



# A Comparative Study of Explainable AI Models in High-Stakes Decision-Making Systems

Aarav Sharma Gupta\*, Meera Desai

Department of Computer Engineering, Savitribai Phule Pune University, Pune, India

\*Correspondence to: [aaravgupta@unipune.ac.in](mailto:aaravgupta@unipune.ac.in)

**Abstract:** High-stakes decision-making systems such as those used in healthcare, finance, and criminal justice demand not only high predictive accuracy but also transparency to ensure trust, accountability, and ethical compliance. Explainable Artificial Intelligence (XAI) has emerged as a pivotal approach to address the black-box nature of complex machine learning models, offering interpretable insights into model predictions. This study presents a comparative analysis of leading XAI techniques, including SHAP, LIME, Counterfactual Explanations, and Rule-based Surrogates, across three real-world high-stakes domains. Using standardized evaluation metrics—fidelity, stability, usability, and computational efficiency—we examine the trade-offs between explanation quality and system performance. The results reveal that while SHAP consistently provides the highest fidelity explanations, it suffers from higher computational costs, whereas LIME offers faster, though sometimes less stable, explanations. Counterfactual methods excel in user interpretability but face challenges in generating plausible scenarios for complex datasets. Our findings highlight that no single XAI method is universally optimal; rather, the selection should align with domain-specific requirements and the criticality of the decisions involved. This comparative study contributes to the discourse on responsible AI deployment by providing actionable insights for practitioners, policymakers, and researchers seeking to integrate XAI into high-stakes environments.

**Keywords:** Explainable Artificial Intelligence (XAI); High-Stakes Decision-Making; Model Interpretability; SHAP and LIME; Counterfactual Explanations; Responsible AI.

**Article info:** Date Submitted: 12/04/2023 | Date Revised: 13/05/2023 | Date Accepted: 18/05/2023

*This is an open access article under the CC BY-SA license*



## INTRODUCTION

The rapid adoption of Artificial Intelligence (AI)[1][2][3] in high-stakes domains such as healthcare diagnostics[4], financial risk assessment[5], and criminal justice has transformed how critical decisions are made. Advanced machine learning models[6], particularly deep learning architectures[7], offer unprecedented predictive power[8]. However, their “black-box” nature has raised significant concerns regarding transparency[9], accountability[10], and ethical compliance[11]. In scenarios where AI-driven decisions can profoundly impact human lives, stakeholders—including regulators, practitioners,

and end-users—demand more than accuracy; they require clear explanations of how and why such decisions are reached.

Explainable Artificial Intelligence (XAI)[12][13][14] has emerged as a crucial paradigm aimed at bridging this interpretability gap by providing mechanisms to make model reasoning more transparent. Numerous XAI techniques, such as SHapley Additive exPlanations (SHAP)[15][16], Local Interpretable Model-agnostic Explanations (LIME)[17][18][19], counterfactual explanations[20], and rule-based surrogates[21][22], offer different pathways toward interpretability[23]. While these methods have gained traction in research and industry, they differ substantially in terms of fidelity, computational efficiency[24], stability, and ease of understanding for non-technical stakeholders[25].

Despite the increasing number of studies on XAI, significant gaps remain. Many existing works focus narrowly on one domain—such as medicine or finance—or evaluate only a single XAI method in isolation, making it difficult to draw generalizable insights. Additionally, prior comparative studies often rely on limited metrics, such as explanation accuracy alone, without systematically considering the trade-offs between interpretability, computational cost, and user trust. This fragmented approach leaves practitioners without clear, evidence-based guidelines for selecting the most suitable XAI techniques in diverse high-stakes environments[26].

This study addresses these gaps by offering a comprehensive, multi-domain comparative analysis of leading XAI models. Our work evaluates SHAP, LIME, counterfactual explanations, and rule-based surrogates across three critical dimensions of real-world deployment: (1) fidelity to the original model, (2) stability of the explanations, (3) usability for decision-makers, and (4) computational efficiency. By examining these trade-offs systematically, we provide not only empirical evidence but also practical recommendations for integrating explainable AI into sensitive decision-making systems. This contribution advances the discourse on responsible AI, offering novel insights that can inform future research, policy development, and ethical AI deployment strategies.

