Core Problem and Challenges of Explainable Artificial Intelligence Model for Data Mining Purposes

Tianrun Qiu
Southern University of Science and Technology
Shenzhen, China
qiutr@mail.sustech.edu.cn

ACM Reference Format:
Tianrun Qiu. 2025. Core Problem and Challenges of Explainable Artificial Intelligence Model for Data Mining Purposes. In *Data Mining '25: SUSTech Data Mining Thesis 2025 Spring, June 03, 2025, Shenzhen, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Utilization of artificial intelligence (AI) models have been an important method for complex data mining tasks [28, 29, 36]. However, AI models are frequently described as "black boxes", which hinders the further application of models in high-stakes fields such as medicine judgment, financial risk assessment and judicial adjudication [10, 27, 51]. To resolve this issue, explainable AI (XAI) methods are proposed to provide crucial insights to make AI results more explainable and trustworthy [54].

However, it remains challenging towards an effective paradigm to design an AI algorithm framework that provides feature contribution explanations that are human-interpretable, without significantly sacrificing model prediction performance [1, 3, 54], which is one of the core scientific problem in the data mining field.

To fulfill this vision, researchers concluded four main hurdles to overcome.

- **Complex Data and Models**: Complex dataset often consists of high-dimensional and non-linear data, which are typically unsolvable using more traditional and simplistic methods, making it harder to explain model decisions [3]. In domains with large feature sets, such as genomic data in medical diagnosis, this complexity is more significant, where thousands of variables may influence the outcome [47].
- **Variability in User Cognition Levels**: The ability and willingness for understanding explained results vary significantly for domain experts (e.g. doctors) and normal users (e.g. patients) [36]. Therefore, flexibility and a balance between precision and intuitiveness remain a critical problem, which also makes it more challenging to provide a universal evaluation standard for XAI researches [40, 48] and further develop a standardized method to evaluate XAI application automatically [24].
- **Computational Overhead and Sacrificed Performance**: In real-time applications (such as autonomous vehicles or real-time financial trading systems), the computational cost of generating explanations may introduce latency, thereby affecting system performance [40], which is more pronounced for larger datasets [20]. However, large datasets are extremely common in data mining circumstances.
- **Misleading Interpretation Possibility**: Explanations may not always align with the model's actual decision-making process, which can lead to incorrect conclusions. For example, traditional XAI methods like LIME suffer from unreliable sampling issues [17]. More recent NLP-based methods will further suffer from hallucination problem [26].

This paper aims to survey and summarize current cutting-edge advancements for resolving the four main challenges to provide a comprehensive sense towards the future of explainable data mining.

2 COMPLEX DATA AND MODELS

Simple data analysis can often be solved by utilizing inherently interpretable model structures, such as decision tres or generalized additive models (GAMs) [41]. However, data mining tasks, especially tasks within domains with large feature sets, often involve analysis of high-dimensional data and non-linear relationships, and are therefore more suitable to analyze using more complex models [3, 47]. Therefore, explainable AI (XAI) methods towards complex models is an important direction for XAI researches.

2.1 Hybrid Explainable Models

A possible solution is hybrid explainable modeling methods, which seek to mix an inherently interpretable modeling technique with a rather sophisticated black-box method [3, 14].

The Contextual Explanation Networks (CEN) [2] operates on a AI framework where inputs must be predicted within specific contexts. Its methodology involves contextual information encoded as probability information into the parameter space of an inherently interpretable model through a complex architecture. Subsequently, the processed data is fed into the CEN model to generate predictions with the explainable feature of simpler models, as shown in Fig. 1. Similarly, BagNets [8] are hybrid interpretable models that classify images by processing partitioned patches through deep neural networks (DNNs), then aggregating local evidence via SoftMax. Therefore, every patch's prediction is explainable to some extents, as discussed in Fig. 2.

However, the limitation is that a specific model must be utilized for explainability, which may lack enough community support. Thus, researchers also started to seek methods to make existing popular modeling techniques more interpretable.

