International Journal of Innovative Research and Scientific Studies, 8(1) 2025, pages: 899-906

IJIRSS International Journal of Innovative Research & Scientific Studies

ISSN: 2617-6548

URL: www.ijirss.com

A study on the improvement of supercomputer energy efficiency based on green500 benchmarking data

Hyungwook Shim[1*], Minho Suh[2]

[1,2]Korea Institute of Science and Technology Information, South Korea.

Corresponding author: Hyungwook Shim (*Email: shw@kisti.re.kr*)

Abstract

With the rapid growth of AI-related industries, the need for reducing and optimizing energy consumption in large-scale computational resources, such as supercomputers, has become increasingly important. This study focuses on supercomputers listed in the Green 500, categorizing existing benchmarking evaluation variables into input and output factors. An energy efficiency objective function was introduced, and DEA was conducted using BCC and SEM models. The study analyzed the relative efficiency levels among supercomputers and identified factors and levels of potential efficiency improvements. The results provide insights into the performance factors of individual supercomputers and their potential for improvement. Furthermore, by comparing the energy efficiency evaluated by the Green 500 with the results of DEA, it demonstrated the potential for utilizing DEA as a new means for efficiency improvement. It also highlighted the necessity of a comprehensive evaluation that incorporates various performance factors, rather than a simple efficiency assessment based solely on energy consumption.

Keywords: AI, Computational resource, DEA, Energy efficiency, Supercomputer.

DOI: 10.53894/ijirss.v8i1.4448
Funding: This research was supported by Korea Institute of Science and Technology Information (KISTI) (Grant number: K25L2M1C1).
History: Received: 12 December 2024/**Revised:** 20 January 2025/**Accepted:** 27 January 2025/**Published:** 3 February 2025

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: Both authors contributed equally to the conception and design of the study. Both authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Publisher: Innovative Research Publishing

1. Introduction

As the AI industry continues to evolve, the exponential growth in data volume and complexity demands increasingly powerful computational capabilities. This is driving significant advancements in supercomputing technologies, enabling rapid processing and analysis of large datasets essential for training and deploying sophisticated AI models. As a result, the global supercomputer market, valued at $1.9 billion in 2023, is expected to grow at a compound annual growth rate (CAGR) of 18.97%, reaching $6.43 billion by 2030. This rapid expansion reflects the strategic importance of supercomputing infrastructure in various domains, including scientific research, national security, healthcare, and industry-driven AI innovation. In this competitive landscape, countries, corporations, and research institutions are heavily investing in the development and deployment of high-performance computing systems to establish leadership in this field. The introduction of El Capitan in November 2024, boasting an Rmax of 2,746 PFLOPS, underscores the race to achieve unprecedented computational speeds. Alongside, global distribution reveals significant disparities, with the United States housing 172 supercomputers, followed by 62 in China and 41 in Germany. This distribution highlights the geopolitical and strategic nature of supercomputing capabilities. However, the growing adoption of supercomputers has raised concerns about their environmental impact, particularly in terms of energy consumption. In Germany, data centers hosting supercomputing infrastructure consume approximately 4% of the nation's total electricity. Similarly, Frontier, the world's second-most powerful supercomputer, requires a staggering 22.7 MW of electricity, a demand comparable to that of a medium-sized solar power plant. Addressing such challenges requires a multi-faceted approach that combines technological innovation with sustainable operational strategies. Current advancements in energy-efficient hardware and cooling technologies are paving the way for more sustainable supercomputing. For instance, NVIDIA's Grace Hopper processor incorporates energy efficiency optimizations, and LUMI employs pre-cooling systems that utilize natural airflow while repurposing waste heat for local district heating grids. Liquid cooling technologies, such as ManicoreSoft's Deep Gadget, further eliminate the need for traditional air-conditioning systems, significantly reducing energy consumption. These technological strides demonstrate the potential for reducing supercomputing's environmental footprint. Despite these advancements, the scope of improving supercomputing energy efficiency extends beyond hardware and cooling systems. Operational strategies, inspired by other industries such as electric vehicles, offer promising avenues for optimization. In the electric vehicle sector, energy efficiency is achieved not only through improved components, such as motor controllers and battery systems, but also through meticulous analysis of configurations, including battery capacity, driving speed, and aerodynamics. Analogously, supercomputers could benefit from a systems-level approach that investigates the interplay of workload distribution, scheduling algorithms, data transfer efficiency, and cooling system configurations. Moreover, the integration of AI-driven monitoring systems can optimize energy consumption in real time. Machine learning algorithms can predict computational loads, enabling dynamic allocation of resources to minimize energy usage during periods of low demand. Research into power-aware scheduling policies can ensure that performance requirements are met while reducing operational costs. For instance, studies have demonstrated that dynamic voltage and frequency scaling (DVFS) techniques can effectively reduce energy consumption without compromising computational throughput. In addition, collaborations between academia, industry, and government are essential to foster the development of comprehensive energy-efficiency frameworks. Standardizing metrics for measuring and comparing supercomputing energy efficiency will be critical to benchmarking progress and identifying best practices. Furthermore, international cooperation can play a pivotal role in addressing global challenges associated with the sustainability of supercomputing infrastructures, encouraging knowledge-sharing and co-development of innovative solutions. As the reliance on supercomputers grows, the need to balance performance and sustainability becomes increasingly urgent. The path forward lies in embracing a holistic approach that combines cutting-edge technology with strategic planning and policy development. By addressing both technical and operational aspects, the supercomputing community can ensure that advancements in computational power are achieved in an environmentally responsible and economically viable manner.

