

Peer Review

Review of: "Explaining the Unexplainable: A Systematic Review of Explainable AI in Finance"

Udo Schlegel¹

1. Universität Konstanz, Germany

This work seeks to deliver a systematic review of Explainable Artificial Intelligence (XAI) within the financial domain. The authors aim to clarify how XAI techniques are applied to support transparency and trust in finance, a sector heavily influenced by regulatory, ethical, and operational concerns. With a stated review of 323 articles and a narrowing down to 30 for deeper analysis, the paper aspires to summarize state-of-the-art developments, current challenges, and future research avenues.

Strengths

- **Topic Relevance:** The subject of XAI in finance is of timely significance, aligning well with both academic inquiry and real-world financial practice. This makes the review highly relevant for a wide audience.
- **Literature Scope:** The initial review of 323 articles from a pool of 5,763 indicates an ambitious attempt to comprehensively survey the field.

Weaknesses

- **Selection Criteria:** The article lacks a transparent explanation of the filtering process from 5,763 papers down to 30. The review would benefit from a detailed account of the inclusion/exclusion criteria, such as publication quality, research design, or relevance to finance-specific applications of XAI.
- **Review Depth:** While Table 1 lists selected papers, the discussion does not adequately analyze their contributions. A critical comparative discussion of these studies is needed—what are their strengths, limitations, and contributions to understanding XAI in finance? How does XAI meaningfully enhance insights in each case?

- **Challenges and Research Directions:** The discussion of challenges facing XAI in finance is brief and underdeveloped. A more thorough exploration—drawing from themes such as regulatory complexity, model interpretability vs. performance trade-offs, and stakeholder-specific explainability needs—would significantly strengthen the paper. Furthermore, a clearer articulation of future research opportunities grounded in the systematic review would be beneficial.

Technical and Editorial Comments

- **Figures:** The figures in the paper would benefit from longer, more explanatory captions that help readers understand the relevance and content at a glance.
- **Figure 2 Misclassification:** The categorization of Integrated Gradients as a post-hoc model-agnostic method is incorrect. Integrated Gradients is a post-hoc technique, but it is model-specific, relying on differentiability and internal gradients—hence applicable only to certain neural architectures.

Missing Literature

The paper would be more robust if it included foundational and recent references in the field of XAI. Notably absent are:

- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). *A survey of methods for explaining black box models*. ACM Computing Surveys (CSUR), 51(5), 1-42.
- Theissler, A., Spinnato, F., Schlegel, U., & Guidotti, R. (2022). *Explainable AI for time series classification: a review, taxonomy and research directions*. IEEE Access, 10, 100700-100724.

These works provide foundational taxonomies and current insights, especially in black-box explanation methods and time-series classification, which are highly relevant for financial applications.

Conclusion

The article contributes to the ongoing conversation about the role of XAI in finance, offering a useful starting point for further exploration. However, its impact would be significantly enhanced by addressing the methodological opacity, expanding the depth of analysis of reviewed works, correcting conceptual inaccuracies, and incorporating essential literature. With these improvements, the paper could serve as a more authoritative guide to researchers and practitioners alike.

Declarations

Potential competing interests: No potential competing interests to declare.