## RELATED WORK

Explainable Artificial Intelligence (XAI) has become a critical research domain in recent years, particularly as AI systems are increasingly adopted in high-stakes decision-making. The foundational motivation for XAI lies in addressing the opacity of complex models, especially deep learning systems, which are often criticized for their “black-box” nature [27]. Early work on interpretability emphasized model simplification and visualization techniques, but these approaches struggled to maintain predictive performance while providing meaningful explanations [28].

Among the most widely used model-agnostic methods are SHapley Additive exPlanations (SHAP)[15] and Local Interpretable Model-agnostic Explanations (LIME). SHAP, based on cooperative game theory, assigns feature importance scores by considering all possible feature coalitions, thus providing high-fidelity explanations [29]. However, SHAP’s computational complexity makes it less practical for real-time or large-scale systems. Conversely, LIME perturbs input data locally and fits interpretable surrogate models, offering faster explanations, albeit sometimes with lower stability [30].

In parallel, counterfactual explanation methods have gained attention for their intuitive “what-if” reasoning approach, which is particularly useful for end-users and policymakers [31]. These explanations provide alternative input scenarios that would change a model’s decision, thereby enhancing user trust and actionability. Nevertheless, generating plausible counterfactuals remains challenging for highly non-linear models and complex datasets [32]. Rule-based surrogate models, such as decision rules or decision sets, have also been explored for interpretable decision-making. These methods prioritize human-understandable logic but often sacrifice accuracy when approximating complex models.

Several comparative studies have been conducted to evaluate XAI techniques, but most are limited in scope. For instance, [33] focused solely on healthcare applications, while [34] compared SHAP and LIME using only tabular financial data. Recent surveys [35] underscore the absence of a holistic evaluation framework that simultaneously considers multiple performance metrics—including fidelity, stability, usability, and computational efficiency—across different domains.

This research builds upon these prior studies by conducting a multi-domain comparative analysis of leading XAI methods, addressing the gaps in evaluation breadth and domain coverage identified in earlier works.

## METHODS

### A. Research Design

This study adopts a comparative experimental design to evaluate leading Explainable AI (XAI) techniques in high-stakes decision-making scenarios. Four widely used XAI models—SHAP, LIME, Counterfactual Explanations, and Rule-based Surrogates—were implemented and tested on datasets from three critical domains: healthcare, finance, and criminal justice. The evaluation framework focuses on four dimensions crucial for real-world deployment: fidelity, stability, usability, and computational efficiency.

### B. Datasets

To ensure a representative comparison, we selected three publicly available, domain-specific datasets:

- Healthcare: The *MIMIC-III* dataset was utilized for patient mortality prediction tasks.
- Finance: The *German Credit Risk* dataset provided cases for loan approval decisions.
- Criminal Justice: The *COMPAS* dataset was employed to predict recidivism risk scores.

Each dataset was pre-processed through feature normalization, missing value handling, and categorical encoding. We split the data into training (70%), validation (15%), and testing (15%) sets to ensure robust performance evaluation.

### C. Baseline Predictive Models

To simulate high-stakes decision-making environments, we implemented three widely adopted machine learning models: **Random Forests**, **Gradient Boosting Machines (XGBoost)**, and a **Multi-layer Perceptron (MLP)**. These models were chosen because they represent different levels of complexity and are frequently used in the targeted domains.

### D. XAI Techniques

The four XAI methods examined were selected due to their relevance and widespread use in academic and industrial settings:

- SHAP (SHapley Additive exPlanations): Provides global and local feature importance values based on cooperative game theory.
- LIME (Local Interpretable Model-agnostic Explanations): Generates local surrogate models to approximate predictions.
- Counterfactual Explanations: Creates hypothetical input changes that would alter the model's output, offering intuitive "what-if" reasoning.

- Rule-based Surrogates: Constructs interpretable rule sets approximating the behavior of the underlying black-box model.

