New Energy Exploitation and Application
https://ojs.ukscip.com/index.php/neea

Review

Sustainable and Efficient Energy Use in Data Center Cooling: Techniques and Innovations

Okan Şen , Obed N. Onsomu * and Bulent Yeşilata

Energy Systems Engineering, Ankara Yıldırım Beyazıt University, Ankara 06090, Turkey
* Correspondence: obedonsomu@gmail.com

Received: 6 October 2025; **Revised:** 10 November 2025; **Accepted:** 18 November 2025; **Published:** 12 December 2025

Abstract: The rapid expansion of high-performance computing (HPC), artificial intelligence (AI), and cloud-based services has significantly increased the energy demand of modern data centers. Among the key challenges in maintaining operational efficiency is managing excess heat from densely packed computing equipment. While liquid-cooling technologies offer superior thermal performance, they are often operated conservatively, leading to excessive energy use through overcooling. This study investigates overcooling in data centers using operational data from the Frontier supercomputer, currently the world's fastest publicly available exascale system. A linear regression model was developed to predict baseline coolant return temperatures using compute power and waste heat as inputs, and its performance was validated using standard regression metrics (R^2 = 0.357, MAE = 2.76 °C). Overcooling was identified when actual return temperatures were at least 1.5 °C below the predicted baseline. The analysis revealed that approximately 6.9% of the cooling effort could be reduced without compromising thermal safety margins. The study also translates the energy implications of avoidable overcooling into public-scale usage equivalents, showing potential annual savings exceeding 2.5 million kilowatt-hours. These findings demonstrate the potential of AI-assisted thermal modeling as a lightweight and interpretable method to improve cooling efficiency, reduce operational costs, and support sustainable data center management.

Keywords: Data Centers; Virtual Power Plant; Power Usage Effectiveness (PUE); Supercomputer; Cooling Systems; Waste Heat Recovery; Machine Learning; Energy Efficiency

1. Introduction

As the backbone of the modern digital ecosystem, data centers enable the computational, storage, and networking capabilities required for critical applications such as artificial intelligence (AI), high-performance computing (HPC), big data analytics, and cloud services. The increasing demand for processing power, driven by scientific research, real-time services, and industrial automation, has significantly elevated the energy footprint of these facilities. One of the most critical challenges associated with this energy demand is thermal management. High-density server architectures generate substantial heat, requiring continuous and reliable cooling to ensure system stability and prevent hardware degradation. Although liquid-cooling systems have become the preferred solution for next-generation data centers and supercomputers due to their superior thermal conductivity and efficiency, their operation is often conservative leading to overcooling, where more cooling is delivered than needed.

Overcooling not only results in energy waste but also limits opportunities for heat recovery and optimization. Yet, detecting and quantifying overcooling under dynamic workload conditions remains a complex task, especially in HPC environments where thermal behavior is non-linear and workload-sensitive. This study addresses this issue by proposing a data-driven framework for overcooling detection using real operational data from the Frontier supercomputer, located at Oak Ridge National Laboratory. Frontier, the world's first exascale system, provides high-resolution measurements of power consumption, coolant flow, and thermal states over a full operational year. By training a lightweight AI-based regression model to estimate the expected return temperature of the coolant from compute power and waste heat, the study identifies instances of avoidable overcooling and evaluates their energy impact. Furthermore, the analysis contextualizes the avoidable energy loss in terms of household electricity usage and solar PV generation equivalents, offering a broader perspective on the sustainability implications. This approach provides a practical and data-driven pathway for improving energy efficiency beyond existing model-based optimization or CFD-driven analyses. Through this approach, the paper demonstrates how machine learning can be effectively integrated into thermal diagnostics to enhance energy efficiency, reduce operational costs, and support sustainable data center management.

Unlike previous studies that primarily focused on predictive control or thermal simulation in data center environments, this work uniquely emphasizes the quantification of overcooling using real operational data from a high-performance computing (HPC) system. By developing a lightweight and interpretable AI-based regression model trained on Frontier's dataset, the study contributes a novel framework for identifying and evaluating unnecessary cooling efforts under real-world conditions. This approach provides a practical and data-driven pathway for improving energy efficiency beyond existing model-based optimization or CFD-driven analyses.

This article is divided into the following sections, Section 1 introduces the background of high-performance computing, Section 2 describes the problem specific to this work. Section 2 gives brief evaluation of recent work on cooling of datacenters, Section 4 presents innovative cooling techniques while Section 5 presents a case study detailing application of AI technique on cooling of HPC, followed by concluding remarks in Section 6.

