

# Does one XAI fit all?

## An Elastic Framework for Explainable Artificial Intelligence

### Auxiliary Reference List

#### ANNOTATED PAPERS

- [1] Z. A. E. Houda, B. Brik, and L. Khoukhi, ““Why Should I Trust Your IDS?”: An Explainable Deep Learning Framework for Intrusion Detection Systems in Internet of Things Networks,” *IEEE Open Journal of the Communications Society*, vol. 3, pp. 1164–1176, 2022.
- [2] E. Pintelas, I. E. Livieris, and P. Pintelas, “A Grey-Box Ensemble Model Exploiting Black-Box Accuracy and White-Box Intrinsic Interpretability,” *Algorithms*, vol. 13, no. 1, p. 17, Jan. 2020.
- [3] C. Chen, K. Lin, C. Rudin, Y. Shaposhnik, S. Wang, and T. Wang, “A holistic approach to interpretability in financial lending: Models, visualizations, and summary-explanations,” *Decision Support Systems*, vol. 152, p. 113647, Jan. 2022.
- [4] K. M. Hasib, F. Rahman, R. Hasnat, and M. G. R. Alam, “A Machine Learning and Explainable AI Approach for Predicting Secondary School Student Performance,” in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*. Las Vegas, NV, USA: IEEE, Jan. 2022, pp. 0399–0405.
- [5] I. A. Khan, N. Moustafa, D. Pi, K. M. Sallam, A. Y. Zomaya, and B. Li, “A New Explainable Deep Learning Framework for Cyber Threat Discovery in Industrial IoT Networks,” *IEEE Internet of Things Journal*, vol. 9, no. 13, pp. 11 604–11 613, Jul. 2022.
- [6] Y. Wang, Z. Wang, X. Kang, and Y. Luo, “A novel interpretable model ensemble multivariate fast iterative filtering and temporal fusion transform for carbon price forecasting,” *Energy Science & Engineering*, vol. 11, no. 3, pp. 1148–1179, 2023.
- [7] J. Adams and H. Hagsras, “A Type-2 Fuzzy Logic Approach to Explainable AI for regulatory compliance, fair customer outcomes and market stability in the Global Financial Sector,” in *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. Glasgow, UK: IEEE, Jul. 2020, pp. 1–8.
- [8] Z. Wang, W. Yan, C. Zeng, Y. Tian, and S. Dong, “A Unified Interpretable Intelligent Learning Diagnosis Framework for Learning Performance Prediction in Intelligent Tutoring Systems,” *International Journal of Intelligent Systems*, vol. 2023, no. 1, p. 4468025, 2023.
- [9] J. Xing, T. Nagata, X. Zou, E. Neftci, and J. L. Krichmar, “Achieving efficient interpretability of reinforcement learning via policy distillation and selective input gradient regularization,” *Neural Networks*, vol. 161, pp. 228–241, Apr. 2023.
- [10] M. Szczepański, M. Choraś, M. Pawlicki, and R. Kozik, “Achieving Explainability of Intrusion Detection System by Hybrid Oracle-Explainer Approach,” in *2020 International Joint Conference on Neural Networks (IJCNN)*. Glasgow, UK: IEEE, Jul. 2020, pp. 1–8.
- [11] T. Zebin, S. Rezvy, and Y. Luo, “An Explainable AI-Based Intrusion Detection System for DNS Over HTTPS (DoH) Attacks,” *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 2339–2349, 2022.
- [12] J. Collenette, K. Atkinson, and T. Bench-Capon, “An Explainable Approach to Deducing Outcomes in European Court of Human Rights Cases Using ADFs,” in *Computational Models of Argument*. Amsterdam, The Netherlands: IOS Press, 2020, pp. 21–32.
- [13] Z. Zhang, C. Wu, S. Qu, and X. Chen, “An explainable artificial intelligence approach for financial distress prediction,” *Information Processing & Management*, vol. 59, no. 4, p. 102988, Jul. 2022.
- [14] A. Oseni, N. Moustafa, G. Creech, N. Sohrabi, A. Strelzoff, Z. Tari, and I. Linkov, “An Explainable Deep Learning Framework for Resilient Intrusion Detection in IoT-Enabled Transportation Networks,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 1, pp. 1000–1014, Jan. 2023.
- [15] M. Keshk, N. Koroniotis, N. Pham, N. Moustafa, B. Turnbull, and A. Y. Zomaya, “An explainable deep learning-enabled intrusion detection framework in IoT networks,” *Information Sciences*, vol. 639, p. 119000, Aug. 2023.
- [16] J. G. C. Krüger, A. d. S. Britto, and J. P. Barddal, “An explainable machine learning approach for student dropout prediction,” *Expert Systems with Applications*, vol. 233, p. 120933, Dec. 2023.
- [17] M. Wang, K. Zheng, Y. Yang, and X. Wang, “An Explainable Machine Learning Framework for Intrusion Detection Systems,” *IEEE Access*, vol. 8, pp. 73 127–73 141, 2020.