Data Mining '25, June 03, 2025, Shenzhen, China

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

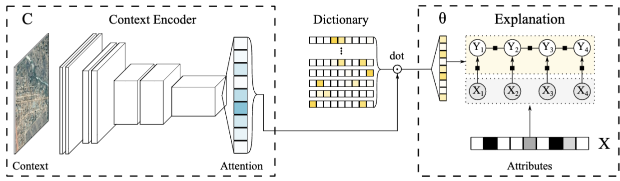

Figure 1: Structure of a CEN network

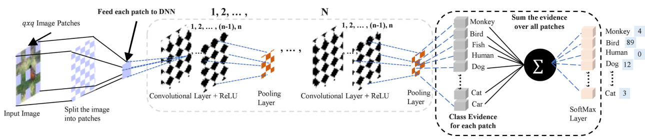

Figure 2: Decision paradigm of BagNets

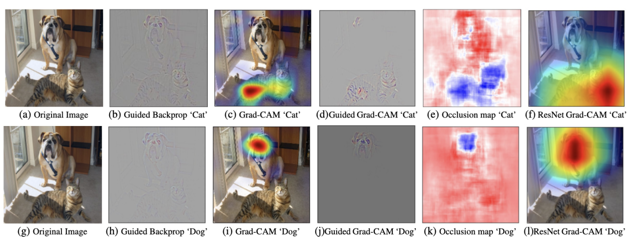

Figure 3: Explanation graph produced via GradCAM method

2.2 Model-specific Methods

To facilitate explainability of popular complex models, researchers must have a sufficient understanding of the operational mechanisms of popular models to enable demonstration of the model's operational logic to more users to provide explainability.

For instance, researchers have perceived the explainable potential of attention mechanism [5]. For attention-based models, an attention map can be created to display a set of weights or scores assigned to the input components, which is usable to determine the most relevant parts in the input that pertain to the successful execution of the specific task under consideration. However, it was also indicated that a diverse set of attention maps may yield identical predictions, which causes controversy over whether such mechanisms provide trustworthy explanations [25]. Similarly, DeepLIFT [46] explains neural network decisions by backpropagating feature contributions relative to reference activations, distinguishing positive or negative impacts, which is based on a profound understanding of the backpropagation theory. Gradient-weighted Class Activation Mapping (GradCAM) [45] provides visual explanations for CNN by highlighting influential regions in images through gradients from the final convolutional layer, further aiding in interpreting convolutional neural network decisions, as shown in Fig. 3. As for Large language model (LLM), Chain-of-Though (CoT) reasoning [53] can be utilized for users to take a glance at the result generation process of the model.

Meanwhile, researchers in visualization field or human-computer interaction (HCI) field have proposed methods that improve model interpretability via interactive exploratory system over neural network structures. Popular examples include CNN Explainer [52], which provides on-demand, dynamic visual explanation views that can promote users' sense of understanding towards CNN networks, as displayed in Fig. 4; and Interactive-Classification [9], which allows users to compare and contrast the image regions which AI models actually utilize for classification via interactive exploration.

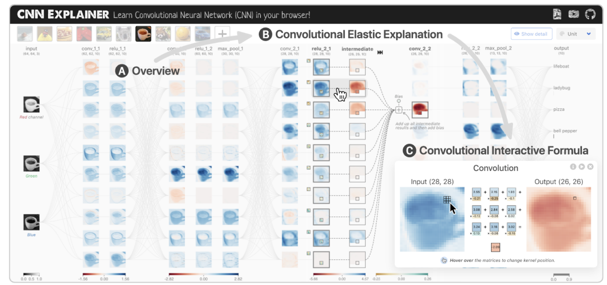

Figure 4: Interface of CNN Explainer

3 VARIABILITY IN USER COGNITION LEVELS

Significant variability exists in users' cognitive abilities and domain expertise, influencing their interpretation and acceptance of these explanations, and may further decrease the effectiveness of Explainable Artificial Intelligence (XAI) systems. For instance, domain experts, such as physicians, often require detailed, technical explanations to inform critical decision-making, whereas laypersons, such as patients, benefit more from simplified, intuitive summaries [36]. This is also proved by Hagras [21], who concluded that when providing easily understandable explanations in layperson-friendly language, the results generated by AI models are more likely to be considered credible by end users. This diversity necessitates flexible XAI systems that balance precision for accuracy with intuitiveness for accessibility, a challenge that is central to advancing XAI research and deployment.

