Exploring Apple Silicon's Potential from Simulation and Optimization Perspective

Karol Struniawski[1][0000−0002−4574−2986],
Aleksandra Konopka[1][0000−0003−1730−5866] and
Ryszard Kozera[1,2][0000−0002−2907−8632]

[1] Institute of Information Technology,
Warsaw University of Life Sciences - SGGW,
ul. Nowoursynowska 159, 02-776 Warsaw, Poland
{karol_struniawski,aleksandra_konopka,ryszard_kozera}@sggw.edu.pl
[2] School of Physics, Mathematics and Computing,
The University of Western Australia,
35 Stirling Highway, Crawley, WA 6009, Perth, Australia
ryszard.kozera@uwa.edu.au

Abstract. This study explores the performance of Apple Silicon processors in real-world research tasks, with a specific focus on optimization and Machine Learning applications. Diverging from conventional benchmarks, various algorithms across fundamental datasets have been assessed using diverse hardware configurations, including Apple's M1 and M2 processors, NVIDIA RTX 3090 GPU and a mid-range laptop. The M2 demonstrates competitiveness in tasks such as *BreastCancer, liver* and *yeast* classification, establishing it as a suitable platform for practical applications. Conversely, the dedicated GPU outperformed M1 and M2 on the *eyestate1* dataset, underscoring its superiority in handling more complex tasks, albeit at the expense of substantial power consumption. With the technology advances, Apple Silicon emerges as a compelling choice for real-world applications, warranting further exploration and research in chip development. This study underscores the critical role of device specifications in evaluating Machine Learning algorithms.

Keywords: Machine Learning, Optimization, Simulation, Apple Silicon, Extreme Learning Machine

1 Introduction

In November 2020, Apple introduced a new line of processors, starting with the M1 chip that adopts a System on a Chip (SoC) design with unified memory. The M1 processor, built using 5nm process technology and containing 16 billion transistors, also integrated the task specific modules like Apple Neural Engine. Over subsequent years, Apple released upgraded versions like the M1 Pro/M1 Max in 2021, Ultra in 2022, M2 in 2022, M2 Pro/Max/Ultra in 2023, and M3, M3 Pro/Max in 2023, all promising improved performance. The Apple M1 SoC is a highly integrated processor unit that includes all of the necessary components for

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_3

a fully working computer while consuming less power in general, making it available to customers in the markets without losing performance. The technological innovations in this field have promising future and grant more investigation and research toward the development of the chips [11, 14].

In addition to noteworthy features such as prolonged battery life and fanless design in MacBook Air models, contributing to their quiet and portable nature, these devices have piqued the interest of researchers due to their potential applications in scientific endeavors. More specifically, researchers can harness the computational capabilities of Apple Silicon for tasks like optimization and Machine Learning (ML) calculations [3]. This study aims to evaluate the practicality and effectiveness of Apple Silicon-powered devices in tasks commonly undertaken by researchers, with a focus on performing calculations. The primary objective in this study is to evaluate various fundamental optimization and ML algorithms across diverse datasets. The notable gap in the existing literature is detected, where performance assessments are conducted on a single dataset [10] or are limited to a single ML method. This paper is an extension of Kasperek et al. [9] suggesting the possibility to further expand their research to CUDA-enabled devices. This study employed a CUDA device - RTX 3090.

2 Apple Silicon Overview

The Apple Silicon processors, such as the M1, utilize a Unified Memory Architecture (UMA) that allows for shared memory access across different modules of the SoC [7]. This means that the RAM is a single pool of memory that all parts of the processor can access, enabling the GPU to utilize more system memory while other parts of the SoC ramp down, without the need to shuttle data between different memory spaces [10]. In contrast, traditional CPU devices have separate memory spaces for the GPU and CPU, requiring data movement between these spaces, which can be inefficient. The benefits of UMA are particularly evident in the context of ML tasks, where the Apple Silicon chip offers hardware acceleration support, making it a tempting option for researchers. Additionally, the use of UMA has been found to be beneficial when only a small random portion of data is accessed for a set of benchmarks, highlighting its efficiency [13]. In the realm of SoC architectures, many-core architectures with shared memory are preferred for flexible and programmable solutions in computationally intensive application domains, including ML and embedded processing [12].