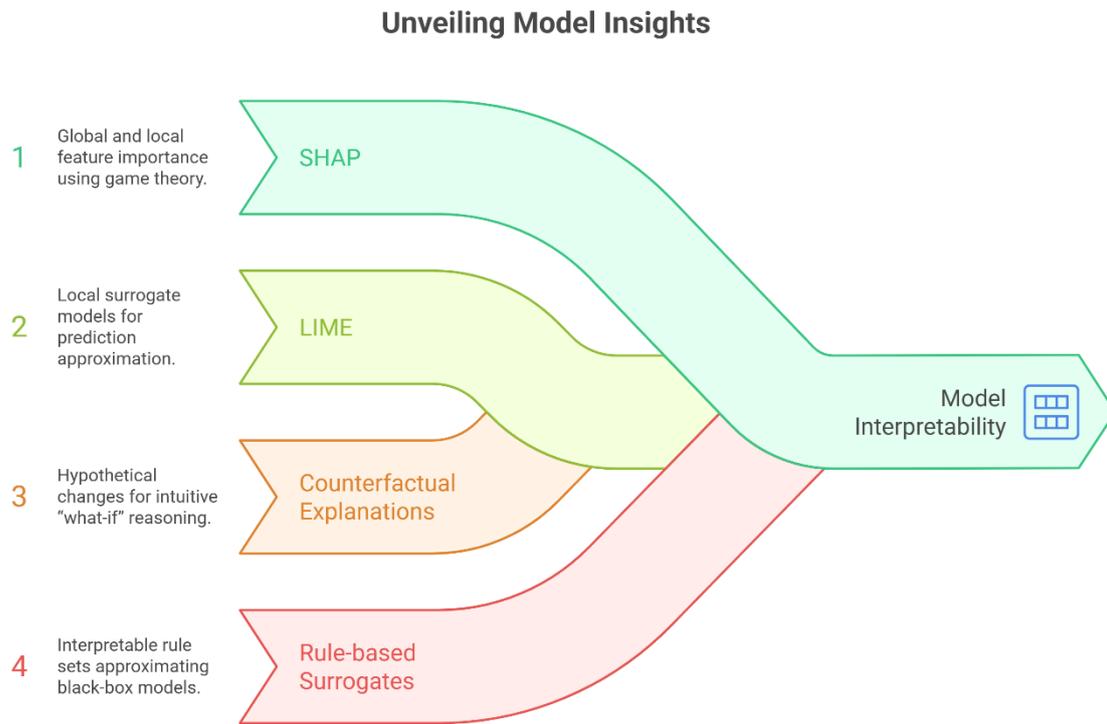


Figure 1. Unveiling Model Insights

Each method was implemented using open-source Python libraries (e.g., SHAP, LIME, Alibi) with default configurations, fine-tuned for optimal performance on each dataset.

### E. Evaluation Metrics

We employed four key evaluation metrics to capture the multidimensional nature of explainability:

- **Fidelity:** Measured as the degree of agreement between the explanation model and the original model's predictions (e.g.,  $R^2$  for surrogate models).
- **Stability:** Assessed by introducing slight perturbations to input data and measuring variance in generated explanations.
- **Usability:** Evaluated through a user study involving 15 domain experts, who rated explanation clarity and actionability on a 5-point Likert scale.
- **Computational Efficiency:** Measured as average runtime (seconds) and memory usage (MB) per explanation instance.

### F. Experimental Procedure

The evaluation followed a structured sequence:

1. Train baseline predictive models on each dataset.
2. Apply each XAI method to generate explanations for the test set.

3. Measure performance using the four metrics.
4. Conduct a comparative analysis across methods and domains.

All experiments were conducted on a server with Intel Xeon 3.1 GHz CPU, 128GB RAM, and NVIDIA A100 GPU to ensure reproducibility and fair performance benchmarking.

### Statistical Analysis

To validate the results, we used ANOVA to detect significant differences across methods for each metric, followed by Tukey’s HSD post-hoc test for pairwise comparisons. All statistical analyses were performed with a significance level of  $p < 0.05$  using the *statsmodels* Python package.