This paper consists of 4 chapters. Chapters 1 and 2 analyze the background of the theory and previous research to derive the necessity and academic value of the study. Chapter 3 explains the research procedure and research data, and presents the efficiency analysis results. Finally, in the conclusion of Chapter 4, the results are summarized, and the implications obtained from the research results and the limitations of this study are discussed.

2. Literature Review and Theoretical Background

León-Vega, et al. [1] provides insight into the comprehensive behavior of high-performance computing (HPC) systems by examining the impact of executed instructions on overall power consumption. We propose two new mathematical models to estimate the energy consumption of a process, which are based on a normalized vector of total node energy, process usage, and probability distributions for the instruction types of CPU and GPU processes. This approach enables energy accounting for specific processes without the need for isolation of individual processes León-Vega, et al. [1]. Prieto, et al. [2] investigates the evolution of this parameter by analyzing high-performance computers from 2008 to 2023 and presents results comparing it to Koomey's law. Comparing the two results, we conclude that energy efficiency will continue to grow exponentially over the study period and into the foreseeable future, but at a slower rate than suggested by Koomey's law (maximum energy efficiency every 1.57 years). Instead of doubling, it doubles every 2.29 years Prieto, et al. [2]. Carastan-Santos, et al. [3] proposes a method to predict and leverage the power consumption of HPC workloads. This method aims to reduce the power consumption of the supercomputer while maintaining the management (scheduling) performance of the Resource Task Management System (RJMS). The proposed method leverages workload submission logs and power monitoring data and uses a mix of lightweight power prediction methods and a heuristic approach inspired by classic EASY Backfilling Carastan-Santos, et al. [3]. Erol, et al. [4] proposes a new modular approach to standard shrouded packaging. This approach aims to achieve better thermo-mechanical performance and act as a reworkable thermal plug. This approach, called CoolStar, uses an ultra-low modulus gel-like gap-filter to attach a fractal heat spreader to a liquid cooler. This paper presents a proof of concept, showing that this is a promising approach for addressing thermo-mechanical stresses at the system level Erol, et al. [4]. Naduvilakath-Mohammed, et al. [5] presents a numerical model of a compact vapor compression refrigeration (VCR) system that cools liquid coolant in a secondary pump circuit used to cool high-performance CPUs. The model utilizes a physical approach and an iterative algorithm to solve the coupled nonlinear equations of the cooling cycle and single-phase cooling loop Naduvilakath-Mohammed, et al. [5]. Song [6] investigates various techniques for optimizing chip performance and energy efficiency through design innovation, materials science, and strategic management of the manufacturing process. By focusing on integrating these innovations into current manufacturing practices, we aim to minimize energy consumption while meeting the demands of HPC. Key innovations include multicore designs and heterogeneous computing, while advances in materials such as graphene and silicon carbide are also reviewed. Additionally, manufacturing technologies such as extreme ultraviolet lithography (EUVL) and 3D additive processes are analyzed [6]. A review of previous studies reveals significant research on improving the energy efficiency of supercomputers, focusing on cooling systems, sealed package technologies that enhance thermo-mechanical performance, and management technologies that predict power consumption under varying workloads while maintaining performance. Additionally, there has been research on models for measuring and evaluating energy consumption. Key insights from these studies are as follows: First, most technological developments have focused on single factors. However, multiple factors influence energy efficiency, and these factors are interconnected, collectively impacting energy consumption. Furthermore, energy efficiency is mutually influenced by computational performance. Thus, strategies for improving the energy efficiency of supercomputers must comprehensively account for multiple factors simultaneously. An optimal strategy should be formulated by considering all relevant factors, and improvement goals for each factor should be established based on this strategy.

