Alessio
davide
chiara
Fatto

Reviewer #1: Manuscript Number: EAAI-24-5135
Title: Interpretability and Generalizability in Extended Isolation Forest Models with Extended Isolation Forest Feature Importance and Enhanced Extended Isolation Forest
This paper focuses on enhancing interpretability in Anomaly Detection (AD) systems, particularly in engineering contexts where understanding the root cause of anomalies is essential. The paper introduces two key contributions:
1. Extended Isolation Forest Feature Importance (ExIFFI): A novel approach designed to explain the predictions made by the Extended Isolation Forest (EIF) model. ExIFFI provides feature importance insights at both global and local levels, helping users understand why certain anomalies are flagged. This is crucial for fostering trust in the model and facilitating root cause analysis.
2. Enhanced Extended Isolation Forest (EIF+): An improved version of EIF that aims to enhance generalization capabilities through a refined splitting hyperplane design strategy. This new variant is designed to perform better across different datasets.

Comment 1: Some sentences are complex and could be simplified for better readability. For example, "Moreover, the end users of these systems are becoming more diverse, including people from various backgrounds who may not have a knowledge in data-driven methods" could be rephrased as "Moreover, the end users of these systems are increasingly diverse, often lacking expertise in data-driven methods." -> Rewritten some parts still can be improved
Comment 2: The sentence structure could be streamlined to improve the flow. For example, "Despite the remarkable recent advancements in eXplainable Artificial Intelligence (XAI), most approaches are designed for supervised tasks, leaving unsupervised tasks, like Anomaly Detection (AD), rarely discussed in the literature" could be split into two sentences for clarity.-> Done the specific sentence.. rewritten introduction and past works
Comment 3: Before introducing ECOD, a sentence that links the discussion from XAD models to ECOD could help readers follow the narrative more easily.
Comment 4: The citation "[16]" should match the format of others like "[17]" or "[18]." -> chiedi Gian
Comment 5: Figure 1a and 1b) should be well-labeled and referenced clearly in the text. Consider adding captions that briefly describe what each figure illustrates. -> DONE
Comment 6: "Induced Imbalance Coefficients," "Cumulative feature importance" briefly define these terms when first introduced or provide references for readers who may need additional context. -> chiedi Gian perched ne parliamo di già… dobbiamo estendere il discorso? Copiaincollapassato
Comment 7: In Section 4 authors are suggested to write benefit from a brief mention of why interpretability is crucial in the context of anomaly detection models. This would set the stage for the discussion that follows. DONE
Comment 8: In Section 4 consider summarizing the main takeaway of DIFFI in simpler terms at the end of the section to reinforce understanding. DONE
Comment 9: In Section 5 the organization of content could be improved. Separating the description of datasets from the performance and interpretability evaluations could enhance clarity. -> DONE
Comment 10: Justify why Scenario II (0 contamination of outliers) is particularly relevant in certain AD model applications could provide more context for the reader.-> DONE
Comment 11: Consider adding a brief explanation of why each dataset was chosen and how its characteristics align with the goals of the evaluation. -> DAVIDE
Comment 12: Providing specific examples or data points that illustrate the computational burden of SHAP would strengthen the argument. -> DAVIDE
Comment 13: The comparison with other models is thorough, the discussion could benefit from deeper insights into why certain models like DIF and AutoEncoder perform poorly in specific scenarios. Providing a more detailed explanation of the underlying reasons for these performance discrepancies could add value to the analysis.
Comment 14: The significance of feature 0 in different scenarios is noted, but discussing how this impacts

the practical application of EIF+ and ExIFFI in real-world scenarios could provide more context.-> WTF
Comment 15: The AI finds its applications in various domains. Authors are suggested to cite recent articles such as Green Fault Tolerant AIoT-enabled Mobile Sink Data Collection Scheme in Sensor Networks, Fault Diagnosis for AIT based Smart Farming Applications, AI Revolutionizing Healthcare: Innovations, Challenges, and Ethical Considerations, Unlocking the power of industrial artificial intelligence towards Industry 5.0:Insights, pathways, and challenges.-> DAVIDE

Reviewer #2: The paper focuses on unsupervised anomaly detection, proposing ExIFFI for interpretability and EIF+ for improved generalization of the EIF model. It compares their performance using synthetic and real-world datasets, demonstrating the effectiveness of ExIFFI and the superiority of EIF+. A metric is introduced for quantitative evaluation, and future research directions are suggested, including exploring information in splitting nodes and deploying the methods in industrial contexts.