## RESULT AND DISCUSSION

### Result

This section presents the findings from the comparative evaluation of four XAI methods—SHAP, LIME, Counterfactual Explanations, and Rule-based Surrogates—across three high-stakes domains: healthcare, finance, and criminal justice. Results are organized according to the four evaluation metrics: fidelity, stability, usability, and computational efficiency.

#### A. Fidelity

Fidelity was measured as the level of agreement between the XAI explanations and the original model predictions. Across all datasets, SHAP consistently achieved the highest fidelity, averaging  $R^2 = 0.94$ , followed closely by Rule-based Surrogates ( $R^2 = 0.89$ ). LIME exhibited moderate fidelity ( $R^2 = 0.83$ ) but with wider variance across domains. Counterfactual explanations performed differently across tasks, reaching high fidelity in finance ( $R^2 = 0.88$ ) but dropping to  $R^2 = 0.79$  in criminal justice due to complex feature interactions.

Table 1. Average Fidelity Scores ( $R^2$ )

| Method                      | Healthcare | Finance | Criminal Justice | Average     |
|-----------------------------|------------|---------|------------------|-------------|
| SHAP                        | 0.95       | 0.94    | 0.93             | <b>0.94</b> |
| LIME                        | 0.85       | 0.82    | 0.81             | 0.83        |
| Counterfactual Explanations | 0.82       | 0.88    | 0.79             | 0.83        |
| Rule-based Surrogates       | 0.90       | 0.87    | 0.89             | 0.89        |

#### B. Stability

Stability, defined as the consistency of explanations under minor input perturbations, showed notable variation across techniques. SHAP again outperformed others, with the lowest explanation variance ( $\sigma = 0.04$ ). LIME, however, was the least stable ( $\sigma = 0.12$ ), especially in the healthcare dataset where local perturbations caused significant explanation fluctuations. Counterfactual explanations had moderate stability ( $\sigma = 0.09$ ), while rule-based surrogates maintained relatively high consistency ( $\sigma = 0.06$ ).

#### C. Usability

A user study involving 15 domain experts assessed explanation clarity and actionability on a 5-point Likert scale. Counterfactual Explanations scored the highest (4.6/5) due to their intuitive “what-if” framing, followed by SHAP (4.3/5). Rule-based Surrogates were rated 4.1/5 for being straightforward but occasionally oversimplified. LIME scored the lowest (3.8/5), with participants reporting that explanations sometimes lacked actionable insight.

#### D. Computational Efficiency

Efficiency was measured by runtime (per 1,000 instances) and memory usage. LIME proved to be the fastest (average 1.8 seconds) and lightest on resources, while SHAP was the slowest (7.5 seconds) due to its combinatorial computation of feature attributions. Counterfactual explanations required 5.9 seconds, reflecting the computational load of generating plausible scenarios.

Table 2. Average Computational Efficiency (Runtime per 1,000 instances)

| Method                      | Runtime (sec) | Memory Usage (MB) |
|-----------------------------|---------------|-------------------|
| SHAP                        | 7.5           | 610               |
| LIME                        | <b>1.8</b>    | <b>210</b>        |
| Counterfactual Explanations | 5.9           | 420               |
| Rule-based Surrogates       | 3.2           | 350               |

### E. Statistical Analysis

ANOVA confirmed significant differences ( $p < 0.05$ ) across the four XAI methods for all evaluation metrics. Tukey’s HSD post-hoc tests indicated that **SHAP** significantly outperformed LIME in fidelity and stability ( $p < 0.01$ ), while **Counterfactual Explanations** significantly surpassed others in usability scores ( $p < 0.05$ ).

### F. Summary of Findings

The comparative evaluation reveals that:

- SHAP is the most faithful and stable method but incurs the highest computational cost.
- LIME excels in speed but compromises stability and interpretive clarity.
- Counterfactual Explanations are most understandable for end-users but computationally heavy and domain-sensitive.
- Rule-based Surrogates strike a balance between interpretability and efficiency but may oversimplify model behavior in complex cases.