The challenge lies in reconciling the demands of high computational performance with sustainability objectives, especially as workloads diversify with the increasing adoption of AI applications. Enhanced cooling solutions, efficient scheduling algorithms, and energy-aware hardware design must be complemented by comprehensive benchmarks such as the Green500. Moreover, adaptive strategies leveraging machine learning could dynamically optimize energy use across varying workloads, ensuring a balance between power efficiency and computational throughput. Prieto, et al. [2] also observed that the rate of technological advancements in energy efficiency has slowed compared to Koomey's Law, highlighting the urgent need for multifaceted approaches. In light of this, this paper analyzes comprehensive strategies for improving energy efficiency by utilizing key factors related to computational performance from the Green500 benchmarking metric data. A synergistic approach involving technical innovation, operational adjustments, and policy-level interventions is essential to ensure sustainable development in the supercomputing domain.

DEA models are divided into several models depending on the assumptions and perspectives on efficiency, and the most widely used models are the Constant Returns to Scale (CRS) model called CCR and the Variant Returns to Scale (VRS) model called BBC. These models are divided into multiplier models, ratio models, and envelope models according to the optimization method of the objective function representing efficiency, and this paper uses an envelope model that provides various information in terms of interpreting the results. First, the input-directed CCR envelope model is analyzed by focusing on Slack, which implies additional room for improvement to reach the state above the efficiency boundary. It can be described as a model that finds the smallest input level θ by reducing all m inputs by a certain percentage while maintaining an output level that is at least equal to or greater than the current output level. A mathematical representation of this model in terms of linear programming would look like Equations 1-4. x, y are the inputs and outputs of a particular DMU (Decision Making Unit), and λ is the weight assigned to each DMU. After creating an efficiency boundary using the linear combination of inputs $\sum_{j=1}^{n} x_{ji}\lambda_j$ and outputs $\sum_{j=1}^{n} y_{jr}\lambda_j$, we impose the constraint that the output of a DMU is less than or equal to the output of this efficiency boundary. And the input is constrained to be greater than or equal to the input above the efficiency boundary. At this time, the value that reduces the distance from the efficiency boundary is calculated as the efficiency score θ.

$$Min\ \theta \tag{1}$$
$$s.t.\ x_{ki}\theta \geq \sum_{j=1}^{n} x_{ji}\lambda_j \ \ (i = 1,2,\cdots,m) \tag{2}$$
$$\sum_{j=1}^{n} y_{jr}\lambda_j \geq y_{kr} \ \ (r = 1,2,\cdots,s) \tag{3}$$
$$\lambda_j \geq 0. (j = 1,2,\cdots,n) \tag{4}$$

In other words, the objective function of the DMU is the problem of minimizing θ. If $\theta = 1$, the DMU is efficient because there are no more inputs to reduce, and if $\theta < 1$, the DMU is inefficient because there is room to reduce inputs.

Next, we need to maximize the sum of the input and output slack while maintaining the efficiency level. This is done by using the current input and output levels and the difference between input and output above the efficiency boundary to satisfy the following conditional Equations 5-7. The margins of the i-th input and j-th output factors are expressed as s_i^-, s_r^+. If θ is 1 and w is 0, it is judged as a DMU that is efficient and satisfies strong efficiency with no slack.

$$Max\ w = \sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \tag{5}$$
$$s_i^- = x_{ki}\theta^* - \sum_{j=1}^{n} x_{ji}\lambda_j \tag{6}$$
$$s_i^+ = \sum_{j=1}^{n} y_{jr}\lambda_j - y_{kr} \tag{7}$$