The introduction of ExIFFI and EIF+ is a noteworthy contribution, enhancing interpretability and generalization in unsupervised anomaly detection. Comprehensive comparisons using diverse datasets offer robust evidence of the models' performance and effectiveness. The AUCFS metric provides a quantitative means to assess the effectiveness of interpretation algorithms, adding objectivity to the evaluation.

However, the article can be further strengthened, so I propose the following comments:

1. When describing the ExIFFI algorithm, please use example data and intermediate calculation results to help readers better understand how the algorithm calculates feature importance. -> DONE
2. Additionally, consider providing pseudocode to make the algorithm steps clearer. -> chiedi chiara come migliorare
3. When explaining the improvements of EIF+ compared to EIF, please quantify and compare the performance differences between the two on several representative datasets, and conduct an in-depth analysis of the fundamental reasons for this difference (i.e., the difference in the hyperplane selection strategy). -> chiedi chiara violini
4. For the generation process of the synthetic datasets, it is best to provide specific data generation functions or algorithm pseudocode, rather than just a simple description. If the dataset involves noise or other complex situations, they also need to be clearly explained.
5. During the experimental evaluation, consider adding some of the latest anomaly detection baseline models for comparison, rather than just limiting it to classic models such as IF, to demonstrate the superiority of the proposed method. -> aggiungeremo ecod
6. In the interpretability evaluation, in addition to using feature selection as a proxy task, artificial datasets can also be designed, and based on the known ground-truth, a quantitative score can be given to the explanations provided by ExIFFI to more directly evaluate the quality of its explanations.
7. For the correlation analysis between the explanations of ExIFFI and the outliers, it can be specifically examined whether the samples with high feature importance scores also correspond to the higher anomaly scores given by the algorithm, and whether there is a clear positive correlation between the two.
8. In addition to analyzing the differences of the model under different data distributions and anomaly types, the differences in the explanation quality of ExIFFI can also be analyzed, such as in which cases the model's explanation is more reliable and more in line with human cognition.

Reviewer #3: This is a relevant article. However, it is still being prepared to be considered as a scientific paper. Below are some key observations:
1. The title and content could be simpler for anyone to read and follow. -> chiedi chiara come accontentarlo
2. The objectives and the focus of the study need to be clarified.

3. The authors spend most of their time describing what they have done without highlighting why it is unique and how it will benefit others. Unless these two questions are answered upfront, I am not sure why someone would read this work.
4. Too lengthy -> chiedi chiara
5. There are too many details [Figures and Tables] that need to highlight their purpose. It reads more like a part of a thesis or a report.
6. The topic is about relevant and if work is presented well may have an impact.
7. The authors should expand their review of relevant work and present it succinctly. Many researchers have widely used advanced data-driven techniques, such as the Bayesian Network-based methods; researchers from C-RISE, Memorial University, Canada, have proposed scores of fault diagnosis and isolation models and methods that are extended to complex engineering system safety and risk analysis. The authors have also ignored these studies.
8. The authors should attempt to benchmark their study with other published work.

Reviewer #4: Include all important numerical results in the abstract.
Better highlight new contributions and novelty of your work in the introduction section.
Enhance linkage to recent applied literature that demonstrates the great potential and usefulness of different machine learning models for forecasting purposes (doi: 10.1108/AJEB-01-2024-0007; 10.1016/j.meaene.2024.100001; 10.1007/s00521-024-09531-2; 10.1108/JM2-12-2023-0315; 10.1177/03019233241254891) to better motivate your present investigation.
Provide a summary of previous studies in terms of what has been done, what is still missing, and correspondingly your contributions.
Improve the quality of the figures to make them clearly visualized by using high-resolution formats.
Improve the organization of the tables to make them clearly laid out. Add more explanations to the figures and tables to make them self-explanatory.
Offer more detailed discussions of the results by linking to the figures and tables.
Use more mathematical formulae and equations to demonstrate your proposed approach. -> matematicizziamo l'aucfs score, synthetic datasets
Conduct further benchmark analysis to demonstrate the advantage or disadvantage of your method.
Elaborate on the potential limitations of your work, and correspondingly, the path to future studies in the conclusion.
More carefully proofread your manuscript to avoid writing issues.

- Scrittura:
  - Troppo generali
    - The title and content could be simpler for anyone to read and follow.
  - Generali
    - Some sentences are complex and could be simplified for better readability.
    - The sentence structure could be streamlined to improve the flow.
    - The objectives and the focus of the study need to be clarified.
    - The authors spend most of their time describing what they have done without highlighting why it is unique and how it will benefit others. Unless these two questions are answered upfront, I am not sure why someone would read this work.
    - Too lengthy
    - There are too many details [Figures and Tables] that need to highlight their purpose. It reads more like a part of a thesis or a report.
    - Use more mathematical formulae and equations to demonstrate your proposed approach.
    - More carefully proofread your manuscript to avoid writing issues.
    - Elaborate on the potential limitations of your work, and correspondingly, the path to future studies in the conclusion.
  - Specifiche
    - Better highlight new contributions and novelty of your work in the introduction section.