3.1 Challenges in Universal Evaluation Standards

The variability in user cognition levels complicates the establishment of universal evaluation standards for XAI systems, and this absence of standardized methods for automatically evaluating XAI applications hinders the development of broadly applicable solutions [40, 48]. This issue is exacerbated by the need to tailor explanations to diverse user groups, making it difficult to devise one-size-fits-all evaluation metrics [24]. There are also methods that are directly aimed at professional usages [37]. The lack of consensus on evaluation criteria underscores the need for flexible, user-centric approaches to assess the effectiveness of XAI systems across different contexts.

3.2 Adaptive Multi-Level Explanation Systems

To accommodate diverse user needs, researchers have proposed adaptive multi-level explanation systems that dynamically adjust the depth and complexity of explanations based on the user's expertise level. For example, such systems might provide comprehensive technical details, including feature importance scores and model mechanics, for data scientists, while offering narrative, high-level explanations for non-experts, which also align with user-centered design principles.

In a notable case study, Salimiparsa et al. [42] applied a human-centered design approach to develop explanation design patterns for clinical decision support systems (CDSS). Their methodology included domain analysis to define the context of explanations, requirements elicitation to identify user needs, and multi-modal interaction design to create intuitive interfaces. The resulting design patterns were tailored to meet the needs of medical professionals, enhancing the understandability of AI-generated recommendations in clinical settings. Similarly, other researches for explainable AI diagnoses [22, 44] have been highlighting the role of user viewpoints and application orientation in fostering trust and understanding in XAI systems.

Multi-level explanation is also valuable in the field of finance, but has been less discussed about. Bhatia et al. [6] pointed that both multi-level explanation for different investors and the explanability of AI can increase the trust of investors to AI-based invest advisor robots, whereas Chen et al. [13] developed a holistic approach to improve the interpretability in financial lending for both bank officers and end users.

However, research on multi-level explanations in other XAI application domains still requires further exploration. A possible solution to resolve this challenge is to utilize powerful large language models (LLM), which is increasingly more capable of generating user-specific explanation based on the same XAI results to suit the needs for different groups of people [33].

3.3 Semantic-Level Explanation Approaches

Another promising avenue in XAI research involves semantic-level explanation approaches, which leverage knowledge graphs and domain ontologies to translate numerical feature contributions into domain-specific concepts. This method enhances the intuitiveness of explanations by mapping abstract model outputs to familiar terms and domain concepts (e.g., medical biomarkers, such as "blood pressure" or "cholesterol levels").

The Explanation Ontology [11] offers a general-purpose semantic representation that supports user-centered explanations by connecting AI method outputs to underlying data and knowledge, as described in Fig. 5. This ontology enables diverse explanations (e.g., data, rationale, fairness) across healthcare, finance, and recommender systems to ensure accurate, context-aware explanations, enhancing user understanding and trust. Similarly, BioKG provides a method to map genetic characteristics to clinic terms to facilitat semantic explanation [49].

4 COMPUTATIONAL OVERHEAD AND SACRIFICED PERFORMANCE

Explainable AI (XAI) aims to make AI decisions transparent, but generating explanations can slow down systems, particularly in real-time applications like autonomous vehicles or financial trading. Methods like SHAP [31] require extensive computations, which can