The MacOS operating system leverages the concept of shared memory to enhance performance by expanding UMA with swap memory, albeit with a trade-off in effectiveness [6]. This approach allows a more flexible allocation of resources, particularly in the context of Apple Silicon devices, where different RAM sizes are available. Therefore, comparing the performance of devices without considering the RAM utilization may lead to incorrect or incomplete conclusions.

The low-energy SoC chip offers clear advantages, notably in terms of extended battery life and optimal performance per watt power. Significantly, the operational efficiency of Apple Silicon devices remains consistent whether operating

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_3

on battery power or when connected to an external power source, a capability not commonly observed in conventional computing systems.

These advantages become even more pronounced given the escalating energy prices in Europe following the aftermath of the conflict in Ukraine [2]. For instance, the M1-powered Mac mini demonstrates an average power consumption ranging from 10W to 31W [5]. In contrast, a PC-class device equipped with an AMD R9 or Intel i9 CPU and dedicated GPU like the NVIDIA RTX 3090 can consume up to 800W at peak performance (calculated based on the cumulative peak power consumption of individual PC components as per manufacturer specifications). The significance of power consumption is underscored by the current global scenario, where electricity demand is outpacing the growth of renewable sources [8]. Highlighting this, the Cinebench R23 Single Package Power Efficiency metric reflects favorably on the SoC, registering 297 points per watt. In comparison, competitors such as the Ryzen 5 5600U score 90.8 and the Intel i5-1240P scores 64 points [1]. This underlines the efficiency and energy-conscious performance of the low-energy SoC chip in a landscape where power consumption considerations are paramount.

3 Methodology

In preceding experiments that compared the NVIDIA V100 and A100 GPUs with the M1 and M1 Ultra, the obtained results were promising, showcasing the superior performance of Apple Silicon over both GPUs [10]. Despite of the aforementioned GPUs produced impressive results, they do not represent the pinnacle of current GPU capabilities, with the NVIDIA V100 providing 14.13 TFLOPS Float32 precision to the RTX 3090's 35.58. To comprehensively assess the performance of selected ML classifiers across diverse hardware platforms and data types the six benchmark datasets are employed [4], wherein the number of samples, features and classes for each task is specified (see Tab. 1). The objective is to measure the execution time of each classifier on three distinct hardware platforms: Apple's M1 with 8GB RAM and M2 with 16GB RAM, a high-performance NVIDIA RTX 3090 GPU with 24GB memory and a mid-range laptop configuration featuring an Intel Core i5 11500h processor and an NVIDIA RTX 3050ti graphics card. The intended experiment also aimed to compare the performance of mobile devices (M1/M2) with an i5-powered laptop. Surprisingly, the unplugged i5 showed four times longer performance on average compared to the plugged-in scenario. On the other hand, the M1/M2 devices maintained consistent computational power on battery. Regrettably, the i5's limited battery life led to the decision to forgo the experiment before completion.

This approach ensures a comprehensive evaluation that extends beyond the previously explored GPUs, providing insights into the real-world performance of the classifiers across a spectrum of hardware configurations. A variety of ML methods, including Extreme Learning Machine (ELM), k-Nearest Neighbors (kNN), Multi-Layer Perceptron (MLP), Random Forest (RF) and Support Vector Machine (SVM), are employed for the datasets (see Tab. 1, column labels)

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_3

	BreastCancer	*eyestate1*	*liver*	*musk*	*waveform*	*yeast*
Samples	699	762	345	1682	500	150
Features	9	14	6	166	21	8
Classes	2	3	4	2	10	10

Table 1. Details of datasets utilized in performance evaluations.

4 Experimental Results

All experiments were conducted in Python 3.11 using Tensorflow 2.15, scikit-learn 1.4.0 and numpy 1.26.3 on MacOS 14.2.1 or Windows 11. The outcomes, illustrated in Figures 2-6, represent the time taken for training and testing during 10 times repeated 10-fold cross-validation, ensuring result significance by mitigating odd observations.