- [18] L. Yan, H.-T. Zhang, J. Goncalves, Y. Xiao, M. Wang, Y. Guo, C. Sun, X. Tang, L. Jing, M. Zhang, X. Huang, Y. Xiao, H. Cao, Y. Chen, T. Ren, F. Wang, Y. Xiao, S. Huang, X. Tan, N. Huang, B. Jiao, C. Cheng, Y. Zhang, A. Luo, L. Mombaerts, J. Jin, Z. Cao, S. Li, H. Xu, and Y. Yuan, “An interpretable mortality prediction model for COVID-19 patients,” *Nature Machine Intelligence*, vol. 2, no. 5, pp. 283–288, May 2020.
- [19] Z. Mingyu, W. Sutong, W. Yanzhang, and W. Dujuan, “An interpretable prediction method for university student academic crisis warning,” *Complex & Intelligent Systems*, vol. 8, no. 1, pp. 323–336, Feb. 2022.
- [20] G. Mitrentsis and H. Lens, “An interpretable probabilistic model for short-term solar power forecasting using natural gradient boosting,” *Applied Energy*, vol. 309, p. 118473, Mar. 2022.
- [21] K. Mridha, M. M. Uddin, J. Shin, S. Khadka, and M. F. Mridha, “An Interpretable Skin Cancer Classification Using Optimized Convolutional Neural Network for a Smart Healthcare System,” *IEEE Access*, vol. 11, pp. 41 003–41 018, 2023.
- [22] G. Singer and I. Cohen, “An Objective-Based Entropy Approach for Interpretable Decision Tree Models in Support of Human Resource Management: The Case of Absenteeism at Work,” *Entropy*, vol. 22, no. 8, p. 821, Aug. 2020.
- [23] T. Zhou, H. Sheng, and I. Howley, “Assessing Post-hoc Explainability of the BKT Algorithm,” in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, ser. AIES ’20. New York, NY, USA: Association for Computing Machinery, Feb. 2020, pp. 407–413.
- [24] S. Solayman, S. A. Aumi, C. S. Mery, M. Mubassir, and R. Khan, “Automatic COVID-19 prediction using explainable machine learning techniques,” *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 36–46, Jun. 2023.
- [25] Y. Shen and Y. Pan, “BIM-supported automatic energy performance analysis for green building design using explainable machine learning and multi-objective optimization,” *Applied Energy*, vol. 333, p. 120575, Mar. 2023.
- [26] A. Kuppaa and N.-A. Le-Khac, “Black Box Attacks on Explainable Artificial Intelligence(XAI) methods in Cyber Security,” in *2020 International Joint Conference on Neural Networks (IJCNN)*. Glasgow, UK: IEEE, Jul. 2020, pp. 1–8.
- [27] M. A. Elsayed and N. Zincir-Heywood, “BoostGuard: Interpretable Misbehavior Detection in Vehicular Communication Networks,” in *NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium*. Budapest, Hungary: IEEE, Apr. 2022, pp. 1–9.
- [28] L. M. Schmidt, G. Kontes, A. Plinge, and C. Mutschler, “Can You Trust Your Autonomous Car? Interpretable and Verifiably Safe Reinforcement Learning,” in *2021 IEEE Intelligent Vehicles Symposium (IV)*. Nagoya, Japan: IEEE, Jul. 2021, pp. 171–178.
- [29] B. M. Maweu, S. Dakshit, R. Shamsuddin, and B. Prabhakaran, “CEFEs: A CNN Explainable Framework for ECG Signals,” *Artificial Intelligence in Medicine*, vol. 115, p. 102059, May 2021.
- [30] D. Nemirovsky, N. Thiebaut, Y. Xu, and A. Gupta, “CounterGAN: Generating counterfactuals for real-time recourse and interpretability using residual GANs,” in *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence*. Eindhoven, The Netherlands: PMLR, Aug. 2022, pp. 1488–1497.
- [31] D. Shome, T. Kar, S. N. Mohanty, P. Tiwari, K. Muhammad, A. Al-Tameem, Y. Zhang, and A. K. J. Soudagar, “COVID-Transformer: Interpretable COVID-19 Detection Using Vision Transformer for Healthcare,” *International Journal of Environmental Research and Public Health*, vol. 18, no. 21, p. 11086, Jan. 2021.

- [32] R. R. Karn, P. Kudva, H. Huang, S. Suneja, and I. M. Elfadel, "Cryptomining Detection in Container Clouds Using System Calls and Explainable Machine Learning," *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 3, pp. 674–691, Mar. 2021.
- [33] Y. Zhang, B. K. Teoh, M. Wu, J. Chen, and L. Zhang, "Data-driven estimation of building energy consumption and GHG emissions using explainable artificial intelligence," *Energy*, vol. 262, p. 125468, Jan. 2023.