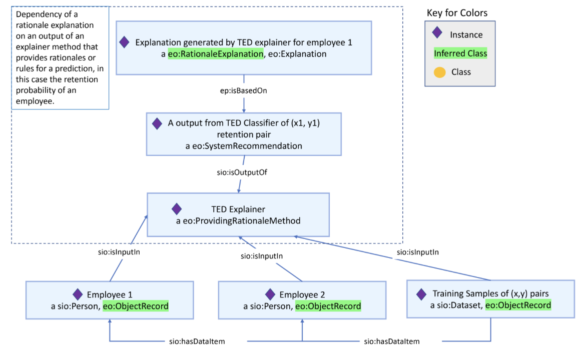

Figure 5: An application case of Explanation Ontology framework

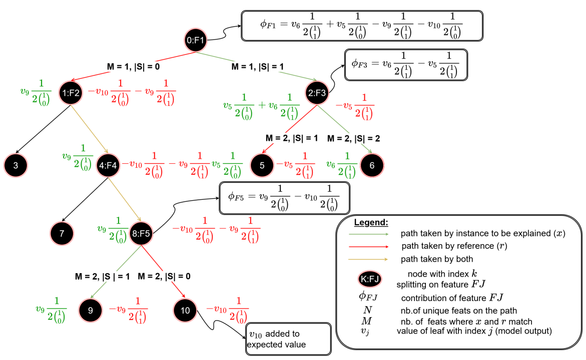

Figure 6: Illustration of the Tree SHAP method

be costly for large datasets, impacting system responsiveness [40], whereas in data mining contexts, the need for scalable XAI methods is evident.

4.1 Mitigating Computational Overhead

To address these challenges, researchers have proposed various optimization strategies to reduce the computational overhead.

An importance example is TreeSHAP [32], which uses tree structures to compute explanations faster, and sampling techniques to approximate results, balancing speed and accuracy, as illustrated in 6. GPUTreeSHAP [35] and Linear TreeSHAP [55] further improved computational speed while assuring high precision. Precomputing and caching can also be utilized to reduce latency to support the application of XAI in real-world industrial scenarios [18].

Non-amortized approaches further focus on instance-specific optimizations to provide extra acceleration, as surveyed by Chuang et al. [15]. These methods includes data-centric acceleration methods (e.g., SHEAR [50], which reduces Shapley value computation via contributive feature coalition selection) and model-centric acceleration (e.g., optimization-driven approximations like antithetical permutation sampling [34]).

However, methods to decrease computational overhead often introduce approximation and trade-offs, which lead to sacrificed model performance. For example, RKHS-SHAP [12], which offer model-agnostic solutions by using weighted linear regression to estimate SHAP values can significantly improve the speed for Sharply values generation but involved ambiguity. Assis et al. also found that opaque models like CNNs achieved up to 98% accuracy on datasets like MNIST, compared to 94% for transparent models that response as fast as CNNs [4]. In the meantime, models that are inherently fast, such as LIME, may be less reliable [17].

To summarize, the balance between explainability and performance remains a dynamic research frontier, with implications for the adoption of XAI in critical domains.

4.2 Advancing Inference Efficiency

Research also indicates that the continuous advancement in the field of High-Performance Computing (HPC) may significantly reduce the computational overhead of XAI methods, though this depends on specific applications and optimization techniques.

For example, Sarma et al. [43] introduced the AI4HPC library, which significantly enhances the training efficiency of AI models for CFD applications on HPC systems through distributed training, mixed-precision optimization, and adaptive gradient aggregation techniques. Its 96% strong scalability and optimization methods for non-uniform datasets provide a scalable parallelization framework, which illustrated the potential for HPC systems to process large-scale datasets, which are common in data mining scenarios. Huerta et al. [23] also pointed out the strong potential for HPC methods to cope with computationally intensive tasks.

Although HPC demonstrates significant potential, direct research on computational overhead in XAI remains quite limited, with the paper of GPUTreeSHAP [35] mentioned above as a precious example. Future efforts may focus on: (1) developing HPC algorithms specifically optimized for XAI, such as parallel SHAP computation or distributed LIME evaluation; (2) exploring how HPC can address the unique challenges of XAI, such as balancing accuracy and interpretability; (3) designing HPC-XAI integration solutions for real-time applications to ensure low latency.