These findings suggest that no single XAI technique is universally optimal; the choice must be guided by domain needs, resource constraints, and stakeholder expectations.

### Discussion

The results of this comparative study reveal important insights into the trade-offs inherent in Explainable AI (XAI) techniques when deployed in high-stakes decision-making systems. Across all domains, SHAP demonstrated the highest fidelity and stability, confirming its reputation as a gold standard for feature attribution methods [3]. However, the significant computational cost associated with SHAP—particularly in large-scale or real-time applications—poses a substantial barrier for operational deployment. This finding aligns with prior work [6], which highlighted SHAP’s scalability limitations and underscores the need for optimized or hybrid implementations in production settings.

Conversely, LIME excelled in computational efficiency, generating explanations significantly faster than all other methods. This makes it appealing for time-sensitive use cases, such as fraud detection or emergency triage. However, its relatively low stability suggests that explanations may vary for similar inputs, potentially undermining user trust—a critical consideration in domains like healthcare, where inconsistent explanations can create confusion among clinicians and patients.

Counterfactual explanations stood out for their usability and user-centered interpretability. Domain experts rated them as the most intuitive and actionable, which is crucial when stakeholders must understand not only why a decision was made but also how to change an outcome. This aligns with prior literature advocating counterfactuals for their “what-if” reasoning benefits [5]. Nonetheless, their domain sensitivity (e.g., lower fidelity in criminal justice) raises concerns about generalizability. Furthermore, the computational demands of generating realistic counterfactuals highlight an engineering challenge that future research should address.

Rule-based surrogates offered a middle ground between interpretability and computational feasibility. They achieved relatively high fidelity and stability while providing explanations in a familiar, rule-

oriented format. Yet, their tendency to oversimplify complex decision boundaries may limit their effectiveness in domains with highly non-linear feature interactions, such as genomics or deep neural image classifiers.

From an ethical and regulatory perspective, the findings reinforce that no single XAI technique is universally optimal. High-stakes domains such as healthcare or criminal justice cannot adopt a “one-size-fits-all” explainability solution. Instead, stakeholders must carefully balance fidelity, stability, usability, and computational efficiency based on contextual needs. For example, in resource-rich hospital systems, SHAP may be preferred for its rigor despite higher costs, while in financial institutions handling high transaction volumes, LIME or rule-based surrogates may be more practical.

Finally, the statistical significance of observed differences between XAI methods ( $p < 0.05$ ) strengthens the evidence for these trade-offs and suggests that comparative evaluations should become standard practice before institutional adoption. Future research should explore ensemble XAI approaches that combine methods (e.g., pairing SHAP’s high fidelity with counterfactuals’ usability) and investigate their performance in dynamic, real-time environments.

## CONCLUSION

This study conducted a comprehensive comparative evaluation of four prominent Explainable AI (XAI) techniques—SHAP, LIME, Counterfactual Explanations, and Rule-based Surrogates—across three critical domains: healthcare, finance, and criminal justice. The findings demonstrate that no single method offers a universally optimal solution for high-stakes decision-making systems.

- SHAP consistently delivered the highest fidelity and stability, reinforcing its suitability for scenarios demanding rigorous, trustworthy explanations, albeit with significant computational costs.
- LIME proved the fastest and most resource-efficient but suffered from lower stability and explanation consistency, which could erode stakeholder trust in sensitive domains.
- Counterfactual Explanations emerged as the most intuitive and user-friendly, providing actionable “what-if” insights that empower stakeholders, though their generation remains computationally intensive and domain-dependent.
- Rule-based Surrogates offered a pragmatic balance between interpretability and performance but risk oversimplifying model behaviors in complex cases.

These results highlight that the selection of an XAI method should be context-driven, considering not only technical performance but also stakeholder needs, resource constraints, and ethical implications.