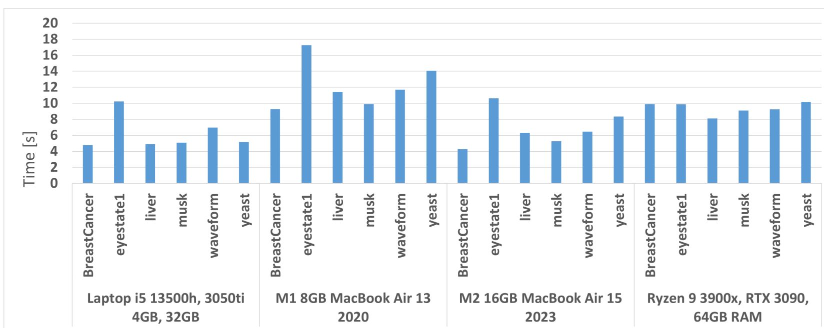

Fig. 1. Extreme Learning Machine with 100 neurons in hidden layer execution time.

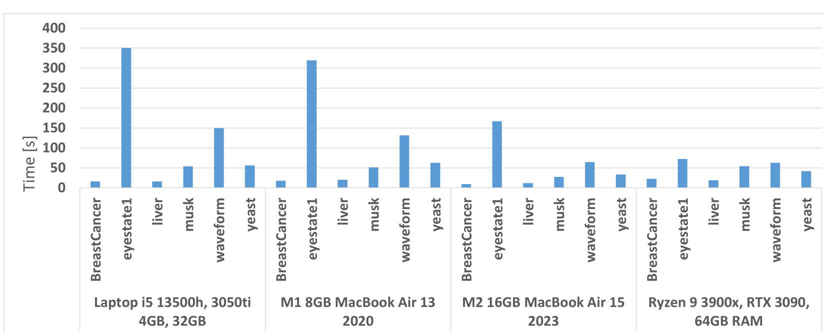

Fig. 2. Extreme Learning Machine with 1000 neurons in hidden layer execution time.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_3

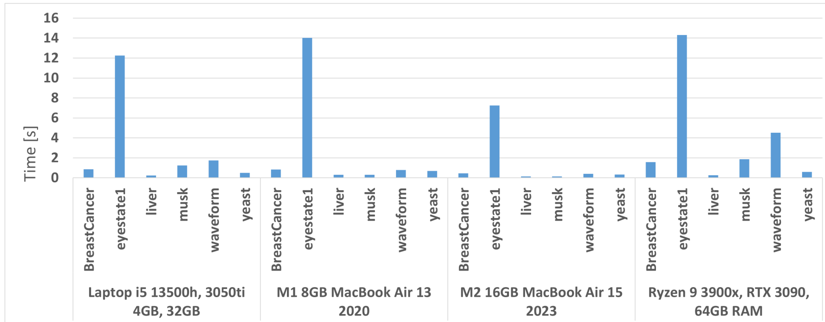

Fig. 3. K-Nearest Neighbours classifier execution time.

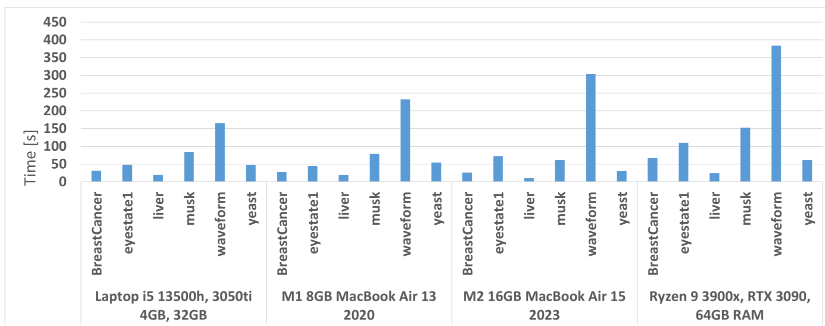

Fig. 4. Multi-Layer Perceptron execution time.

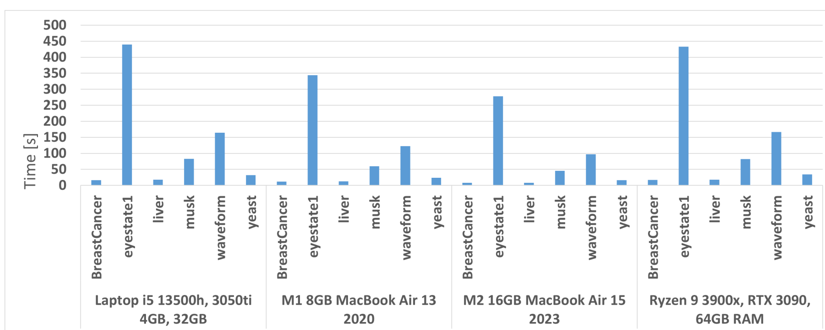

Fig. 5. Random Forest execution time.