- [34] J. M. Dolezal, R. Wolk, H. M. Hieromnimon, F. M. Howard, A. Srisuwanakorn, D. Karpeyev, S. Ramesh, S. Kochanny, J. W. Kwon, M. Agni, R. C. Simon, C. Desai, R. Kherallah, T. D. Nguyen, J. J. Schulte, K. Cole, G. Khramtsova, M. C. Garassino, A. N. Husain, H. Li, R. Grossman, N. A. Cipriani, and A. T. Pearson, "Deep learning generates synthetic cancer history for explainability and education," *npj Precision Oncology*, vol. 7, no. 1, pp. 1–13, May 2023.
- [35] A. Gaudio, A. Smailagic, C. Faloutsos, S. Mohan, E. Johnson, Y. Liu, P. Costa, and A. Campilho, "DeepFixCX: Explainable privacy-preserving image compression for medical image analysis," *WIREs Data Mining and Knowledge Discovery*, vol. 13, no. 4, p. e1495, 2023.
- [36] A. Raza, K. P. Tran, L. Koehl, and S. Li, "Designing ECG monitoring healthcare system with federated transfer learning and explainable AI," *Knowledge-Based Systems*, vol. 236, p. 107763, Jan. 2022.
- [37] J. Dong, S. Chen, M. Miralinaghi, T. Chen, and S. Labi, "Development and testing of an image transformer for explainable autonomous driving systems," *Journal of Intelligent and Connected Vehicles*, vol. 5, no. 3, pp. 235–249, Jul. 2022.
- [38] M. Johnson, A. Albizri, A. Harfouche, and S. Tutun, "Digital transformation to mitigate emergency situations: Increasing opioid overdose survival rates through explainable artificial intelligence," *Industrial Management & Data Systems*, vol. 123, no. 1, pp. 324–344, Oct. 2021.
- [39] S.-L. Lin and X. Jin, "Does ESG Predict Systemic Banking Crises? A Computational Economics Model of Early Warning Systems with Interpretable Multi-Variable LSTM based on Mixture Attention," *Mathematics*, vol. 11, no. 2, p. 410, Jan. 2023.
- [40] N. Stef, H. Başağaoğlu, D. Chakraborty, and S. Ben Jabeur, "Does institutional quality affect CO2 emissions? Evidence from explainable artificial intelligence models," *Energy Economics*, vol. 124, p. 106822, Aug. 2023.
- [41] H.-C. Thorsen-Meyer, A. B. Nielsen, A. P. Nielsen, B. S. Kaas-Hansen, P. Toft, J. Schierbeck, T. Strøm, P. J. Chmura, M. Heimann, L. Dybdahl, L. Spangsege, P. Hulsén, K. Belling, S. Brunak, and A. Perner, "Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: A retrospective study of high-frequency data in electronic patient records," *The Lancet Digital Health*, vol. 2, no. 4, pp. e179–e191, Apr. 2020.
- [42] S. Chowdhury, J.-E. Sian, D. Prasanta Kumar, B. Sudeshna, and A. Kharlamov, "Embedding transparency in artificial intelligence machine learning models: Managerial implications on predicting and explaining employee turnover," *The International Journal of Human Resource Management*, vol. 34, no. 14, pp. 2732–2764, Aug. 2023.
- [43] A. Wolanin, G. Mateo-García, G. Camps-Valls, L. Gómez-Chova, M. Meroni, G. Duveiller, Y. Liangzhi, and L. Guanter, "Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt," *Environmental Research Letters*, vol. 15, no. 2, p. 024019, Feb. 2020.
- [44] V. Swamy, B. Radmehr, N. Krco, M. Marras, and T. Käser, "Evaluating the Explainers: Black-Box Explainable Machine Learning for Student Success Prediction in MOOCs," in *Proceedings of the 15th International Conference on Educational Data Mining*. Durham, UK: IEDMS, 2022, p. 98.
- [45] M. J. Ariza-Garzon, J. Arroyo, A. Caparrini, and M.-J. Segovia-Vargas, "Explainability of a Machine Learning Granting Scoring Model in Peer-to-Peer Lending," *IEEE Access*, vol. 8, pp. 64 873–64 890, 2020.
- [46] C. Belloni, A. Balleri, N. Aouf, J.-M. Le Caillec, and T. Merlet, "Explainability of Deep SAR ATR Through Feature Analysis," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 1, pp. 659–673, Feb. 2021.
- [47] M. S. Hossain, G. Muhammad, and N. Guizani, "Explainable AI and Mass Surveillance System-Based Healthcare Framework to Combat COVID-19 Like Pandemics," *IEEE Network*, vol. 34, no. 4, pp. 126–132, Jul. 2020.
- [48] A. P. Carrieri, N. Haiminen, S. Maudsley-Barton, L.-J. Gardiner, B. Murphy, A. E. Mayes, S. Paterson, S. Grimshaw, M. Winn, C. Shand, P. Hadjidakos, W. P. M. Rowe, S. Hawkins, A. MacGuire-Flanagan, J. Tazzioli, J. G. Kenny, L. Parida, M. Hoptroff, and E. O. Pyzer-Knapp, "Explainable AI reveals changes in skin microbiome composition linked to phenotypic differences," *Scientific Reports*, vol. 11, no. 1, p. 4565, Feb. 2021.