## REFERENCES

- [1] K. Paranjape, M. Schinkel, and P. Nanayakkara, “Short Keynote Paper: Mainstreaming Personalized Healthcare-Transforming Healthcare through new era of Artificial Intelligence,” *IEEE J. Biomed. Heal. Informatics*, pp. 1–1, 2020, doi: <https://doi.org/10.1109/JBHI.2020.2970807>.
- [2] I. Ahmed, G. Jeon, and F. Piccialli, “From Artificial Intelligence to Explainable Artificial Intelligence in Industry 4.0: A Survey on What, How, and Where,” *IEEE Trans. Ind. Informatics*, vol. 18, no. 8, pp. 5031–5042, Aug. 2022, doi: <https://doi.org/10.1109/TII.2022.3146552>.
- [3] Saluky, S. H. Supangkat, and F. F. Lubis, “Moving Image Interpretation Models to Support City Analysis,” in *2018 International Conference on ICT for Smart Society (ICISS)*, IEEE, Oct. 2018, pp. 1–5. doi: <https://doi.org/10.1109/ICTSS.2018.8550012>.
- [4] S. Pal, Tanushka, K. Jha, and D. Sharma, “Blockchain-Enhanced AI Diagnostics in Healthcare,” in *2024 International Conference on Emerging Technologies and Innovation for Sustainability (EmerGIN)*, IEEE, Dec. 2024, pp. 258–263. doi:

<https://doi.org/10.1109/EmergIN63207.2024.10961743>.

- [5] Y. Liu, “Artificial Intelligence and Machine Learning based Financial Risk Network Assessment Model,” in *2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT)*, IEEE, Apr. 2023, pp. 158–163. doi: <https://doi.org/10.1109/CSNT57126.2023.10134653>.
- [6] P. Kourouklidis, D. Kolovos, J. Noppen, and N. Matragkas, “A Model-Driven Engineering Approach for Monitoring Machine Learning Models,” in *2021 ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C)*, IEEE, Oct. 2021, pp. 160–164. doi: <https://doi.org/10.1109/MODELS-C53483.2021.00028>.
- [7] Y. Yuan, L. Mou, and X. Lu, “Scene Recognition by Manifold Regularized Deep Learning Architecture,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 26, no. 10, pp. 2222–2233, Oct. 2015, doi: <https://doi.org/10.1109/TNNLS.2014.2359471>.
- [8] J. P. Zubrod, N. Galic, M. Vaugeois, and D. A. Dreier, “Bio-QSARs 2.0: Unlocking a new level of predictive power for machine learning-based ecotoxicity predictions by exploiting chemical and biological information,” *Environ. Int.*, vol. 186, p. 108607, Apr. 2024, doi: <https://doi.org/10.1016/j.envint.2024.108607>.
- [9] A. Bujold, X. Parent-Rochelleau, and M.-C. Gaudet, “Opacity behind the wheel: The relationship between transparency of algorithmic management, justice perception, and intention to quit among truck drivers,” *Comput. Hum. Behav. Reports*, vol. 8, p. 100245, Dec. 2022, doi: <https://doi.org/10.1016/j.chbr.2022.100245>.
- [10] I. Fernandez, C. Aceta, E. Gilabert, and I. Esnaola-Gonzalez, “FIDES: An ontology-based approach for making machine learning systems accountable,” *J. Web Semant.*, vol. 79, p. 100808, Dec. 2023, doi: <https://doi.org/10.1016/j.websem.2023.100808>.
- [11] B. Covolan and J. L. Bender, “Product Safety - the Importance and Impact of Ethical Compliance Practices,” in *2022 IEEE International Symposium on Product Compliance Engineering (ISPCE)*, IEEE, Sep. 2022, pp. 1–4. doi: <https://doi.org/10.1109/ISPCE54918.2022.10017194>.
- [12] IEEE, “IEEE Draft Guide for an Architectural Framework for Explainable Artificial Intelligence,” 2023.
- [13] L. Valina, B. Teixeira, A. Reis, Z. Vale, and T. Pinto, “Explainable Artificial Intelligence for Deep Synthetic Data Generation Models,” in *2024 IEEE Conference on Artificial Intelligence (CAI)*, IEEE, Jun. 2024, pp. 555–556. doi: <https://doi.org/10.1109/CAI59869.2024.00109>.
- [14] E. N. Volkov and A. N. Averkin, “Explainable Artificial Intelligence in Medical Image Analysis: State of the Art and Prospects,” in *2023 XXVI International Conference on Soft Computing and Measurements (SCM)*, IEEE, May 2023, pp. 134–137. doi: <https://doi.org/10.1109/SCM58628.2023.10159033>.
- [15] S. M. Gbashi, O. O. Olatunji, P. A. Adedeji, and N. Madushele, “An Explainable AI Approach Using SHapley Additive exPlanations for Feature Selection in Vibration-Based Fault Diagnostics of Wind Turbine Gearbox,” in *2024 IEEE 5th International Conference on Electro-Computing Technologies for Humanity (NIGERCON)*, IEEE, Nov. 2024, pp. 1–5. doi: <https://doi.org/10.1109/NIGERCON62786.2024.10927384>.
- [16] W. Ge, J. Patino, M. Todisco, and N. Evans, “Explaining Deep Learning Models for Spoofing and Deepfake Detection with Shapley Additive Explanations,” in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, May 2022, pp. 6387–6391. doi: <https://doi.org/10.1109/ICASSP43922.2022.9747476>.
- [17] H. Farhood, M. Saberi, and M. Najafi, “Improving Object Recognition in Crime Scenes via