5 MISLEADING INTERPRETATION POSSIBILITY

Explainable AI (XAI) aims to make AI decisions transparent, but explanations can sometimes mislead, especially in sensitive areas like healthcare or finance. Research shows that explanations are often context-dependent and can change over time, leading to distrust if they do not align with user expectations [7, 16, 19]. For example, traditional methods like LIME can have unreliable sampling [17], and newer NLP methods may hallucinate [26], leading to misleading explanations. This is a concern in high-stakes domains where trust is crucial, as misleading interpretations can erode confidence.

To solve this problem, recent research has proposed several metrics that address different dimensions of explainability [38], including faithfulness, monotonicity, stability, user satisfaction, etc. However, not every criterion can be computer-evaluated, and it was also noted that the proliferation of metrics can lead to possible challenges and may cause confusion [39].

Research also noted that SHAP [31] offers a balanced approach by providing detailed insights while maintaining theoretical rigor, which is more likely to mitigate potential misleading issues than other methods, such as LIME. Létoffé et al. [30] further integrate axiomatic aggregations derived from ML models into Sharply values to improve the robustness of SHAP method. However, limited research has been published to resolve the misleading explanation fault for cutting-edge XAI solutions, which remains a challenge for the academia to conquer.

6 CONCLUSION

This survey addresses the core scientific challenge of designing explainable AI (XAI) frameworks that maintain high predictive performance while providing human-understandable explanations in data mining applications. The four primary hurdles - complex data and model interactions [3], user cognition variability [36, 42], computational overhead [40], and explanation fidelity risks [17] — require integrated solutions. The development of adaptive multi-level explanation systems and semantic mapping through knowledge graphs [11] offers pathways to bridge technical explanations with domain-specific user needs. Future research directions should prioritize HPC-accelerated XAI implementations and LLM-powered explanation personalization, establish standardized evaluation metrics for explanation quality and conduct more research about new XAI algorithms and model-specific XAI methods. These advancements will enable XAI systems to achieve the critical balance between computational efficiency, predictive accuracy, and human-centric interpretability required for high-stakes applications.

REFERENCES

[1] Amina Adadi and Mohammed Berrada. 2018. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access* 6 (2018), 52138–52160.

[2] Maruan Al-Shedivat, Avinava Dubey, and Eric Xing. 2020. Contextual explanation networks. *Journal of Machine Learning Research* 21, 194 (2020), 1–44.

[3] Sajid Ali, Tamer Abuhmed, Shaker El-Sappagh, Khan Muhammad, Jose M. Alonso-Moral, Roberto Confalonieri, Riccardo Guidotti, Javier Del Ser, Natalia Díaz-Rodríguez, and Francisco Herrera. 2023. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion* 99 (2023), 101805. https://doi.org/10.1016/j.inffus.2023.101805

[4] André Assis, Douglas Véras, and Ermeson Andrade. 2023. Explainable Artificial Intelligence - An Analysis of the Trade-offs Between Performance and Explainability. In *2023 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*. 1–6. https://doi.org/10.1109/LA-CCI58595.2023.10409462

[5] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate.

[6] A. Bhatia, A. Chandani, and J. Chhateja. 2020. Robo advisory and its potential in addressing the behavioral biases of investors - a qualitative study in Indian context. *Journal of Behavioral and Experimental Finance* 25 (2020). forthcoming.

[7] Sebastian Breden, Florian Hinterwimmer, Sarah Consalvo, Jan Neumann, Carolin Knebel, Rüdiger Eisenhart-Rothe, Rainer Burgkart, and Ulrich Lenze. 2023. Deep Learning-Based Detection of Bone Tumors around the Knee in X-rays of Children. *Journal of Clinical Medicine* 12 (09 2023), 5960. https://doi.org/10.3390/jcm12185960

[8] Wieland Brendel and Matthias Bethge. 2019. Approximating cnns with bag-of-local-features models works surprisingly well on imagenet. *arXiv preprint arXiv:1904.00760* (2019).

