

EdgeAIMetric Benchmark de avaliação de desempenho e consumo energético de computadores de placa única para ambientes de IA na Borda

Gustavo Martins Conceição¹, Rodrigo da Rosa Righi²

¹Curso de Ciência da Computação – Universidade do Vale do Rio dos Sinos (Unisinos)
Av. Unisinos, 950 – 93022-750 – São Leopoldo – RS – Brazil

²Departamento de Sistemas e Computação
Universidade do Vale do Rio dos Sinos – São Leopoldo, RS – Brazil

gustavomccb@gmail.com

Resumo. *Com o crescimento da IoT, grandes volumes de dados são gerados, aumentando latência e desafios de segurança. A computação de borda reduz esses problemas ao permitir execução de IA próxima à fonte de dados. SBCs desempenham um papel essencial nesse cenário, mas seu desempenho e consumo energético precisam ser avaliados. Este artigo apresenta o EdgeAIMetric, um benchmark que mede CPU, RAM e consumo energético ao executar Decision Tree, K-Means, Naïve Bayes, SVM e CNN em diferentes conjuntos de dados. Os resultados experimentais evidenciam os trade-offs entre desempenho e eficiência, auxiliando a implementação de IA na borda.*

1. Introdução

Com o crescimento da Internet das Coisas (IoT), grandes volumes de dados são gerados e enviados para a nuvem para processamento. No entanto, essa abordagem tradicional enfrenta desafios como latência e problemas de privacidade. A computação de borda surge como solução, trazendo processamento mais próximo da origem dos dados, o que possibilita aplicações em tempo real, como Edge AI, reduzindo o consumo de rede e melhorando a eficiência. Os computadores de placa única (SBCs) vêm sendo amplamente utilizados para Inteligência Artificial na borda, tornando essencial avaliar seu desempenho. Nesse contexto, o modelo benchmark EdgeAIMetric é proposto para medir desempenho e consumo energético de SBCs ao executar diferentes algoritmos de IA. Esse modelo visa preencher uma lacuna na avaliação desses dispositivos, permitindo testes variados com diferentes cenários e parâmetros. Este trabalho tem como objetivo o desenvolvimento do EdgeAIMetric, possibilitando análises objetivas sobre o comportamento dos SBCs em aplicações de IA na borda. O artigo apresenta a fundamentação teórica, o modelo proposto, a metodologia de avaliação, os resultados obtidos e, por fim, a conclusão e sugestões para trabalhos futuros.

2. Trabalhos Relacionados

Neste capítulo, será apresentado o processo de seleção dos trabalhos relacionados que foram analisados, seguido de uma análise comparativa destacando suas características, contribuições e oportunidades identificadas no estado-da-arte. Foram selecionados trabalhos utilizados como base para a pesquisa realizada, estabelecendo uma base sólida de embasamento teórico e comparação relevante para o objetivo deste estudo. A identificação

dos trabalhos foi feita por consulta em bases de dados do IEEE, Google Acadêmico, Springer e Elsevier. A seleção foi realizada utilizando palavras-chave como "Edge AI", "Edge Computing", "Desempenho", "Eficiência Energética", "Single Board Computer" e "Benchmarking". A combinação dessas palavras nas bases de dados resultou em 200 artigos para revisão inicial, dos quais 12 foram selecionados para estudo. A análise focou em identificar trabalhos voltados ao desenvolvimento de benchmarks para avaliação de dispositivos para computação e inteligência artificial na borda. Para a análise comparativa, foram considerados critérios específicos que impactam o desempenho e a eficiência energética na implementação de algoritmos de inteligência artificial em computadores de placa única. Os principais tópicos analisados incluem a metodologia ou algoritmo utilizado no artigo, identificação do dispositivo de computador de placa única utilizado no estudo, quantidade de testes que foram realizados e peculiaridades dos testes. Também foi feita avaliação das métricas específicas utilizadas nos estudos, incluindo consumo e eficiência energética, desempenho da unidade de processamento central, unidade de processamento gráfico e memória e análise da relação entre as métricas de desempenho avaliadas.

