

Peer Review

Review of: "Optimizing Edge AI: A Comprehensive Survey on Data, Model, and System Strategies"

Karima Moumane¹

1. Mohammed V University, Rabat, Morocco

Review of the Paper: "Optimizing Edge AI: A Comprehensive Survey on Data, Model, and System Strategies"

General Comments

This paper provides a comprehensive survey on optimizing Edge AI through a data-model-system optimization framework. It effectively highlights the challenges of deploying AI models on resource-constrained edge devices and presents relevant optimization strategies. However, there are areas that require further improvement, including:

1. Comparative Analysis: A deeper comparison with existing systematic literature reviews on AI, cloud, and DevOps is missing.
2. Practical Applications: The paper lacks discussion on how Edge AI integrates into real-world domains such as healthcare and enterprise systems.
3. Usability and Performance Evaluation: There is limited reference to ISO 9126 usability frameworks for evaluating Edge AI efficiency and effectiveness.

1. Related Work – Missing Key References

The paper does not sufficiently compare its findings with existing systematic reviews on Edge AI optimization. To strengthen this section, I suggest including the following papers:

- Recommended Reference:
- "A Systematic Literature Review on Agile, Cloud, and DevOps Integration: Challenges, Benefits <https://www.sciencedirect.com/science/article/pii/S0950584924001745> " – This paper offers a

structured analysis of AI deployment challenges, which could improve the discussion on how Edge AI optimizations fit into broader enterprise environments, and the review of using AI and edge in smart cities: Opportunities, Applications, and Challenges of Edge-AI Enabled Video Analytics in Smart Cities: A Systematic Review <https://ieeexplore.ieee.org/abstract/document/10198424>.

2. Usability and Performance Metrics – Strengthen Discussion

The paper discusses model compression and system acceleration but does not introduce standard usability evaluation metrics to assess Edge AI performance in real-world applications.

- Recommended Reference:
- "An Empirical Evaluation of Mobile Software Usability Using ISO 9126 and QoS DiffServ Model <https://ebooks.iospress.nl/volumearticle/44439> " – This study presents a structured usability model, which could be used to evaluate Edge AI performance in low-latency environments.

as examples:

Understandability (Ease of learning & user guidance)	How easy is it for developers and data scientists to deploy Edge AI models?	- Measure the time required for first-time deployment.
---	---	--

Learnability (Ease of adoption)	How quickly can users understand Edge AI outputs?	- Assess how well explainable AI (XAI) techniques enhance model interpretability. - Conduct user tests with domain experts (e.g., healthcare professionals) to measure the time-to-understanding of AI decisions.
--	---	--

3. Real-World Integration of Edge AI – Practical Use Cases Needed

While the paper provides a technical review of Edge AI optimization, it does not explore domain-specific applications.

- Recommended Reference:
- "Enhancing the Processing of Healthcare Data Streams using Fog Computing <https://ieeexplore.ieee.org/abstract/document/10198424> " – This study discusses how Edge AI can

improve real-time healthcare data processing, which aligns well with data-driven optimizations in the reviewed paper.

General and Global Recommendations

1. Expand Related Work: Compare existing systematic reviews on AI, cloud, and DevOps.
2. Integrate Standardized Usability Frameworks: Introduce ISO 9126 usability metrics to assess Edge AI performance.
3. Add Practical Applications: Discuss real-world Edge AI use cases, particularly in healthcare, enterprise AI deployment, and smart cities.
4. Clarity and Structure
 - Improve the logical flow between sections, especially transitioning from challenges to optimization strategies.
 - Ensure consistent terminology (e.g., "Edge AI optimization" vs. "AI acceleration on the edge").
5. Depth of Analysis
 - Expand on trade-offs between optimization techniques (e.g., compression vs. quantization).
 - Provide more quantitative comparisons of existing Edge AI models.
6. Figures and Tables
 - Improve figure captions by explaining their relevance to Edge AI deployment.
 - Add a comparative table summarizing optimization strategies, benefits, and limitations.
7. Explainability and Trust
 - Discuss the interpretability of AI models on edge devices (e.g., how to ensure model decisions are transparent and reliable).
8. Future Work Section
 - Suggest how Edge AI can evolve with federated learning and 5G integration.
 - Address ethical concerns and data privacy in Edge AI applications.

Declarations

Potential competing interests: No potential competing interests to declare.