[9] Ángel Cabrera, Fred Hohman, Jason Lin, and Duen Horng Chau. 2018. Interactive Classification for Deep Learning Interpretation. *Demo, IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018).

[10] Diogo V. Carvalho, Eduardo M. Pereira, and Jaime S. Cardoso. 2019. Machine Learning Interpretability: A Survey on Methods and Metrics. *Electronics* 8, 8 (2019). https://doi.org/10.3390/electronics8080832

[11] Shruthi Chari, Oshani Seneviratne, Mohamed Ghalwash, Sola Shirai, Daniel M. Gruen, Pablo Meyer, Prithwish Chakraborty, and Deborah L. McGuinness. 2024. Explanation Ontology: A general-purpose, semantic representation for supporting user-centered explanations. *Semantic Web* 15, 4 (2024), 959–989. https://doi.org/10.3233/SW-233282

[12] Siu Lun Chau, Robert Hu, Javier Gonzalez, and Dino Sejdinovic. 2022. RKHS-SHAP: Shapley Values for Kernel Methods. arXiv:2110.09167 [stat.ML]

[13] Chaofan Chen, Kangcheng Lin, Cynthia Rudin, Yaron Shaposhnik, Sijia Wang, and Tong Wang. 2021. A Holistic Approach to Interpretability in Financial Lending: Models, Visualizations, and Summary-Explanations. arXiv:2106.02605 [cs.LG]

[14] Tsung-Nan Chou. 2019. An Explainable Hybrid Model for Bankruptcy Prediction Based on the Decision Tree and Deep Neural Network. In *2019 IEEE 2nd International Conference on Knowledge Innovation and Invention (ICKII)*. 122–125. https://doi.org/10.1109/ICKII46306.2019.9042639

[15] Yu-Neng Chuang, Guanchu Wang, Fan Yang, Zirui Liu, Xuanting Cai, Mengnan Du, and Xia Hu. 2023. Efficient XAI Techniques: A Taxonomic Survey. arXiv:2302.03225 [cs.LG]

[16] Hans de Bruijn, Martijn Warnier, and Marijn Janssen. 2022. The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making. *Government Information Quarterly* 39, 2 (2022), 101666.

[17] Jürgen Dieber and Sabrina Kirrane. 2020. Why model why? Assessing the strengths and limitations of LIME. arXiv:2012.00093 [cs.LG]

[18] Krishna Gade, Sahin Cem Geyik, Krishnaram Kenthapadi, Varun Mithal, and Ankur Taly. 2019. Explainable AI in Industry *(KDD '19)*. Association for Computing Machinery, New York, NY, USA, 3203–3204. https://doi.org/10.1145/3292500.3332281

[19] Bhimaja Goonatillaka and Prasanna S. Haddela. 2024. Exploring Radiologists' Reluctance Towards Machine Learning Models and Explainable AI in Brain Tumor Detection. *2024 6th International Conference on Advancements in Computing (ICAC)* (2024), 25–30.

[20] Alexey Guryanov. 2021. Efficient Computation of SHAP Values for Piecewise-Linear Decision Trees. In *2021 International Conference on Information Technology and Nanotechnology (ITNT)*. 1–4. https://doi.org/10.1109/ITNT52450.2021.9649051

[21] Hani Hagras. 2018. Toward Human-Understandable, Explainable AI. *Computer* 51, 9 (2018), 28–36. https://doi.org/10.1109/MC.2018.3620965

[22] Xin He, Yeyi Hong, Xi Zheng, and Yong Zhang. 2022. What Are the Users' Needs? Design of a User-Centered Explainable Artificial Intelligence Diagnostic System. *International Journal of Human-Computer Interaction* 39 (07 2022), 1–24. https://doi.org/10.1080/10447318.2022.2095093

[23] E. A. Huerta, Asad Khan, Eliu Davis, Hongyang Zhu, Erik Johansson, Larry Titus, Daniel Heydari, Zhe Zhao, Yifan Zhao, Shantenu Jha, et al. 2020. Convergence of artificial intelligence and high performance computing on NSF-supported cyberinfrastructure. *Journal of Big Data* 7, 1 (2020), 88. https://doi.org/10.1186/s40537-020-00361-2

[24] Alon Jacovi and Yoav Goldberg. 2020. Towards faithfully interpretable NLP systems: How should we define and evaluate faithfulness? *arXiv preprint arXiv:2004.03685* (2020).