Observando os estudos apresentados, é possível identificar lacunas no desenvolvimento dos benchmarks estudados. Por exemplo, [Tamburello et al. 2023] analisaram apenas um algoritmo, limitando a diversidade dos testes. Além disso, trabalhos como [Sánchez Sánchez et al. 2023] e [Baller et al. 2021] executaram algoritmos sem alteração dos parâmetros, o que pode impactar a análise da variabilidade dos desempenhos. Estudos como [Baller et al. 2021] e [Hadidi et al. 2019] não consideraram medições de consumo de CPU e memória RAM, métricas fundamentais para avaliar o desempenho em ambientes de borda. Por outro lado, algumas pesquisas enfatizam medições de desempenho da unidade de processamento central e de outros componentes do computador, como [Muhoza et al. 2023]. No entanto, não foi realizado o monitoramento simultâneo do consumo energético. Já [Sánchez Sánchez et al. 2023] utilizaram múltiplos dispositivos, mas todos pertencentes à mesma família, limitando a abrangência da comparação. Diante dessas observações, conclui-se que há oportunidades para o desenvolvimento de um modelo de benchmark que teste dispositivos em múltiplos cenários, com diferentes conjuntos de parâmetros, além de executar testes em diversos algoritmos, fornecendo uma avaliação mais abrangente do desempenho de computadores de placa única.

3. Modelo EdgeAIMetric

O EdgeAIMetric é um modelo de benchmark projetado para avaliar o desempenho e consumo energético de computadores de placa única ao executar algoritmos de inteligência artificial. Segundo [Grambow et al. 2019], benchmarks permitem a comparação entre diferentes sistemas e configurações. Esta seção apresenta as decisões de projeto, o funcionamento do benchmark e o ambiente de testes. Os testes avaliam o desempenho de computadores de placa única executando algoritmos de IA em diferentes cenários. O treinamento será realizado diretamente nos dispositivos, com armazenamento interno ou conectado. O modelo inclui módulos específicos: o Módulo de Gerenciamento de Dados, que coleta métricas de desempenho, e o Módulo de Controle de Execução, que garante a execução padronizada dos testes. São analisados CPU, memória RAM e disco, além do consumo energético, obtido a partir da medição de corrente e tensão elétrica. A potência elétrica é calculada como o produto dessas grandezas, e a energia elétrica como o produto

da potência pelo tempo de execução. Os algoritmos utilizados incluem K-Means, SVM, Naïve Bayes, Árvore de Decisão e Redes Neurais Convolucionais, escolhidos com base em literatura relevante e ampla documentação.

3.1. Funcionamento

Os testes são realizados em quatro cenários distintos, cada um variando a configuração dos algoritmos e o tamanho da base de dados MNIST. O primeiro e segundo cenários utilizam o conjunto completo, enquanto o terceiro e quarto utilizam metade dos dados. O Módulo de Controle de Execução carrega os dados armazenados em micro SD, normaliza-os e divide em 80% para treinamento e 20% para teste. Durante a execução, o Módulo de Controle monitora o teste, avançando para o próximo algoritmo caso o dispositivo não consiga concluir a execução. Para garantir a reprodutibilidade, os testes serão realizados em ambientes controlados, mantendo temperatura e umidade constantes.

3.2. Algoritmos Utilizados

Os testes nos computadores de placa única utilizam os algoritmos K-Means, SVM, Naïve Bayes, Árvore de Decisão e Rede Neural Convolucional, selecionados com base nos trabalhos estudados na seção Trabalhos Relacionados. A escolha considerou as tecnologias mais utilizadas no mercado e sua aplicabilidade a diferentes cenários de IA. Para a implementação, são empregadas as bibliotecas Scikit-learn e TensorFlow Lite. Os parâmetros dos algoritmos são definidos com base na literatura e nas documentações das ferramentas utilizadas. Ao final dos testes, o Módulo de Gerenciamento de Dados armazenará os resultados para análise comparativa do desempenho dos dispositivos em diferentes cenários e algoritmos.