- [49] Y. Ramon, R. A. Farrokhnia, S. C. Matz, and D. Martens, "Explainable AI for Psychological Profiling from Behavioral Data: An Application to Big Five Personality Predictions from Financial Transaction Records," *Information*, vol. 12, no. 12, p. 518, Dec. 2021.
- [50] G. Rjoub, J. Bentahar, and O. A. Wahab, "Explainable AI-based Federated Deep Reinforcement Learning for Trusted Autonomous Driving," in *2022 International Wireless Communications and Mobile Computing (IWCMC)*. Dubrovnik, Croatia: IEEE, May 2022, pp. 318–323.
- [51] S. Knapic, A. Malhi, R. Saluja, and K. Främling, "Explainable Artificial Intelligence for Human Decision Support System in the Medical Domain," *Machine Learning and Knowledge Extraction*, vol. 3, no. 3, pp. 740–770, Sep. 2021.
- [52] C. K. Leung, A. G. Pazdor, and J. Souza, "Explainable Artificial Intelligence for Data Science on Customer Churn," in *2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA)*. Porto, Portugal: IEEE, Oct. 2021, pp. 1–10.
- [53] B. Mahbooba, M. Timilsina, R. Sahal, and M. Serrano, "Explainable Artificial Intelligence (XAI) to Enhance Trust Management in Intrusion Detection Systems Using Decision Tree Model," *Complexity*, vol. 2021, no. 1, p. 6634811, 2021.
- [54] M. Ryo, "Explainable artificial intelligence and interpretable machine learning for agricultural data analysis," *Artificial Intelligence in Agriculture*, vol. 6, pp. 257–265, Jan. 2022.
- [55] L. Zhou, X. Shi, Y. Bao, L. Gao, and C. Ma, "Explainable artificial intelligence for digital finance and consumption upgrading," *Finance Research Letters*, vol. 58, p. 104489, Dec. 2023.
- [56] G. Babaei, P. Giudici, and E. Raffinetti, "Explainable FinTech lending," *Journal of Economics and Business*, vol. 125–126, p. 106126, May 2023.
- [57] A. Dahal and L. Lombardo, "Explainable artificial intelligence in geoscience: A glimpse into the future of landslide susceptibility modeling," *Computers & Geosciences*, vol. 176, p. 105364, Jul. 2023.
- [58] Ł. Górski and S. Ramakrishna, "Explainable artificial intelligence, lawyer's perspective," in *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law*, ser. ICAIL '21. New York, NY, USA: Association for Computing Machinery, Jul. 2021, pp. 60–68.
- [59] L. He, N. Aouf, and B. Song, "Explainable Deep Reinforcement Learning for UAV autonomous path planning," *Aerospace Science and Technology*, vol. 118, p. 107052, Nov. 2021.
- [60] S. M. Lauritsen, M. Kristensen, M. V. Olsen, M. S. Larsen, K. M. Lauritsen, M. J. Jørgensen, J. Lange, and B. Thiesson, "Explainable artificial intelligence model to predict acute critical illness from electronic health records," *Nature Communications*, vol. 11, no. 1, p. 3852, Jul. 2020.
- [61] G. D. Luca and Y. Chen, "Explainable Artificial Intelligence for Workflow Verification in Visual IoT/Robotics Programming Language Environment," *Journal of Artificial Intelligence and Technology*, vol. 1, no. 1, pp. 21–27, 2021.
- [62] J. Barria-Pineda, K. Akhuseyinoglu, S. Želem-Čelap, P. Brusilovsky, A. K. Milicevic, and M. Ivanovic, "Explainable Recommendations in a Personalized Programming Practice System," in *Artificial Intelligence in Education*, I. Roll, D. McNamara, S. Sosnovsky, R. Luckin, and V. Dimitrova, Eds. Cham: Springer International Publishing, 2021, pp. 64–76.
- [63] R. Bautista-Montesano, R. Bustamante-Bello, and R. A. Ramirez-Mendoza, "Explainable navigation system using fuzzy reinforcement learning," *International Journal on Interactive Design and Manufacturing (IIJIDeM)*, vol. 14, no. 4, pp. 1411–1428, Dec. 2020.
- [64] M. Onishi and T. Ise, "Explainable identification and mapping of trees using UAV RGB image and deep learning," *Scientific Reports*, vol. 11, no. 1, p. 903, Jan. 2021.
- [65] M. Saarela, V. Heilala, P. Jääskelä, A. Rantakaulio, and T. Kärkkäinen, "Explainable Student Agency Analytics," *IEEE Access*, vol. 9, pp. 137 444–137 459, 2021.
- [66] K. L. Tran, H. A. Le, T. H. Nguyen, and D. T. Nguyen, "Explainable Machine Learning for Financial Distress Prediction: Evidence from Vietnam," *Data*, vol. 7, no. 11, p. 160, Nov. 2022.
- [67] W. Yu, Z. Sun, J. Xu, Z. Dong, X. Chen, H. Xu, and J.-R. Wen, "Explainable Legal Case Matching via Inverse Optimal Transport-based Rationale Extraction," in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '22. New York, NY, USA: Association for Computing Machinery, Jul. 2022, pp. 657–668.