[25] Sarthak Jain and Byron C Wallace. 2019. Attention is not explanation.

[26] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Yejin Bang, Delong Chen, Wenliang Dai, Andrea Madotto, and Pascale Fung. 2022. Survey of Hallucination in Natural Language Generation. *Comput. Surveys* 55 (2022), 1 – 38.

[27] Taotao Jing, Haifeng Xia, Renran Tian, Haoran Ding, Xiao Luo, Joshua Domeyer, Rini Sherony, and Zhengming Ding. 2022. InAction: Interpretable Action Decision Making for Autonomous Driving. In *Guide Proceedings*. Springer-Verlag, Berlin, Germany, 370–387. https://doi.org/10.1007/978-3-031-19839-7_22

[28] Nagappan Krishnaveni and V. Radha. 2019. Feature Selection Algorithms for Data Mining Classification: A Survey. *Indian Journal of Science and Technology* (2019).

[29] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521 (May 2015), 436–444. https://doi.org/10.1038/nature14539

[30] Olivier Létoffé, Xuanxiang Huang, and Joao Marques-Silva. 2025. Towards Trustable SHAP Scores. *AAAI* 39, 17 (Apr 2025), 18198–18208. https://doi.org/10.1609/aaai.v39i17.34002

[31] Scott Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. arXiv:1705.07874 [cs.AI]

[32] Scott M. Lundberg, Gabriel G. Erion, and Su-In Lee. 2019. Consistent Individualized Feature Attribution for Tree Ensembles. arXiv:1802.03888 [cs.LG]

[33] Philip Mavrepis, Georgios Makridis, Georgios Fatouros, Vasileios Koukos, Maria Margarita Separdani, and Dimosthenis Kyriazis. 2024. XAI for All: Can Large Language Models Simplify Explainable AI? arXiv:2401.13110 [cs.AI]

[34] Rory Mitchell, Joshua Cooper, Eibe Frank, and Geoffrey Holmes. 2022. Sampling Permutations for Shapley Value Estimation. *Journal of Machine Learning Research* 23, 43 (2022), 1–46.

[35] Rory Mitchell, Eibe Frank, and Geoffrey Holmes. 2022. GPUTreeShap: Massively Parallel Exact Calculation of SHAP Scores for Tree Ensembles. arXiv:2010.13972 [cs.LG]

[36] Saja Ataallah Muhammed and Laith R. Flaih. 2024. Predictive Modeling in Healthcare: A Survey of Data Mining Applications. *5TH INTERNATIONAL CONFERENCE ON COMMUNICATION ENGINEERING AND COMPUTER SCIENCE (CIC-COCOS'24)* (2024).

[37] Ricardo Müller, Marco Schreyer, Timur Sattarov, and Damian Borth. 2022. RESHAPE: Explaining Accounting Anomalies in Financial Statement Audits by enhancing SHapley Additive exPlanations. In *Proceedings of the Third ACM International Conference on AI in Finance* (New York, NY, USA) *(ICAIF '22)*. Association for Computing Machinery, New York, NY, USA, 174–182. https://doi.org/10.1145/3533271.3561667

[38] Alexandr Oblizanov, Natalya Shevskaya, Anatoliy Kazak, Marina Rudenko, and Anna Dorofeeva. 2023. Evaluation Metrics Research for Explainable Artificial Intelligence Global Methods Using Synthetic Data. *Applied System Innovation* 6, 1 (2023).

[39] Marek Pawlicki, Aleksandra Pawlicka, Federica Uccello, Sebastian Szelest, Salvatore D'Antonio, Rafał Kozik, and Michał Choraś. 2024. Evaluating the necessity of the multiple metrics for assessing explainable AI: A critical examination. *Neurocomputing* 602 (2024), 128282.