5 Discussion

In the case of ELM with 100 hidden layer units, subtle differences emerge, with the most notable discrepancy found in the execution time on the M1 8GB device, which is approximately twice as long as the M2 16GB counterpart (see Fig. 1). Surprisingly, the i5 laptop yields comparable results to the M2 16GB.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_3

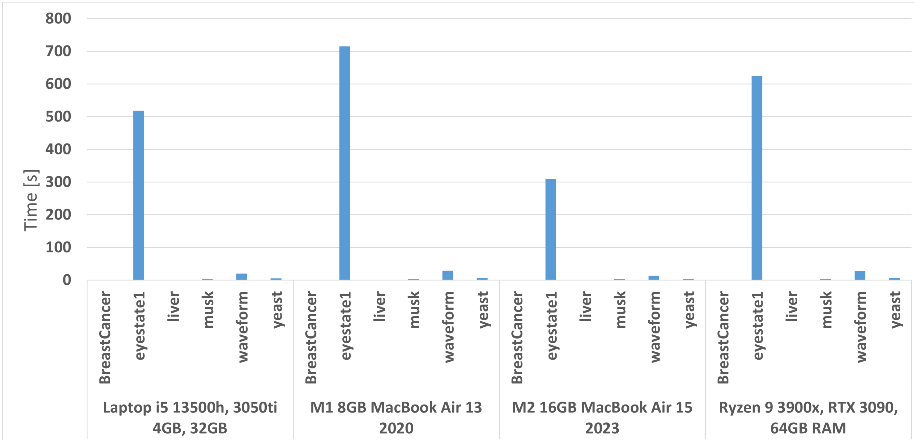

Fig. 6. Support Vector Machine classifier execution time.

Device	*bc*	*eyestate1*	*liver*	*musk*	*waveform*	*yeast*	sum
Laptop i5 11500h, 3050ti, 32GB	657	16210	670	2516	46898	3360	70311
M1 8GB MacBook Air 13 2020	661	69000	700	17777	38621	3034	129792
M2 16GB MacBook Air 15 2023	331	38477	342	3055	22585	1582	66372
Ryzen 9 3900x, RTX 3090, 64GB	401	6058	376	1170	17816	1633	27454

Table 2. Performance evaluation results, where *bc* states as *BreastCancer*.

Notably, the performance of the RTX 3090 is unexpectedly inferior to the M2 on each dataset. The differences in results across various datasets for the RTX 3090 are minimal, indicating that the GPU's memory allocation necessitated longer processing time. Despite this, the RTX 3090's rapid CUDA cores and ample 24GB memory mitigate the dataset size impact for this classifier. The *eyestate1* dataset requires the most time for processing by the classifier, with both the i5 and M1 machines struggling for ELM with 1000 neurons (see Fig. 2). Specifically, the M1 requires times longer than the M2 and the M2 takes twice as long to execute compared to the RTX 3090. Conversely, for the other datasets, the differences between M2 and the RTX 3090 are less pronounced and appear comparable.

Moving to kNN, M2 emerges as the fastest across all devices, surpassing the RTX 3090 by a few times for the *waveform* dataset. Inexplicably, the RTX 3090 delivers suboptimal results despite having updated drivers and configurations, consistent across repeated experiments (see Fig. 3). Similar patterns are observed with MLP with topology (10, 10), where the RTX 3090 consistently produces the worst results, yet for the *waveform* dataset, the i5 device is the fastest, followed by the M1 and then the M2 (see Fig. 4).

In the context of RF, the RTX 3090 once again yields subpar results, while the M2 proves to be the fastest (see Fig. 5). Similar trends persist for the SVM

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_3

method, with the RTX 3090 delivering suboptimal results and the M2 demonstrating the fastest performance (see Fig. 6).