- [68] B. Albreiki, T. Habuza, and N. Zaki, "Framework for automatically suggesting remedial actions to help students at risk based on explainable ML and rule-based models," *International Journal of Educational Technology in Higher Education*, vol. 19, no. 1, p. 49, Sep. 2022.
- [69] H. Anysz, Ł. Brzozowski, W. Kretowicz, and P. Narloch, "Feature Importance of Stabilised Rammed Earth Components Affecting the Compressive Strength Calculated with Explainable Artificial Intelligence Tools," *Materials*, vol. 13, no. 10, p. 2317, Jan. 2020.
- [70] V. Bellandi, P. Ceravolo, S. Maghool, and S. Siccardi, "Graph embeddings in criminal investigation: Towards combining precision, generalization and transparency," *World Wide Web*, vol. 25, no. 6, pp. 2379–2402, Nov. 2022.
- [71] G. D'Angelo, D. Della-Morte, D. Pastore, G. Donadel, A. De Stefano, and F. Palmieri, "Identifying patterns in multiple biomarkers to diagnose diabetic foot using an explainable genetic programming-based approach," *Future Generation Computer Systems*, vol. 140, pp. 138–150, Mar. 2023.
- [72] H. Liu, C. Zhong, A. Alnusair, and S. R. Islam, "FAIXID: A Framework for Enhancing AI Explainability of Intrusion Detection Results Using Data Cleaning Techniques," *Journal of Network and Systems Management*, vol. 29, no. 4, p. 40, May 2021.
- [73] J. Chen, Z. Liu, S. Huang, Q. Liu, and W. Luo, "Improving Interpretability of Deep Sequential Knowledge Tracing Models with Question-centric Cognitive Representations," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 12, pp. 14196–14204, Jun. 2023.
- [74] A. Dikshit and B. Pradhan, "Interpretable and explainable AI (XAI) model for spatial drought prediction," *Science of The Total Environment*, vol. 801, p. 149797, Dec. 2021.
- [75] R. ElShawi, Y. Sherif, M. Al-Mallah, and S. Sakr, "Interpretability in healthcare: A comparative study of local machine learning interpretability techniques," *Computational Intelligence*, vol. 37, no. 4, pp. 1633–1650, 2021.
- [76] S. Mohammadinejad, J. V. Deshmukh, A. G. Puranic, M. Vazquez-Chanlatte, and A. Donzé, "Interpretable classification of time-series data using efficient enumerative techniques," in *Proceedings of the 23rd International Conference on Hybrid Systems: Computation and Control*, ser. HSCC '20. New York, NY, USA: Association for Computing Machinery, Apr. 2020, pp. 1–10.
- [77] M. Todorovic, N. Stanisic, M. Zivkovic, N. Bacanin, V. Simic, and E. B. Tirkolaee, "Improving audit opinion prediction accuracy using metaheuristics-tuned XGBoost algorithm with interpretable results through SHAP value analysis," *Applied Soft Computing*, vol. 149, p. 110955, Dec. 2023.
- [78] H. Wang, H. Gao, S. Yuan, H. Zhao, K. Wang, X. Wang, K. Li, and D. Li, "Interpretable Decision-Making for Autonomous Vehicles at Highway On-Ramps With Latent Space Reinforcement Learning," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 9, pp. 8707–8719, Sep. 2021.
- [79] M. Baranyi, M. Nagy, and R. Molontay, "Interpretable Deep Learning for University Dropout Prediction," in *Proceedings of the 21st Annual Conference on Information Technology Education*, ser. SIGITE '20. New York, NY, USA: Association for Computing Machinery, Oct. 2020, pp. 13–19.
- [80] J. Chen, S. E. Li, and M. Tomizuka, "Interpretable End-to-End Urban Autonomous Driving With Latent Deep Reinforcement Learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 6, pp. 5068–5078, Jun. 2022.
- [81] Y. Y. Hsin, T. S. Dai, Y. W. Ti, and M. C. Huang, "Interpretable Electronic Transfer Fraud Detection with Expert Feature Constructions: 2021 International Conference on Information and Knowledge Management Workshops, CIKMW 2021," *CEUR Workshop Proceedings*, vol. 3052, 2021.
- [82] A. Kuehlkamp, A. Boyd, A. Czajka, K. Bowyer, P. Flynn, D. Chute, and E. Benjamin, "Interpretable Deep Learning-Based Forensic Iris Segmentation and Recognition," in *2022 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*. Waikoloa, HI, USA: IEEE, Jan. 2022, pp. 359–368.
- [83] M. Nagy and R. Molontay, "Interpretable Dropout Prediction: Towards XAI-Based Personalized Intervention," *International Journal of Artificial Intelligence in Education*, vol. 34, no. 2, pp. 274–300, Jun. 2024.
- [84] X. Zhang, N. Wang, H. Shen, S. Ji, X. Luo, and T. Wang, "Interpretable deep learning under fire," in *Proceedings of the 29th USENIX Conference on Security Symposium*, ser. SEC'20. USA: USENIX Association, Aug. 2020, pp. 1659–1676.