[40] Romila Pradhan, Aditya Lahiri, Sainyam Galhotra, and Babak Salimi. 2022. Explainable ai: Foundations, applications, opportunities for data management research. In *Proceedings of the 2022 International Conference on Management of Data*. 2452–2457.

[41] Robert A. Rigby and D. M. Stasinopoulos. 2005. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54 (2005).

[42] Mozhgan Salimiparsa, Daniel Lizotte, and Kamran Sedig. 2021. A User-Centered Design of Explainable AI for Clinical Decision Support. *Proceedings of the Canadian Conference on Artificial Intelligence* (06 2021). https://doi.org/10.21428/594757db.62860442

[43] Rakesh Sarma, Eray Inanc, Marcel Aach, and Andreas Lintermann. 2024. Parallel and scalable AI in HPC systems for CFD applications and beyond. *Frontiers in High Performance Computing* Volume 2 - 2024 (2024). https://doi.org/10.3389/fhpcp.2024.1444337

[44] Tjeerd A.J. Schoonderwoerd, Wiard Jorritsma, Mark A. Neerincx, and Karel van den Bosch. 2021. Human-centered XAI: Developing design patterns for explanations of clinical decision support systems. *International Journal of Human-Computer Studies* 154 (2021), 102684. https://doi.org/10.1016/j.ijhcs.2021.102684

[45] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2019. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision* 128, 2 (Oct. 2019), 336–359. https://doi.org/10.1007/s11263-019-01228-7

[46] Avanti Shrikumar, Peyton Greenside, and Anshul Kundaje. 2019. Learning Important Features Through Propagating Activation Differences. arXiv:1704.02685 [cs.CV]

[47] Erico Tjoa and Cuntai Guan. 2020. A survey on explainable artificial intelligence (xai): Toward medical xai. *IEEE transactions on neural networks and learning systems* 32, 11 (2020), 4793–4813.

[48] Giulia Vilone and Luca Longo. 2021. Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion* 76 (2021), 89–106. https://doi.org/10.1016/j.inffus.2021.05.009

[49] Brian Walsh, Sameh K. Mohamed, and Vít Nováček. 2020. BioKG: A Knowledge Graph for Relational Learning On Biological Data. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (Virtual Event, Ireland) *(CIKM '20)*. 3173–3180. https://doi.org/10.1145/3340531.3412776

[50] Guanchu Wang, Yu-Neng Chuang, Mengnan Du, Fan Yang, Quan Zhou, Pushkar Tripathi, Xuanting Cai, and Xia Hu. 2022. Accelerating shapley explanation via contributive cooperator selection. In *International Conference on Machine Learning*. PMLR, 22576–22590.

[51] Q. Wang and H. Xia. 2025. An Explainable Multi-Level Approach for Financial Fraud Detection. In *Proceedings of the 6th International Conference on Structural Health Monitoring and Integrity Management* (Zhengzhou, 2025-11-08/2025-11-10). https://doi.org/10.58286/31004

[52] Zijie J. Wang, Robert Turko, Omar Shaikh, Haekyu Park, Nilaksh Das, Fred Hohman, Minsuk Kahng, and Duen Horng Polo Chau. 2021. CNN Explainer: Learning Convolutional Neural Networks with Interactive Visualization. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (Feb. 2021), 1396–1406. https://doi.org/10.1109/tvcg.2020.3030418

[53] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. 2023. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. arXiv:2201.11903 [cs.CL]

[54] Haoyi Xiong, Xuhong Li, Xiaofei Zhang, Jiamin Chen, Xinhao Sun, Yuchen Li, Zeyi Sun, and Mengnan Du. 2024. Towards Explainable Artificial Intelligence (XAI): A Data Mining Perspective. arXiv:2401.04374 [cs.AI]

[55] Peng Yu, Chao Xu, Albert Bifet, and Jesse Read. 2023. Linear TreeShap. arXiv:2209.08192 [cs.LG]