4. Metodologia e Resultados

A avaliação do desempenho e consumo energético de computadores de placa única foi conduzida por meio de um benchmark testado em três dispositivos: Raspberry Pi 5, Raspberry Pi 4 e Orange Pi 3. Utilizando as bibliotecas Scikit-learn e TensorFlow, os algoritmos foram executados e seus resultados registrados em arquivos CSV. O consumo energético foi medido com um multímetro TC66, capturando tensão e corrente elétrica para análise da potência elétrica em watts. Durante os testes, o uso de CPU, RAM e disco foi monitorado a cada segundo, garantindo uma avaliação detalhada do impacto computacional. Cada experimento foi repetido cinco vezes para assegurar consistência nos resultados. Os testes foram organizados em quatro cenários distintos, considerando diferentes configurações de parâmetros e frações do dataset MNIST. No Cenário 1, utilizando o dataset completo com parâmetros moderados, apenas o Raspberry Pi 5 e o Orange Pi 3 conseguiram executar os algoritmos, com exceção do KNN e da Rede Neural Convolucional no Orange Pi. O Raspberry Pi 5 apresentou maior consumo energético, possivelmente devido ao seu desempenho superior. Já no Cenário 2, com parâmetros otimizados, o Orange Pi 3 conseguiu rodar apenas Árvore de Decisão e Máquina de Vetores de Suporte, enquanto o Raspberry Pi 5 não executou KNN, evidenciando as limitações de hardware.

Com metade do dataset MNIST, os Cenários 3 e 4 mostraram melhor desempenho geral. No Cenário 3, os dispositivos conseguiram executar Árvore de Decisão e Naïve Bayes, com consumo energético proporcional ao uso de CPU e RAM. No Cenário 4, o Orange Pi 3 manteve limitações semelhantes, rodando apenas Árvore de Decisão e Naïve

Bayes, enquanto o Raspberry Pi 5 também executou Rede Neural Convolutacional. Comparando os cenários, reduzir o volume de dados permitiu execuções mais rápidas, especialmente para algoritmos mais exigentes como CNN. Os resultados mostraram padrões distintos de consumo energético e uso de hardware. No Naive Bayes, o Orange Pi 3 apresentou uso mais estável de RAM, sem ultrapassar 80%. Já na Árvore de Decisão, observou-se que o consumo energético aumentava conforme o uso de CPU. O Raspberry Pi 4, por sua vez, não conseguiu executar nenhum dos cenários, evidenciando limitações críticas para esse tipo de aplicação. A análise comparativa demonstrou que o Raspberry Pi 5 possui maior capacidade para lidar com uma gama mais ampla de algoritmos, enquanto o Orange Pi 3 apresenta restrições significativas ao executar cargas mais pesadas.

5. Conclusão

O trabalho apresentou o modelo EdgeAIMetric, uma abordagem versátil para avaliar o desempenho e o consumo energético de computadores de placa única em diferentes cenários de Inteligência Artificial. O modelo permite testar diversos algoritmos com diferentes parâmetros e tamanhos de conjuntos de dados, oferecendo uma análise abrangente do comportamento dos dispositivos. Os resultados destacam a capacidade do EdgeAIMetric de identificar limitações de hardware, comparando o desempenho e o consumo energético entre dispositivos, como demonstrado na análise do uso de memória no Algoritmo Árvore de Decisão. Além disso, o modelo revelou tendências no uso de CPU, RAM, disco e potência elétrica, garantindo confiabilidade por meio da repetição dos testes. O estudo contribui para a literatura com um benchmark detalhado e pode auxiliar na escolha do melhor dispositivo para aplicações específicas.

Referências

- Baller, S. P., Jindal, A., Chadha, M., and Gerndt, M. (2021). Deepedgebench: Benchmarking deep neural networks on edge devices. In *2021 IEEE International Conference on Cloud Engineering (IC2E)*, pages 20–30.
- Grambow, M., Lehmann, F., and Bermbach, D. (2019). Continuous benchmarking: Using system benchmarking in build pipelines. In *2019 IEEE International Conference on Cloud Engineering (IC2E)*, pages 241–246.
- Hadidi, R., Cao, J., Xie, Y., Asgari, B., Krishna, T., and Kim, H. (2019). Characterizing the deployment of deep neural networks on commercial edge devices. In *2019 IEEE International Symposium on Workload Characterization (IISWC)*, pages 35–48.
- Muhoza, A. C., Bergeret, E., Brdys, C., and Gary, F. (2023). Power consumption reduction for iot devices thanks to edge-ai: Application to human activity recognition. volume 24, page 100930.
- Sánchez Sánchez, P. M., Jorquera Valero, J. M., Huertas Celdrán, A., Bovet, G., Gil Pérez, M., and Martínez Pérez, G. (2023). Lwhbench: A low-level hardware component benchmark and dataset for single board computers. volume 22, page 100764.
- Tamburello, M., Caruso, G., Adami, D., and Giordano, S. (2023). Experimental comparison between sbc and fpga for embedded neural network acceleration. In *ICC 2023 - IEEE International Conference on Communications*, pages 6078–6083.