- [85] R. Bhambhoria, H. Liu, S. Dahan, and X. Zhu, "Interpretable Low-Resource Legal Decision Making," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 11, pp. 11 819–11 827, Jun. 2022.
- [86] X. Chang, W. Li, J. Ma, T. Yang, and A. Y. Zomaya, "Interpretable Machine Learning In Sustainable Edge Computing: A Case Study of Short-Term Photovoltaic Power Output Prediction," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Barcelona, Spain: IEEE, May 2020, pp. 8981–8985.
- [87] G. Kostopoulos, T. Panagiotakopoulos, S. Kotsiantis, C. Pierrakeas, and A. Kameas, "Interpretable Models for Early Prediction of Certification in MOOCs: A Case Study on a MOOC for Smart City Professionals," *IEEE Access*, vol. 9, pp. 165 881–165 891, 2021.
- [88] N. Kundu, G. Rani, V. S. Dhaka, K. Gupta, S. C. Nayak, S. Verma, M. F. Ijaz, and M. Woźniak, "IoT and Interpretable Machine Learning Based Framework for Disease Prediction in Pearl Millet," *Sensors*, vol. 21, no. 16, p. 5386, Jan. 2021.
- [89] L. Li, J. Qiao, G. Yu, L. Wang, H.-Y. Li, C. Liao, and Z. Zhu, "Interpretable tree-based ensemble model for predicting beach water quality," *Water Research*, vol. 211, p. 118078, Mar. 2022.
- [90] J. Liu, C. Li, P. Ouyang, J. Liu, and C. Wu, "Interpreting the prediction results of the tree-based gradient boosting models for financial distress prediction with an explainable machine learning approach," *Journal of Forecasting*, vol. 42, no. 5, pp. 1112–1137, 2023.
- [91] B. Wu, L. Wang, and Y.-R. Zeng, "Interpretable wind speed prediction with multivariate time series and temporal fusion transformers," *Energy*, vol. 252, p. 123990, Aug. 2022.
- [92] C. Zhao, C. Wang, G. Hu, H. Chen, C. Liu, and J. Tang, "ISTVT: Interpretable Spatial-Temporal Video Transformer for Deepfake Detection," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 1335–1348, 2023.
- [93] J. Ault, J. P. Hanna, and G. Sharon, "Learning an Interpretable Traffic Signal Control Policy," in *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, ser. AAMAS '20. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, May 2020, pp. 88–96.
- [94] J. Chakraborty, K. Peng, and T. Menzies, "Making fair ML software using trustworthy explanation," in *Proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering*, ser. ASE '20. New York, NY, USA: Association for Computing Machinery, Jan. 2021, pp. 1229–1233.
- [95] B. Ivanovic, A. Elhafi, G. Rosman, A. Gaidon, and M. Pavone, "MATs: An Interpretable Trajectory Forecasting Representation for Planning and Control," in *Proceedings of the 2020 Conference on Robot Learning*. Massachusetts, USA: PMLR, Oct. 2021, pp. 2243–2256.
- [96] H. Mankodiya, D. Jadav, R. Gupta, S. Tanwar, W.-C. Hong, and R. Sharma, "OD-XAI: Explainable AI-Based Semantic Object Detection for Autonomous Vehicles," *Applied Sciences*, vol. 12, no. 11, p. 5310, Jan. 2022.
- [97] C. Schorr, P. Goodarzi, F. Chen, and T. Dahmen, "Neuroscope: An Explainable AI Toolbox for Semantic Segmentation and Image Classification of Convolutional Neural Nets," *Applied Sciences*, vol. 11, no. 5, p. 2199, Jan. 2021.
- [98] H. Wang, P. Cai, Y. Sun, L. Wang, and M. Liu, "Learning Interpretable End-to-End Vision-Based Motion Planning for Autonomous Driving with Optical Flow Distillation," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. Xi'an, China: IEEE, May 2021, pp. 13 731–13 737.
- [99] L. Lomazzi, S. Fabiano, M. Parziale, M. Giglio, and F. Cadini, "On the explainability of convolutional neural networks processing ultrasonic guided waves for damage diagnosis," *Mechanical Systems and Signal Processing*, vol. 183, p. 109642, Jan. 2023.
- [100] E. Melo, I. Silva, D. G. Costa, C. M. D. Viegas, and T. M. Barros, "On the Use of eXplainable Artificial Intelligence to Evaluate School Dropout," *Education Sciences*, vol. 12, no. 12, p. 845, Dec. 2022.
- [101] A. Yasodhara, A. Asgarian, D. Huang, and P. Sobhani, "On the Trustworthiness of Tree Ensemble Explainability Methods," in *Machine Learning and Knowledge Extraction*, A. Holzinger, P. Kieseberg, A. M. Tjoa, and E. Weippl, Eds. Cham: Springer International Publishing, 2021, pp. 293–308.
- [102] W. Guo, "Partially Explainable Big Data Driven Deep Reinforcement Learning for Green 5G UAV," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*. Dublin, Ireland: IEEE, Jun. 2020, pp. 1–7.