Considering the real-world scenario where all classifiers run on a given dataset, the aim is to compare the models' overall performance. Combining the total time required for a device to run all classifiers on different datasets, along with the ultimate sum of running all classifiers on all datasets, reveals interesting insights (see Tab. 2). The RTX 3090 emerges as the leader with a combined time of 27454 seconds, whereas the M2 is twice as slow. The i5 device demonstrates comparable performance to the M2, while the M1 lags as the slowest due to memory-intensive tasks. A closer examination highlights the substantial impact of the *eyestate1* dataset, where the RTX 3090 outperforms the M2 sixfold, showcasing the dedicated GPU's potential for more complex datasets. Conversely, M2 excels in tasks such as *BreastCancer*, *liver* and *yeast*, underscoring its competitive edge in certain scenarios against an 800W machine with a 30W device.

6 Conclusion

In conclusion, the results reveal nuanced variations in the performance of ML classifiers across diverse datasets and hardware configurations. The ELM with 100 hidden layer units showcases subtle differences, with notable disparities in execution time between devices. ELM with 1000 neurons introduces new dynamics, impacting performance across datasets.

In specific algorithms like kNN, MLP, RF and SVM, the Apple M2 processor consistently demonstrates promising performance compared to the Nvidia RTX 3090 GPU, highlighting the efficacy of Apple Silicon in real-world applications, especially taking into account the performance per watt power.

These findings underscore the importance of considering device specifications and configurations when assessing the practicality and effectiveness of ML algorithms. The competitive edge of Apple Silicon, particularly the M2 processor, is evident in various scenarios, showcasing its potential for tasks such as *Breast-Cancer*, *liver* and *yeast*, even against higher-power GPU counterparts.

References

1. Cinebench scores, https://nanoreview.net/en/cpu-list/cinebench-scores
2. Ari, A., et al.: Surging energy prices in Europe in the aftermath of the war: how to support the vulnerable and speed up the transition away from fossil fuels. IMF Working Papers **2022**(152), A001 (2022). https://doi.org/10.5089/9798400214592.001.A001
3. Dalakoti, V., Chakraborty, D.: Apple M1 chip vs Intel (X86). EPRA International Journal of Research and Development (IJRD) **7**(5), 207–211 (2022)
4. Dheeru, D., Karra Taniskidou, E.: UCI machine learning repository (2017), http://archive.ics.uci.edu/ml
5. Frumusanu, A.: The 2020 Mac Mini unleashed: Putting Apple Silicon M1 to the test (2020), https://www.anandtech.com/show/16252/mac-mini-apple-m1-tested

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_3

6. Golden, R., Case, A.: In lieu of swap: Analyzing compressed RAM in Mac OS X and Linux. Digi. Invest. **11**, S3–S12 (2014). https://doi.org/https://doi.org/10.1016/j.diin.2014.05.011
7. Hart, C.: CDRAM in a Unified Memory Architecture. In: Proc. COMPCON. pp. 261–266 (1994). https://doi.org/10.1109/CMPCON.1994.282913
8. IEA: Electricity market report - july 2021 (2021), https://www.iea.org/reports/electricity-market-report-july-2021
9. Kasperek, D., Podpora, M., Kawala-Sterniuk, A.: Comparison of the usability of Apple M1 processors for various Machine Learning tasks. Sensors **22**(20) (2022). https://doi.org/10.3390/s22208005
10. Kenyon, C., Capano, C.: Apple silicon performance in scientific computing. arXiv (2022). https://doi.org/10.48550/arXiv.2211.00720
11. Liao, X., Li, B., Li, J.: Impacts of Apple's M1 SoC on the Technology Industry. In: Proceedings of the 2022 7th International Conference on Financial Innovation and Economic Development (ICFIED 2022). pp. 355–360 (2022). https://doi.org/10.2991/aebmr.k.220307.056
12. Luan, H., Gatherer, A.: Combinatorics and geometry for the many-ported, distributed and shared memory architecture. In: 2020 14th IEEE/ACM International Symposium on Networks-on-Chip (NOCS). pp. 1–6 (2020). https://doi.org/10.1109/NOCS50636.2020.9241708
13. Xu, H., Lin, P.H., Emani, M., et al.: Xunified: a framework for guiding optimal use of GPU Unified Memory. IEEE Access **10**, 82614–82625 (2022). https://doi.org/10.1109/ACCESS.2022.3196008
14. Zhang, Z.: Analysis of the advantages of the M1 CPU and its impact on the future development of Apple. In: 2021 2nd International Conference on Big Data and Artificial Intelligence and Software Engineering (ICBASE). pp. 732–735 (2021). https://doi.org/10.1109/ICBASE53849.2021.00143

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_3