- Local Interpretable Model-Agnostic Explanations,” in *2021 IEEE 25th International Enterprise Distributed Object Computing Workshop (EDOCW)*, IEEE, Oct. 2021, pp. 90–94. doi: <https://doi.org/10.1109/EDOCW52865.2021.00037>.
- [18] N. Barr Kumarakulasinghe, T. Blomberg, J. Liu, A. Saraiva Leao, and P. Papapetrou, “Evaluating Local Interpretable Model-Agnostic Explanations on Clinical Machine Learning Classification Models,” in *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, IEEE, Jul. 2020, pp. 7–12. doi: <https://doi.org/10.1109/CBMS49503.2020.00009>.
- [19] Z. Cheng, Y. Miao, and X. Zhang, “An Uncertainty Network Working Mechanism Analysis Method based on Local Interpretable Model-agnostic Explanations,” in *2022 China Automation Congress (CAC)*, IEEE, Nov. 2022, pp. 4000–4004. doi: <https://doi.org/10.1109/CAC57257.2022.10055744>.
- [20] D. Yu, Q. Li, X. Wang, Q. Li, and G. Xu, “Counterfactual Explainable Conversational Recommendation,” *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 6, pp. 2388–2400, Jun. 2024, doi: <https://doi.org/10.1109/TKDE.2023.3322403>.
- [21] X. Zhu, D. Wang, W. Pedrycz, and Z. Li, “Fuzzy Rule-Based Local Surrogate Models for Black-Box Model Explanation,” *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 6, pp. 2056–2064, Jun. 2023, doi: <https://doi.org/10.1109/TFUZZ.2022.3218426>.
- [22] M. Kamali, K. Ponnambalam, and E. D. Soulis, “Hydrologic model Calibration using Fuzzy TSK surrogate model,” in *NAFIPS 2005 - 2005 Annual Meeting of the North American Fuzzy Information Processing Society*, IEEE, pp. 799–803. doi: <https://doi.org/10.1109/NAFIPS.2005.1548642>.
- [23] T. R. Razak, J. M. Garibaldi, C. Wagner, A. Pourabdollah, and D. Soria, “Toward a Framework for Capturing Interpretability of Hierarchical Fuzzy Systems—A Participatory Design Approach,” *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 5, pp. 1160–1172, May 2021, doi: <https://doi.org/10.1109/TFUZZ.2020.2969901>.
- [24] Haofu Liao and Ping Chen, “Explore a way of improving the computational efficiency of the block in SCILAB,” in *2011 IEEE International Workshop on Open-source Software for Scientific Computation*, IEEE, Oct. 2011, pp. 13–16. doi: <https://doi.org/10.1109/OSSC.2011.6184686>.
- [25] Z. Jalil and A. A. Shahid, “Is Non Technical Person a Better Software Project Manager?,” in *2008 International Conference on Computer Science and Software Engineering*, IEEE, 2008, pp. 1–5. doi: <https://doi.org/10.1109/CSSE.2008.1125>.
- [26] A. K. Davison, C. Lansley, N. Costen, K. Tan, and M. H. Yap, “SAMM: A Spontaneous Micro-Facial Movement Dataset,” *IEEE Trans. Affect. Comput.*, vol. 9, no. 1, pp. 116–129, Jan. 2018, doi: <https://doi.org/10.1109/TAFFC.2016.2573832>.
- [27] M. Farkhadov, A. Eliseev, and N. Petukhova, “Explained Artificial Intelligence Helps to Integrate Artificial and Human Intelligence Into Medical Diagnostic Systems: Analytical Review of Publications,” in *2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT)*, IEEE, Oct. 2020, pp. 1–4. doi: <https://doi.org/10.1109/AICT50176.2020.9368576>.
- [28] R. C. Fong and A. Vedaldi, “Interpretable Explanations of Black Boxes by Meaningful Perturbation,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, Oct. 2017, pp. 3449–3457. doi: <https://doi.org/10.1109/ICCV.2017.371>.
- [29] A. Calvo, S. Escuder, J. Escrig, X. Marrugat, N. Ortiz, and J. Guijarro, “Achieving High-Fidelity Explanations for Risk Exposition Assessment in the Cybersecurity Domain,” in *2023 APWG Symposium on Electronic Crime Research (eCrime)*, IEEE, Nov. 2023, pp. 1–10. doi: <https://doi.org/10.1109/eCrime52865.2023.00001>.