Second, the input-directed BBC envelope is given by Equations 8-12. Compared to the CCR envelope, the constraint $\sum_{j=1}^{n} \lambda_j = 1$ is added.

$$Min\ \theta_k - \epsilon(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+) \tag{8}$$
$$s.t.\ \theta x_{ik} - \sum_{j=1}^{n} x_{ji}\lambda_j - s_i^- \ (i = 1,2,\cdots,m) \tag{9}$$
$$y_{rk} - \sum_{j=1}^{n} y_{jr}\lambda_j + s_r^+ \ \ (r = 1,2,\cdots,s) \tag{10}$$
$$\sum_{j=1}^{n} \lambda_j{=}1 \tag{11}$$
$$\lambda_j \geq 0, s_i^- \geq 0, s_i^+ \geq 0 \tag{12}$$

Imposing the constraint $0 \leq \sum_{j=1}^{n} \lambda_j \leq 1$ assumes a declining return to scale (DRS) and imposing $\sum_{j=1}^{n} \lambda_j \geq 1$ assumes an increasing return to scale (IRS) [7-10].

The Scale efficiency (SE) is a useful metric that can be obtained from the BBC model and can be defined as the following Equation 13. SE takes on values greater than 0 and less than 1, with values closer to 1 indicating no efficiency losses due to scale. In the case of harvest invariance to scale, the SE is equal to 1, indicating that the DMU is at the most optimal scale level.

$$SE_k = CCR(\theta_k^*)/BCC(\theta_k^*) \tag{13}$$

The super-efficiency model (SEM) solves the problem that the previous models have: when many DMUs have an efficiency value of 1, it is not possible to evaluate the relative value between them. The super-efficiency model can be applied to both CCR and BBC models, but only the SEM of the BBC model will be analyzed [11].

3. Energy Efficiency Analysis

3.1. Research Procedure

The research procedure was meticulously designed to ensure a comprehensive analysis of energy efficiency in supercomputers. The foundational step involved obtaining data from the Green500 list, which ranks the most energy-efficient supercomputers globally. Variables relevant to computing performance and energy efficiency were meticulously selected from this dataset, ensuring alignment with the study's objectives. The analysis focused on 10 supercomputers with high priority, meeting specific conditions based on their architecture, operational characteristics, and energy efficiency metrics. The detailed selection process is elaborated upon in the Research Data chapter. The methodological framework employed Data Envelopment Analysis (DEA), specifically using the BCC (Banker-Charnes-Cooper) model, to evaluate the relative efficiency of Decision-Making Units (DMUs). The DEA approach identified inefficient DMUs and provided quantitative insights into potential improvement measures. For efficient DMUs, a Structural Equation Model (SEM) analysis was conducted to explore deeper relationships among variables and derive actionable implications for enhancing energy efficiency. The DEA analysis was executed using Banxia Software's Frontier Analyst program, a specialized tool for benchmarking and efficiency analysis. This ensured precise modeling of input-output relationships and reliable efficiency scores. To support this, descriptive statistics and correlation analysis were performed using IBM SPSS, enabling a clearer understanding of variable distributions and interdependencies. These complementary methods facilitated a robust, multi-dimensional analysis of energy efficiency and performance metrics in supercomputers.

3.2. Research Data

Table 1 presents the variables selected for the efficiency analysis of supercomputers. These variables were derived from the Green500 list, published in November 2024 [12] which provided a rich dataset encompassing critical attributes of energy-efficient supercomputers. A total of seven variables were chosen, carefully categorized into input and output variables, and coded from A to G for clarity and effective presentation in the subsequent analysis. The input variables were selected based on their direct relationship to power consumption and computational performance.