- [103] R. Liu, F. Mai, Z. Shan, and Y. Wu, "Predicting shareholder litigation on insider trading from financial text: An interpretable deep learning approach," *Information & Management*, vol. 57, no. 8, p. 103387, Dec. 2020.

- [104] G. A. Lyngdoh, M. Zaki, N. M. A. Krishnan, and S. Das, "Prediction of concrete strengths enabled by missing data imputation and interpretable machine learning," *Cement and Concrete Composites*, vol. 128, p. 104414, Apr. 2022.
- [105] L. Munkhdalai, T. Munkhdalai, V.-H. Pham, M. Li, K. H. Ryu, and N. Theera-Umpon, "Recurrent Neural Network-Augmented Locally Adaptive Interpretable Regression for Multivariate Time-Series Forecasting," *IEEE Access*, vol. 10, pp. 11 871–11 885, 2022.
- [106] D. Shah, V. Trivedi, V. Sheth, A. Shah, and U. Chauhan, "ResTS: Residual Deep interpretable architecture for plant disease detection," *Information Processing in Agriculture*, vol. 9, no. 2, pp. 212–223, Jun. 2022.
- [107] W. Wang, Y. Pei, S.-H. Wang, J. manuel Gorz, and Y.-D. Zhang, "PSTCNN: Explainable COVID-19 diagnosis using PSO-guided self-tuning CNN," *Biocell : official journal of the Sociedades Latinoamericanas de Microscopia Electronica ... et. al*, vol. 47, no. 2, pp. 373–384, 2023.
- [108] K. Yang, H. Yuan, and R. Y. K. Lau, "PsyCredit: An interpretable deep learning-based credit assessment approach facilitated by psychometric natural language processing," *Expert Systems with Applications*, vol. 198, p. 116847, Jul. 2022.
- [109] L. K. Branting, C. Pfeifer, B. Brown, L. Ferro, J. Aberdeen, B. Weiss, M. Pfaff, and B. Liao, "Scalable and explainable legal prediction," *Artificial Intelligence and Law*, vol. 29, no. 2, pp. 213–238, Jun. 2021.
- [110] D. Chakraborty, A. Alam, S. Chaudhuri, H. Başağaoğlu, T. Sulbaran, and S. Langar, "Scenario-based prediction of climate change impacts on building cooling energy consumption with explainable artificial intelligence," *Applied Energy*, vol. 291, p. 116807, Jun. 2021.
- [111] P. Giudici and E. Raffinetti, "Shapley-Lorenz eXplainable Artificial Intelligence," *Expert Systems with Applications*, vol. 167, p. 114104, Apr. 2021.
- [112] T. Hickling, N. Aouf, and P. Spencer, "Robust Adversarial Attacks Detection Based on Explainable Deep Reinforcement Learning for UAV Guidance and Planning," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 10, pp. 4381–4394, Oct. 2023.
- [113] Q. Liu, D. Gui, L. Zhang, J. Niu, H. Dai, G. Wei, and B. X. Hu, "Simulation of regional groundwater levels in arid regions using interpretable machine learning models," *Science of The Total Environment*, vol. 831, p. 154902, Jul. 2022.
- [114] B. Pradhan, S. Lee, A. Dikshit, and H. Kim, "Spatial flood susceptibility mapping using an explainable artificial intelligence (XAI) model," *Geoscience Frontiers*, vol. 14, no. 6, p. 101625, Nov. 2023.
- [115] L. R. Rodríguez Oconitrillo, J. J. Vargas, A. Camacho, Á. Burgos, and J. M. Corchado, "RYEL: An Experimental Study in the Behavioral Response of Judges Using a Novel Technique for Acquiring Higher-Order Thinking Based on Explainable Artificial Intelligence and Case-Based Reasoning," *Electronics*, vol. 10, no. 12, p. 1500, Jan. 2021.
- [116] H. Shao, L. Wang, R. Chen, H. Li, and Y. Liu, "Safety-Enhanced Autonomous Driving Using Interpretable Sensor Fusion Transformer," in *Proceedings of The 6th Conference on Robot Learning*. Auckland, New Zealand: PMLR, Mar. 2023, pp. 726–737.
- [117] D. L. Aguilar, M. A. Medina-Pérez, O. Loyola-González, K.-K. R. Choo, and E. Bucheli-Susarrey, "Towards an Interpretable Autoencoder: A Decision-Tree-Based Autoencoder and its Application in Anomaly Detection," *IEEE Transactions on Dependable and Secure Computing*, vol. 20, no. 2, pp. 1048–1059, Mar. 2023.
- [118] J. M. Alonso, "Teaching Explainable Artificial Intelligence to High School Students," *International Journal of Computational Intelligence Systems*, vol. 13, no. 1, pp. 974–987, Jan. 2020.
- [119] G. Iadarola, F. Martinelli, F. Mercaldo, and A. Santone, "Towards an interpretable deep learning model for mobile malware detection and family identification," *Computers & Security*, vol. 105, p. 102198, Jun. 2021.