- Before introducing ECOD, a sentence that links the discussion from XAD models to ECOD could help readers follow the narrative more easily.
- The citation "[16]" should match the format of others like "[17]" or "[18].”
- Figure 1a and 1b) should be well-labeled and referenced clearly in the text. Consider adding captions that briefly describe what each figure illustrates.
- "Induced Imbalance Coefficients," "Cumulative feature importance" briefly define these terms when first introduced or provide references for readers who may need additional context.
- When describing the ExIFFI algorithm, please use example data and intermediate calculation results to help readers better understand how the algorithm calculates feature importance.
- Additionally, consider providing pseudocode to make the algorithm steps clearer.
- For the generation process of the synthetic datasets, it is best to provide specific data generation functions or algorithm pseudocode, rather than just a simple description.
- If the dataset involves noise or other complex situations, they also need to be clearly explained.
- Improve the quality of the figures to make them clearly visualized by using high-resolution formats. Improve the organization of the tables to make them clearly laid out. Add more explanations to the figures and tables to make them self-explanatory.

- Contenuto
  - Conduct further benchmark analysis to demonstrate the advantage or disadvantage of your method.
  - In Section 4 authors are suggested to write benefit from a brief mention of why interpretability is crucial in the context of anomaly detection models. This would set the stage for the discussion that follows.
  - In Section 4 consider summarizing the main takeaway of DIFFI in simpler terms at the end of the section to reinforce understanding.
  - In Section 5 the organization of content could be improved. Separating the description of datasets from the performance and interpretability evaluations could enhance clarity.
  - Justify why Scenario II (0 contamination of outliers) is particularly relevant in certain AD model applications could provide more context for the reader.
  - Consider adding a brief explanation of why each dataset was chosen and how its characteristics align with the goals of the evaluation.
  - The comparison with other models is thorough, the discussion could benefit from deeper insights into why certain models like DIF and AutoEncoder perform poorly in specific scenarios. Providing a more detailed explanation of the underlying reasons for these performance discrepancies could add value to the analysis.
  - The significance of feature 0 in different scenarios is noted, but discussing how this impacts the practical application of EIF+ and ExIFFI in real-world scenarios could provide more context.
- Citazioni
  - Provide a summary of previous studies in terms of what has been done, what is still missing, and correspondingly your contributions.
  - The AI finds its applications in various domains. Authors are suggested to cite recent articles such as Green Fault Tolerant AIoT-enabled Mobile Sink Data Collection Scheme in Sensor Networks, Fault Diagnosis for AIT based Smart Farming Applications, AI Revolutionizing Healthcare: Innovations, Challenges, and Ethical Considerations, Unlocking the power of industrial artificial intelligence towards Industry 5.0:Insights, pathways, and challenges.
  - The authors should expand their review of relevant work and present it succinctly. Many researchers have widely used advanced data-driven techniques, such as the Bayesian Network-based methods; researchers from C-RISE, Memorial University, Canada, have proposed scores of fault diagnosis and isolation models and methods that are extended to complex engineering system safety and risk analysis. The authors have also ignored these studies.
  - Enhance linkage to recent applied literature that demonstrates the great potential and usefulness of different machine learning models for forecasting purposes (doi: 10.1108/AJEB-01-2024-0007; 10.1016/j.meaene.2024.100001; 10.1007/s00521-024-09531-2; 10.1108/JM2-12-2023-0315; 10.1177/03019233241254891) to better motivate your present investigation.

- Cose che mancano?
  - Non mancano secondo me

- Conduct further benchmark analysis to demonstrate the advantage or disadvantage of your method.
- During the experimental evaluation, consider adding some of the latest anomaly detection baseline models for comparison, rather than just limiting it to classic models such as IF, to demonstrate the superiority of the proposed method. (DIF Autoencoder)
- In the interpretability evaluation, in addition to using feature selection as a proxy task, artificial datasets can also be designed, and based on the known ground-truth, a quantitative score can be given to the explanations provided by ExIFFI to more directly evaluate the quality of its explanations. (Taxis bisect etc)
- In addition to analyzing the differences of the model under different data distributions and anomaly types, the differences in the explanation quality of ExIFFI can also be analyzed, such as in which cases the model's explanation is more reliable and more in line with human cognition. (Glass dataset analysis)

- Mancano
  - For the correlation analysis between the explanations of ExIFFI and the outliers, it can be specifically examined whether the samples with high feature importance scores also correspond to the higher anomaly scores given by the algorithm, and whether there is a clear positive correlation between the two.
  - The authors should attempt to benchmark their study with other published work.
  - Include all important numerical results in the abstract.
  - Providing specific examples or data points that illustrate the computational burden of SHAP would strengthen the argument.