- Total Number of Cores: As the primary electricity consumers in a supercomputer, the total number of cores directly correlates with energy usage.
- Number of Accelerator Cores: These cores, often GPUs or specialized accelerators, are known for their high power consumption and their critical role in enhancing computational speed.
- Number of Cores per Socket: This variable reflects the density of computational resources and their impact on power distribution and thermal efficiency.
- Data Processing Speed of Cores: High processing speeds are associated with greater computational throughput but also higher energy requirements.
- Power Supply Capacity (Power): The size of the power supply indicates the system's capability to sustain operations under peak loads, ensuring that energy demands are consistently met.
- Output Variables:
  Two key output variables were selected to evaluate performance and energy efficiency comprehensively:
- Energy Efficiency: This metric represents the ratio of computational work done per unit of energy consumed, providing a direct measure of efficiency.
- Rmax Metric: As a widely recognized performance indicator in HPC systems, Rmax measures the maximum achievable performance (in PFLOPS) under ideal conditions. Including Rmax enabled the study to examine the interplay between energy efficiency and performance.
- Selection of Decision-Making Units (DMUs):

Table 2 details the DMUs chosen for the analysis. From the top 100 systems listed in the Green500, 10 supercomputers were selected based on specific criteria. These criteria prioritized systems with a cluster architecture and those equipped with both CPUs and accelerators, as such configurations are representative of modern HPC systems. The selection also considered the ranking of energy-efficient systems, ensuring that the analysis captured a diverse range of high-performing and energy-efficient supercomputers. The selected DMUs included notable systems such as JEDI, JETI, and Henri, which exhibit cutting-edge architectures and operational strategies.

By combining input variables that represent energy-consuming components with output variables that reflect energy efficiency and performance, this research provides a nuanced understanding of the factors influencing supercomputer efficiency. Additionally, focusing on systems with hybrid architectures enables insights into the unique challenges and opportunities presented by modern HPC designs. The selection process not only ensures relevance to the study's objectives but also lays the groundwork for robust, data-driven recommendations for improving energy efficiency in the HPC sector.

Table 1.
Variable list.

	Name	Unit	Code
Input	Total cores	Ea	A
	Accelerator cores	Ea	B
	Power	KW	C
	Processor speed	MHz	D
	Cores per socket	Ea	E
Output	Rmax	TFlop/s	F
	Energy efficiency	GFlops/Watts	G

Table 2.
DMU list.

DMU	Rank	Name
1	1	JEDI
2	6	JETI - JUPITER exascale transition instrument
3	8	Henri
4	27	Snellius phase 3 GPU
5	28	CEA-HE
6	31	MareNostrum 5 ACC
7	41	Gefion
8	49	Dhabi Supercomputer
9	60	MeluXina - accelerator module
10	66	JUWELS booster module

Table 3 presents the descriptive statistics for seven datasets (A–G), including the sample size (N), range, minimum and maximum values, mean, standard deviation, and variance. These statistics summarize the central tendency, dispersion, and variability of the data.

Table 3.
Descriptive Statistics.

	N	Range	Min.	Max.	Avr	SD
A	10	654752.0	8288.0	663040.0	233516.8	227410.1
B	10	583968.0	7392.0	591360.0	189019.2	193017.7
C	10	4114.8	44.1	4158.9	1115.6	1271.1
D	10	1000.0	2000.0	3000.0	2640.0	347.9
E	10	48.0	24.0	72.0	48.8	20.1
F	10	172418.0	2882.0	175300.0	47011.0	54171.7
G	10	47.7	25.0	72.7	49.2	16.5

3.3. Analysis Results

The input-oriented efficiency analysis results for improving energy efficiency are shown in Table 4. The BCC analysis results indicate that only DMU5 is inefficient, with an efficiency score of 0.915. Its reference group (peers) includes DMU2, 7, 8, and 9, and the RTS is in an IRS (Increasing Returns to Scale) state.

Table 4.
DEA(BCC) Results (Input-oriented model).

DMU	Name	TE	SE	SEM	Peers
1	JEDI	1.000	CRS	1.000	-
2	JETI	1.000	CRS	1.000	-
3	Henri	1.000	CRS	0.236	1
4	Snellius phase 3 GPU	1.000	IRS	0.102	2, 3, 6, 7, 9
5	CEA-HE	0.915	IRS	0.092	2, 7, 8, 9
6	MareNostrum 5 ACC	1.000	CRS	1.000	-
7	Gefion	1.000	CRS	0.122	3, 6, 8
8	Dhabi supercomputer	1.000	IRS	0.109	3, 9
9	MeluXina - Accelerator module	1.000	IRS	0.113	3, 6, 7
10	JUWELS booster module	1.000	IRS	0.133	3, 6

Therefore, DMU5 should expand its scale to improve efficiency, as the increase in outputs would be larger than the increase in inputs. Looking at the SE values, which indicate the relative efficiency between efficient groups, DMU1, 2, and 6 were found to be the most efficient. In contrast, DMU3 (0.236) and DMU10 (0.133) showed significant differences in efficiency values compared to these efficient DMUs. Therefore, DMU3, 4, 7, 8, 9, and 10 also need to examine their Potential Improvement (PI%) to enhance efficiency.