- [120] M. E. Irrarázaval, S. Maldonado, J. Pérez, and C. Vairetti, "Telecom traffic pumping analytics via explainable data science," *Decision Support Systems*, vol. 150, p. 113559, Nov. 2021.
- [121] N. Mirchi, V. Bissonnette, R. Yilmaz, N. Ledwos, A. Winkler-Schwartz, and R. F. D. Maestro, "The Virtual Operative Assistant: An explainable artificial intelligence tool for simulation-based training in surgery and medicine," *PLOS ONE*, vol. 15, no. 2, p. e0229596, Feb. 2020.
- [122] M. Norkute, N. Herger, L. Michalak, A. Mulder, and S. Gao, "Towards Explainable AI: Assessing the Usefulness and Impact of Added Explainability Features in Legal Document Summarization," in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '21. New York, NY, USA: Association for Computing Machinery, May 2021, pp. 1–7.
- [123] C. Sardanios, I. Varlamis, C. Chronis, G. Dimitrakopoulos, A. Al-salemi, Y. Himeur, F. Bensaali, and A. Amira, "The emergence of explainability of intelligent systems: Delivering explainable and personalized recommendations for energy efficiency," *International Journal of Intelligent Systems*, vol. 36, no. 2, pp. 656–680, 2021.
- [124] H. Tan, B. Zhang, H. Zhang, and R. Li, "The Sentencing-Element-Aware Model for Explainable Term-of-Penalty Prediction," in *Natural Language Processing and Chinese Computing*, X. Zhu, M. Zhang, Y. Hong, and R. He, Eds. Cham: Springer International Publishing, 2020, pp. 16–27.
- [125] M. Xiong, H. Wang, C. Che, and R. Lin, "Toward safer aviation: Application of GA-XGBoost-SHAP for incident cognition and model explainability," *Proceedings of the Institution of Mechanical Engineers, Part O*, vol. 238, no. 6, pp. 1195–1208, Dec. 2024.
- [126] M. N. Islam, M. Hasan, M. K. Hossain, M. G. R. Alam, M. Z. Uddin, and A. Soylu, "Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from CT-radiography," *Scientific Reports*, vol. 12, no. 1, p. 11440, Jul. 2022.
- [127] X.-B. Jin, Z.-Y. Wang, W.-T. Gong, J.-L. Kong, Y.-T. Bai, T.-L. Su, H.-J. Ma, and P. Chakrabarti, "Variational Bayesian Network with Information Interpretability Filtering for Air Quality Forecasting," *Mathematics*, vol. 11, no. 4, p. 837, Jan. 2023.
- [128] G. I. Melsión, I. Torre, E. Vidal, and I. Leite, "Using Explainability to Help Children Understand Gender Bias in AI," in *Proceedings of the 20th Annual ACM Interaction Design and Children Conference*, ser. IDC '21. New York, NY, USA: Association for Computing Machinery, Jun. 2021, pp. 87–99.
- [129] N. L. Tsakiridis, T. Diamantopoulos, A. L. Symeonidis, J. B. Theocharis, A. Iossifides, P. Chatzimisios, G. Pratos, and D. Kouvas, "Versatile Internet of Things for Agriculture: An eXplainable AI Approach," in *Artificial Intelligence Applications and Innovations*, I. Maglogiannis, L. Iliadis, and E. Pimenidis, Eds. Cham: Springer International Publishing, 2020, pp. 180–191.
- [130] J. Wang, Y. Li, Z. Zhou, C. Wang, Y. Hou, L. Zhang, X. Xue, M. Kamp, X. L. Zhang, and S. Chen, "When, Where and How Does it Fail? A Spatial-Temporal Visual Analytics Approach for Interpretable Object Detection in Autonomous Driving," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 12, pp. 5033–5049, Dec. 2023.
- [131] G. Yang, Q. Ye, and J. Xia, "Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond," *Information Fusion*, vol. 77, pp. 29–52, Jan. 2022.
- [132] X. Zhou, K. Jin, Y. Shang, and G. Guo, "Visually Interpretable Representation Learning for Depression Recognition from Facial Images," *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 542–552, Jul. 2020.
- [133] J. Dong, S. Chen, M. Miralinaghi, T. Chen, P. Li, and S. Labi, "Why did the AI make that decision? Towards an explainable artificial intelligence (XAI) for autonomous driving systems," *Transportation Research Part C: Emerging Technologies*, vol. 156, p. 104358, Nov. 2023.
- [134] I. A. Khan, N. Moustafa, I. Razzak, M. Tanveer, D. Pi, Y. Pan, and B. S. Ali, "XSRU-IoMT: Explainable simple recurrent units for threat detection in Internet of Medical Things networks," *Future Generation Computer Systems*, vol. 127, pp. 181–193, Feb. 2022.
- [135] A. K. Mondal, A. Bhattacharjee, P. Singla, and A. P. Prathosh, "xViT-COS: Explainable Vision Transformer Based COVID-19 Screening Using Radiography," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 10, pp. 1–10, 2022.