeline in Table 2. Applying XAI techniques can indeed provide numerous benefits across various aspects of machine learning models.

Table 2. An overview of selective XAI approaches across various stages of the ML pipeline.

ML Pipeline	XAI Methods	Type of Explanation	Inter.	Trans.	Data Type			
					T	I	Te	G
Data Collection	Counterfactual	Instance-Based	o	o	✓	✓	✓	✓
	Forward Prop.	Prop. Based	o	x	✓	✓	✓	✓
Feature Engineering and Selection	Van. Grad	Saliency Based	o	o	✓	✓	✓	✓
	Integ. Grad.	Saliency Based	o	o	✓	✓	✓	✓
	SmoothGrad	Saliency Based	x	o		✓	✓	
	DeepRed	Autom. Rule Extr.	o	o	✓			
	DIFFI	Isolation Forest	o	o	✓			✓
	FuzzyRules	Autom. Rule Extr.	o	o	✓			
	Advers.Ex.	Insatnce Based	o	o	✓	✓	✓	✓
Model Training	LRP	Gradient Based	o	o	✓	✓	✓	✓
	DeepLift	Gradient Based	o	o	✓	✓	✓	
	Atten. Rollouts	Attention Based	o	x	✓		✓	
	SHAP	(Game Theory)	o	x	✓	✓	✓	✓
	LIME	MA Perturb. based	o	x	✓	✓	✓	
	Decision Trees	Intr. Rule Based	o	o	✓	✓	✓	
	Capture attention	Self-atten. Net.	o	o	✓			✓
Model Eval.	CAV/TCAV	Knowledge Based	o	x	✓	✓	✓	
Model Optimization and Tuning	Semantically corr.	Knowledge Based	o	x	✓	✓	✓	
	GradCAM	Gradient Based	o	o		✓		
	TEXTVQA	Multimodel Expla.	o	x			✓	
Deployment	Human Alliance	Experience Based	o	x	-	-	-	-
	Policy Abstraction	Experience Based	o	x	-	-	-	-
Monitoring	PFI	Feature Importance	o	o	✓		✓	

Through XAI, data quality can be evaluated, quantified, and remedied. Explanations aid in data selection, identifying valuable information for improvement, and enabling adjustments to ensure fairness and equity in the model’s performance. XAI techniques have the tendency to support data discretization and feature interaction analysis. The integration of XAI with data discretization and feature interaction analysis enhances our understanding and enhances the reliability of the machine learning models. Moreover, XAI techniques contribute to privacy protection. XAI techniques allow for the interpretation of models without compromising individual privacy. In conclusion, applications of XAI techniques bring several benefits to machine learning models. XAI techniques enhance transparency, fairness, and reliability. By leveraging XAI, organizations can make informed decisions, address ethical concerns, and build robust and trustworthy AI systems.

5 Conclusion and Future Directions

In this paper, a comprehensive and systematic review of the development of XAI approaches for the machine learning pipeline is presented. XAI posed several

challenges, from complex infrastructure to computational cost, but the strategic choices of the explanation techniques with the defined objective are beneficial and can mitigate the risks associated with the high-stake application. Deploying the right approach for the explanation of the model decisions can not only enrich the business processes but help to build faith in the system results. In these times of generative AI and the foundation models, systems are suffering from the inaccessibility to system understanding, and XAI can fill the gap in human understanding and model decision-making in high stake decisions. The explanation generation and evaluation framework for the machine learning pipeline can strengthen the downstream applications even derived from the foundation models. Appropriate XAI techniques with the relevant metrics for computational and cognitive evaluations of the model are a key step to proceed.

Acknowledgement. We acknowledge the Centre for Applied Artificial Intelligence at Macquarie University, Sydney, Australia, for funding this research.

References

1. Adadi, A., et al.: Peeking inside the black-box: a survey on Explainable Artificial Intelligence (XAI). *IEEE Access* **6**, 52138–52160 (2018)
2. Alvarez-Melis, D., et al.: On the Robustness of Interpretability Methods. arXiv preprint [arXiv:1806.08049](https://arxiv.org/abs/1806.08049) (2018)
3. Amir, O., et al.: Summarizing agent strategies. *Auton. Agent. Multi-Agent Syst.* **33**(5), 628–644 (2019)
4. Aytekin, C.: Neural Networks are Decision Trees. arXiv preprint [arXiv:2210.05189](https://arxiv.org/abs/2210.05189) (2022)
5. Arrieta, A.B., et al.: Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **58**, 82–115 (2020)
6. Cabitza, F., Campagner, A., Ciucci, D.: New frontiers in explainable AI: understanding the GI to interpret the GO. In: Holzinger, A., Kieseberg, P., Tjoa, A.M., Weippl, E. (eds.) *CD-MAKE 2019*. LNCS, vol. 11713, pp. 27–47. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-29726-8_3
7. Doshi-Velez, F., et al.: Towards A Rigorous Science of Interpretable Machine Learning. arXiv preprint [arXiv:1702.08608](https://arxiv.org/abs/1702.08608) (2017)
8. Došilović, F.K., Brčić, M., Hlupić, N.: Explainable artificial intelligence: a survey. In: *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, Croatia, 2018, pp. 0210–0215 (2018). <https://doi.org/10.23919/MIPRO.2018.8400040>
9. Ehsan, U., et al.: Automated rationale generation: a technique for explainable AI and its effects on human perceptions. In: *Proceedings of IUI*. ACM (2019)
10. Fong, R.C., et al.: Interpretable explanations of black boxes by meaningful perturbation. In: *Proceedings of the IEEE ICCV*, pp. 3429–3437 (2017)
11. Fumagalli, F., et al.: Incremental Permutation Feature Importance (iPFI): Towards Online Explanations on Data Streams. arXiv preprint [arXiv:2209.01939](https://arxiv.org/abs/2209.01939) (2022)
12. Gaur, M., et al.: Semantics of the black-box: can knowledge graphs help make deep learning systems more interpretable and explainable? *IEEE Internet Comput.* **25**(1), 51–59 (2021)