The output-oriented efficiency analysis results are shown in Table 5. Similar to the input-oriented analysis, DMU5 was found to be inefficient (reference group: 1, 2). The RTS is in a DRS (Decreasing Returns to Scale) state, meaning efficiency can be improved by reducing the scale. Looking at the SE (Scale Efficiency) values, there was a significant difference between the efficient DMUs (3, 6, 7, 8, 9, 10) and the inefficient DMUs (1, 2, 4), which is similar to the input-oriented results. Therefore, it is also meaningful to check Potential Improvement (PI%) to enhance efficiency.

Table 5.
DEA(BCC) results (Output-oriented model)

DMU	Name	TE	SE	SEM	Peers
1	JEDI	1.000	CRS	0.111	2, 3
2	JETI	1.000	CRS	0.126	1, 5, 6
3	Henri	1.000	CRS	1.000	-
4	Snellius phase 3 GPU	1.000	DRS	0.118	3, 6, 9
5	CEA-HE	0.822	DRS	0.082	1, 2
6	MareNostrum 5 ACC	1.000	CRS	1.000	-
7	Gefion	1.000	CRS	1.000	-
8	Dhabi supercomputer	1.000	DRS	1.000	-
9	MeluXina - Accelerator module	1.000	DRS	1.000	-
10	JUWELS booster module	1.000	DRS	1.000	-

In efficiency analysis, the contribution of input and output variables quantifies the extent to which each input and output variable contributes to overall efficiency. The contribution rates of input and output variables for the inefficient DMU5 are shown in Table 6. The benchmarking peers were selected as JEDI and JETI. In the input-oriented case, the input variable D had the highest contribution to efficiency at 61.97%, followed by C at 33.99% and B at 4.04%. Variables A and E showed no contribution. For the output variable, G contributed 100%. In the output-oriented case, the contribution of input variables showed that C had a significantly higher value at 97.15% compared to D at 2.84%. The output variable G again contributed 100%, as in the input-oriented case. Results where a single variable exerts extreme influence are not considered.

Table 6.
Input, output contribution.

DMU	Variable		Contributions (%)		Peers
			Input-Oriented	Output-Oriented	2, 7, 8, 9
5	Input	A	-	-	
		B	4.04	-	
		C	33.99	97.15	
		D	61.97	2.84	
		E	-	-	
	Output	F	-	-	1, 2
		G	100.00	100.00	

The Potential Improvement (PI) results are shown in Table 7. When considering only the analysis results, the input-oriented model shows that all input variables have potential for improvement through reduction, with E showing the largest improvement potential at -23.59%. For the output-oriented model, it would be ideal to increase F by 30.85% and G by 21.62% compared to the reference group.

Table 7.
PI results (DMU 5).

Variable	Input-oriented model			Output-oriented model		
	Actual	Target	PI (%)	Actual	Target	PI (%)
A	251856	202650	-19.54	251856	238391	-5.35
B	72	66	-8.5	72	72	-
C	1233	1128	-8.5	1233	1233	-
D	3000	2745	-8.5	3000	3000	-
E	389232	297423	-23.59	389232	78225	-5.35
F	52.17	57.46	10.14	52.17	68.26	30.85
G	64320	64320	-	64320	78225	21.62

From the previous efficiency analysis results, it can be determined that many DMUs have established efficient operational plans. However, since this analysis performs a relative comparison of efficiency among groups, deriving a priority ranking of efficiency among the efficient groups holds additional significance. Therefore, SEM was conducted for the efficient DMUs. Table 8 and 9 show the SEM results. In the input-oriented SEM, DMU 3, 7, 9, and 10 showed relatively high potential for improvement. Taking DMU 3 as an example, the improvement levels for input variables A and E were significantly high at 136.29% and 125%, respectively. Similarly, the output variable G for DMU 9 and 10 showed high values of 127.97% and 145.17%, respectively.

Table 8.
PI results (Input-oriented, SEM).