- Hooray!
  - The topic is about relevant and if work is presented well may have an impact.

To structure the revision of your paper, let's organize the feedback from the reviewers from the most complex tasks (which involve more substantial changes to the content or structure) to the finer details (such as formatting and minor language improvements).

1. **Clarify Objectives and Contributions (Reviewers 3 and 4)**
- **Clarify Objectives and Focus:** Simplify the title and objectives (Reviewer 3, Points 1, 2, 3). Explicitly highlight why the work is unique and its benefits. (Reviewer 3, Points 3).
- **Highlight Contributions:** Improve the introduction to clearly showcase the novelty of your work, linking to recent literature and explaining what has been done, what's missing, and your contributions (Reviewer 4, Points 2, 3, 4).
- **Summary of Previous Studies:** Add a summary of previous studies, emphasizing gaps that your work addresses (Reviewer 4, Point 4).

2. **Enhance Content Structure (Reviewers 1, 2, 4)**
- **Reorganize Sections for Clarity:** Consider reorganizing Section 5 by separating the description of datasets from performance evaluations for better clarity (Reviewer 1, Comment 9).
- **Benchmarking and Comparisons:** Include more benchmark analysis and comparison with both classic and recent models (Reviewer 2, Point 4; Reviewer 4, Points 9, 10). Ensure to address the fundamental reasons for performance differences in EIF+ and EIF (Reviewer 2, Point 2).
- **Extend Review of Related Work:** Expand your review of related work to include advanced data-driven techniques, Bayesian Networks, and studies from other prominent researchers (Reviewer 3, Point 7). This should also be summarized succinctly in the paper (Reviewer 4, Point 4).
- **Improve Figures and Tables:** Enhance the quality and organization of figures and tables, ensuring they are self-explanatory (Reviewer 4, Point 7; Reviewer 1, Comments 5, 9).
- **Add More Mathematical Rigor:** Introduce more mathematical formulae and equations to better demonstrate your proposed approach (Reviewer 4, Point 8).

3. **Incorporate Detailed Explanations and Examples (Reviewers 1 and 2)**
- **Algorithm Explanations and Pseudocode:** Provide example data, intermediate results, and pseudocode for the ExIFFI algorithm (Reviewer 2, Point 1). Include pseudocode for synthetic dataset generation (Reviewer 2, Point 3).
- **Linking Explanations to Real-World Applications:** Discuss how the findings, particularly feature importance and model performance, apply to real-world scenarios (Reviewer 1, Comment 14; Reviewer 2, Point 6).
- **Add Specific Explanations of Terms and Concepts:** Briefly define terms such as "Induced Imbalance Coefficients" and "Cumulative feature importance" when first introduced (Reviewer 1, Comment 6).

4. **Improve Discussion and Conclusion (Reviewers 1 and 4)**
- **Deepen Insights into Model Performance:** Provide deeper insights into why certain models, like DIF and AutoEncoder, perform poorly in specific scenarios (Reviewer 1, Comment 13). Discuss scenarios where ExIFFI explanations are reliable and aligned with human cognition (Reviewer 2, Point 7).
- **Elaborate on Limitations and Future Work:** Clearly state the limitations of your work and potential paths for future research in the conclusion (Reviewer 4, Point 11).
- **Enhance Discussions Linked to Figures and Tables:** Offer more detailed discussions of the results, closely linking them to figures and tables (Reviewer 4, Point 7).

5. **Simplify Language and Proofreading (Reviewers 1 and 4)**
- **Simplify Complex Sentences:** Revise complex sentences for clarity and readability, as suggested in Reviewer 1, Comments 1 and 2.
- **Proofreading:** Carefully proofread the manuscript to avoid grammatical and writing issues (Reviewer 4, Point 12).

6. **Finalize Formatting and Minor Details (Reviewer 1)**
- **Ensure Consistency in Citations:** Make sure all citations are consistent (Reviewer 1, Comment 4).
- **Improve Labeling and Referencing of Figures:** Ensure all figures are well-labeled, clearly referenced in the text, and have descriptive captions (Reviewer 1, Comment 5).
- **Include Important Numerical Results in the Abstract:** Summarize key numerical findings in the abstract to provide a comprehensive overview of the study (Reviewer 4, Point 1).

By following this plan, you will address the most complex and impactful feedback first, progressively working down to the finer details and formatting, ensuring your paper is both comprehensive and well-polished.