13. Gilpin, L.H., et al.: Explaining explanations: an overview of interpretability of machine learning. In: IEEE 5th International Conference on DSAA (2019)
14. Guidotti, R., et al.: A survey of methods for explaining black box models. *ACM Comput. Surv. (CSUR)* **51**(5), 1–42 (2018)
15. Gunning, D., et al.: XAI-Explainable Artificial Intelligence. *Sci. Robot.* **4**(37), 7120 (2019)
16. Hanif, A., et al.: A survey on explainable artificial intelligence techniques and challenges. In: IEEE 25th EDOCW, pp. 81–89. IEEE (2021)
17. Kim, B., et al.: Examples are not enough, learn to criticize! criticism for interpretability. In: Advances in NIPS, vol. 29 (2016)
18. Lundberg, S.M., et al.: A unified approach to interpreting model predictions. In: Advances in NIPS, Long Beach, CA, vol. 30 (2017)
19. Ma, W., et al.: Jointly learning explainable rules for recommendation with knowledge graph. In: Proceedings of the WWW, pp. 1210–1221 (2019)
20. Markus, A.F., et al.: The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. *JBHI* **113**, 103655 (2021)
21. Miller, T.: Explanation in artificial intelligence: insights from the social sciences. *Artif. Intell.* **267**, 1–38 (2019)
22. Rao, V.N., et al.: A first look: towards explainable TextVQA models via visual and textual explanations. In: Proceedings of the Third MAI-Workshop, pp. 19–29. ACL (2021)
23. Pouriyeh, S., et al.: A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease. In: IEEE ISCC, pp. 204–207 (2017)
24. Raju, C., et al.: A survey on predicting heart disease using data mining techniques. In: ICEDSS, pp. 253–255 (2018)
25. Ras, G., et al.: Explainable deep learning: a field guide for the uninitiated. *J. Artif. Intell. Res.* **73**, 329–397 (2022)
26. Ribeiro, M.T., et al.: “Why should I trust you?” Explaining the predictions of any classifier. In: Proceedings of the ACM SIGKDD, KDD 2016, pp. 1135–1144 (2016)
27. Romei, A., et al.: A multidisciplinary survey on discrimination analysis. *KER* **29**(5), 582–638 (2014)
28. Saeed, W., et al.: Explainable AI (XAI): a systematic meta-survey of current challenges and future opportunities. *KBS* **263**, 110273 (2023)
29. Sarker, I.H.: Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SNCS* **2**(6), 420 (2021)
30. Selvaraju, R.R., et al.: Grad-CAM: visual explanations from deep networks via gradient-based localization. *IJCV* **128**(2), 336–359 (2017)
31. Shrikumar, A., et al.: Learning important features through propagating activation differences. In: 34th ICML, vol. 7, pp. 4844–4866 (2017)
32. Smilkov, D., et al.: SmoothGrad: removing noise by adding noise. arXiv (2017)
33. Sridharan, M., et al.: Towards a theory of explanations for human-robot collaboration. *KI Künstliche Intell.* **33**(4), 331–342 (2019)
34. Sundararajan, M., et al.: Axiomatic attribution for deep networks. In: 34th ICML 2017, vol. 7, pp. 5109–5118 (2017)
35. Tjoa, E., et al.: A survey on explainable artificial intelligence (XAI): towards medical XAI. *IEEE Trans. Neural Netw. Learn.* **14**(8), 1–21 (2019)
36. Vaswani, A., et al.: Attention is all you need. In: Advances in NIPS (2017)
37. Wachter, S., et al.: Counterfactual explanations without opening the black box: automated decisions and the GDPR. *Harvard JOLT* **31**, 841 (2017)

38. Wang, Y.-X., et al.: Using data mining and machine learning techniques for system design space exploration and automatized optimization. In: ICASI, pp. 1079–1082 (2017)
39. Wells, L., et al.: Explainable AI and reinforcement learning—a systematic review of current approaches and trends. *Front. Artif.* **4**, 550030 (2021)
40. Yuan, X., et al.: Adversarial examples: attacks and defenses for deep learning. *IEEE Trans. Neural Netw. Learn.* **30**(9), 2805–2824 (2018)
41. Zhang, Z., et al.: Deep learning on graphs: a survey. *IEEE Trans. Knowl. Data Eng.* **34**(1), 249–270 (2022)
42. Zhou, B., et al.: Learning deep features for discriminative localization. In: *IEEE CVPR*, pp. 2921–2929 (2016)
43. Zilke, J.R., Loza Mencía, E., Janssen, F.: DeepRED – rule extraction from deep neural networks. In: Calders, T., Ceci, M., Malerba, D. (eds.) *DS 2016. LNCS (LNAI)*, vol. 9956, pp. 457–473. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46307-0_29