	3	4	7	8	9	10	AVR
A	136.29	1.89	21.94	-8.4	-37.94	-63.3	8.41
B	71.43	-0.95	20.09	-8.05	-35.91	-63.63	-2.84
C	52.73	1.89	-5.81	8.62	12.8	-41.72	4.75
D	7.14	1.89	21.94	8.62	12.8	-4.27	8.02
E	125	1.89	-14.84	-50	12.8	33.33	18.03
F	56.28	0	4.17	2.85	72.35	0	22.61
G	11.22	19.32	0	46.01	127.97	145.17	58.28

In the output-oriented SEM, DMU 1 showed relatively high improvement potential for input variables A and E at -21.76% and -54.54%, respectively. Similarly, the output variables F and G for DMU 2 showed values of -20.82%. However, since negative directional improvement for G is not appropriate, it is not considered.

Table 9.
PI results (Output-oriented, SEM).

	1	2	4	AVR
A	-21.76	-4.89	0	-8.88
B	-6.04	-0.84	-0.78	-2.55
C	0	0	-5.3	-1.77
D	-6.54	-1.48	0	-2.67
E	-54.54	-3.51	0	-19.35
F	-3.32	-20.82	-15.19	-13.11
G	-10.02	-20.82	10.52	-6.77

The comparative analysis of the original Green500 rankings and the efficiency rankings of the DMUs is shown in Table 10. DMU6 to DMU10 were positioned in lower ranks in the Green500 due to their high energy efficiency. However, the DEA efficiency results, which considered detailed performance factors, showed that they achieved the highest efficiency scores. On the other hand, DMU1, 2, 4, and 5, despite being ranked within the top 30 for high energy efficiency, were relatively lower in the DEA results. This highlights the discovery of new opportunities to improve energy efficiency and emphasizes the need to consider multiple performance indicators comprehensively rather than solely relying on energy consumption as a measure of efficiency.

Table 10.
Efficiency rank.

DMU	Name	Green500	Rank
1	JEDI	1	9
2	JETI - JUPITER exascale transition instrument	6	7
3	Henri	8	1
4	Snellius phase 3 GPU	27	8
5	CEA-HE	28	10
6	MareNostrum 5 ACC	31	1
7	Gefion	41	1
8	Dhabi Supercomputer	49	1
9	MeluXina - Accelerator module	60	1
10	JUWELS booster module	66	1

4. Conclusion

Until now, it has been generally accepted that as the computational resources of supercomputers increase, energy efficiency tends to decrease slightly. However, significant limitations have persisted in identifying and explaining the specific factors influencing this trade-off, while the reliability of predictive models has remained limited due to insufficiently comprehensive analyses. In this study, we addressed these gaps by leveraging extensive performance data from supercomputers to identify seven critical factors influencing energy efficiency. By quantitatively deriving improvement measures, we have introduced a novel, data-driven approach to enhance the energy efficiency of supercomputing systems. The most noteworthy contribution of this research lies in its comprehensive framework for addressing energy efficiency improvements at the component level. Unlike previous approaches, which primarily relied on singular strategies, this study highlights the interconnectedness of multiple factors influencing energy efficiency and their relative contributions. This systemic understanding not only allows for targeted interventions to improve specific components but also provides an integrated perspective that is essential for optimizing energy-efficient systems during the design phase of next-generation supercomputers. Despite these contributions, the study is not without its limitations. Due to the inherent characteristics of Data Envelopment Analysis (DEA), the findings are based on a relative efficiency evaluation among Decision-Making Units (DMUs), which inherently introduces a degree of relativity. The scope of the analysis was also limited to 10 supercomputers, selected based on specific criteria, which may restrict the generalizability of the results. Future research will address this limitation by expanding the scope to include data from the Green500 spanning at least five years, enabling a longitudinal analysis that captures trends and patterns over time. Moreover, this study did not account for external factors beyond the Green500 indicators that could significantly influence energy efficiency. For instance, elements requiring detailed power system analysis—such as heat loss from line heating, voltage drops during frequency regulation in power conversion systems, and cooling inefficiencies—were not controlled. These factors represent important avenues for future research to further refine the understanding of energy efficiency in supercomputers.