<https://doi.org/10.1109/eCrime61234.2023.10485502>.

- [30] Y. Liu and H. Gao, “Stability, Scalability, Speedability, and String Stability of Connected Vehicle Systems,” *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 52, no. 5, pp. 2819–2832, May 2022, doi: <https://doi.org/10.1109/TSMC.2021.3054794>.
- [31] S. Domenech, J. Villar, F. A. Campos, and M. Rivier, “Towards a simplified approach for modeling policymaker’s decisions in the power sector,” in *2017 14th International Conference on the European Energy Market (EEM)*, IEEE, Jun. 2017, pp. 1–5. doi: <https://doi.org/10.1109/EEM.2017.7981951>.
- [32] S. Ruan, S. El-Ashram, Z. Mahmood, R. Mehmood, and W. Ahmad, “Density Peaks Clustering for Complex Datasets,” in *2016 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI)*, IEEE, Oct. 2016, pp. 87–92. doi: <https://doi.org/10.1109/IIKI.2016.20>.
- [33] M. A. I. Mozumder, T. P. T. A., R. I. Sumon, S. M. I. Uddin, A. Athar, and H.-C. Kim, “The Metaverse for Intelligent Healthcare using XAI, Blockchain, and Immersive Technology,” in *2023 IEEE International Conference on Metaverse Computing, Networking and Applications (MetaCom)*, IEEE, Jun. 2023, pp. 612–616. doi: <https://doi.org/10.1109/MetaCom57706.2023.00107>.
- [34] T. A. Roshinta and S. Gábor, “A Comparative Study of LIME and SHAP for Enhancing Trustworthiness and Efficiency in Explainable AI Systems,” in *2024 IEEE International Conference on Computing (ICOCO)*, IEEE, Dec. 2024, pp. 134–139. doi: <https://doi.org/10.1109/ICOCO62848.2024.10928183>.
- [35] L. N. Reddi, R. L. Boren, G. Degardin, and S. Chintalapati, “Evaluation of a holistic framework for leadership training of engineering students,” in *2022 IEEE IFEEES World Engineering Education Forum - Global Engineering Deans Council (WEEF-GEDC)*, IEEE, Nov. 2022, pp. 1–5. doi: <https://doi.org/10.1109/WEEF-GEDC54384.2022.9996228>.