Nonetheless, the significance of this study lies in its comprehensive attempt to analyze the factors affecting supercomputer performance and energy efficiency. By identifying key variables and quantifying their impact, this research provides a foundation for developing systematic strategies to mitigate the massive power consumption associated with supercomputing systems. In doing so, it contributes to the broader field of systems engineering by offering practical insights for optimizing the design and operation of increasingly sophisticated supercomputers. Looking forward, the results of this research are expected to have substantial implications for both academia and industry. By addressing the challenge of excessive power consumption, this study aligns with the global push toward sustainable computing and the development of eco-friendly artificial intelligence (AI) environments. We anticipate that these findings will not only help reduce the operational costs and environmental footprint of supercomputers but also accelerate technological progress in AI and high-performance computing. Ultimately, this research lays the groundwork for the design of future supercomputing systems that are both highly efficient and environmentally sustainable, fostering innovation in an era where computational demands continue to rise exponentially.

References

[1] L. G. León-Vega, N. Tosato, and S. Cozzini, "A comprehensive analysis of process energy consumption on multi-socket systems with GPUs," *arXiv preprint arXiv:2409.04941*, 2024. https://doi.org/10.48550/arXiv.2409.04941

[2] A. Prieto, B. Prieto, J. J. Escobar, and T. Lampert, "Evolution of computing energy efficiency: Koomey's law revisited," *Cluster Computing*, vol. 28, no. 1, p. 42, 2024. https://doi.org/10.1007/s10586-024-04767-y

[3] D. Carastan-Santos, G. Da Costa, M. Poquet, P. Stolf, and D. Trystram, "Light-weight prediction for improving energy consumption in HPC platforms," in *European Conference on Parallel Processing, Springer. https://doi.org/10.1007/978-3-031-69577-3_11*, 2024, pp. 152-165.

[4] B. Erol, D. May, and B. Wunderle, "Numerical Investigation of a Novel Lid Design for Automotive HPC Cooling," in *2024 30th International Workshop on Thermal Investigations of ICs and Systems (THERMINIC). https://doi.org/10.1109/EuroSimE60745.2024.10491471*, 2024: IEEE, pp. 1-10.

[5] F. Naduvilakath-Mohammed, M. Lebon, and A. Robinson, "Numerical modelling of a hybrid vapor compression refrigeration assisted closed loop liquid cooling system for high-performance computing systems," in *Journal of Physics: Conference Series, IOP Publishing. https://doi.org/10.1088/1742-6596/2766/1/012078*, 2024, vol. 2766, no. 1, p. 012078.

[6] C. Song, "Optimizing Management Strategies for Enhanced Performance and Energy Efficiency in Modern Computing Systems," *Academic Journal of Sociology and Management,* vol. 2, no. 4, pp. 57-64, 2024. https://doi.org/10.5281/zenodo.12760026

[7] V. A. Novikov, O. Sapun, and L. Shipulina, "Accounting the scale and synergies in the DEA-Analysis," *LogForum,* vol. 12, no. 2, pp. 123--128, 2016. https://doi.org/10.17270/J.LOG.2016.2.2

[8] D. Niu, R. Wang, X. Zhang, and W. Song, "Comparative studies on efficiency of China's regional innovation system on the basis of cooperation measurement model," *Journal of Emerging Trends in Economics and Management Sciences,* vol. 4, no. 2, pp. 147-157, 2013. https://doi.org/10.10520/EJC135867

[9] M. E. F. E. Atta, M. A. Baafi, F. Liu, and K. Zeng, "Dynamic operating efficiency and its determining factors of listed real-estate companies in China: A hierarchical slack-based DEA-OLS approach," *International Journal of Finance & Economics,* vol. 26, no. 3, pp. 3352-3376, 2021. https://doi.org/10.1002/ijfe.1965

[10] W. W. Cooper, L. M. Seiford, and J. Zhu, *Handbook on data envelopment analysis*. US, New York: Springer. http://dx.doi.org/10.1007/978-1-4419-6151-8_1, 2011.

[11] H. Shim and J. Hahm, "A study on the operational efficiency of national supercomputer resources based on DEA model: focusing on specialized center resources by field," *International Journal of Asian Business and Information Management,* vol. 15, no. 1, pp. 1-14, 2024. https://doi.org/10.4018/IJABIM.336561

[12] Top500.org, "TOP500 supercomputer sites," Retrieved: https://www.top500.org, 2024.