2. Problem Description

Modern data centers house thousands of servers, storage units, and networking devices powered by electricity and generating substantial amounts of heat during operation. The rapid expansion of workloads driven by artificial intelligence (AI), high-performance computing (HPC), and cloud services has intensified the need for effective thermal management. Between 2010 and 2018, global IP traffic increased tenfold, data center storage capacity grew 25-fold, and the number of compute instances rose more than sixfold [1].

Cooling systems are essential to prevent thermal degradation, maintain system reliability, and ensure service continuity. Industry guidelines, such as those from the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), recommend maintaining server inlet temperatures between 18 °C and 27 °C and relative humidity between 20% and 80% [2]. However, many data centers operate well below these limits, often maintaining server room temperatures around 19 °C to 21 °C, in an effort to avoid any thermal risk.

Figures 1 and **2** illustrate this relationship based on data collected from multiple Google data centers. **Figure 1** shows the probability of latent sector errors as a function of temperature, while **Figure 2** presents the correlation between temperature and hard disk failure rates [3].

Here, an alternative strategy called the "expected failure" approach has been adopted by some hyperscale data center operators. Rather than overcooling to prevent hardware failures, these systems allow servers to operate at higher inlet temperatures often near 27 °C, increasing the risk of failure in exchange for lower cooling costs. Failed servers are replaced rapidly, and system reliability is maintained through redundancy and software-based fault tolerance. This model emphasizes energy savings over hardware preservation and illustrates the trade-offs involved in thermal management strategies.

Conservative approach to cooling maintaining lower than necessary temperatures is known as overcooling. While it can improve perceived safety margins, it results in unnecessary energy consumption, increased operational costs, and reduced sustainability. The importance of thermal control goes beyond efficiency. Elevated operating temperatures have been empirically linked to increased hardware failure rates, particularly in storage devices. Latent sector errors, which render portions of disks unreadable, and complete hard disk failures both become more probable at higher temperatures. These risks reinforce the widespread tendency to overcool, even when not strictly necessary.

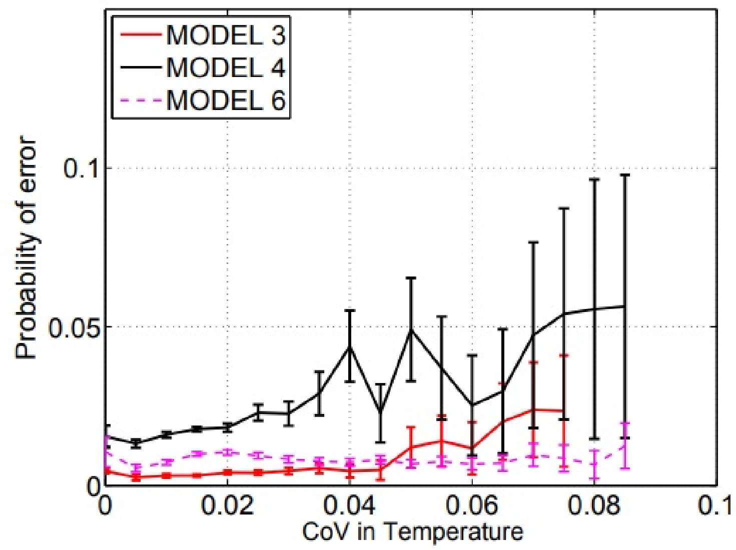

Figure 1. Probability of latent sector errors occurrence.

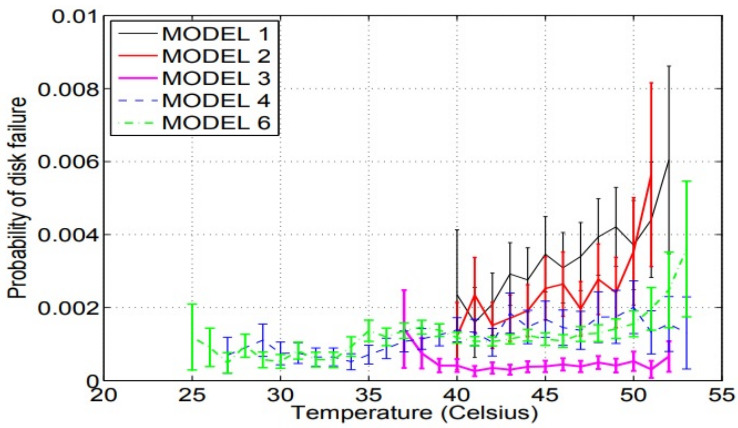

Figure 2. Probability of hard disk failures.

Although such data-driven insights justify the need for reliable cooling, they also raise the question: how much cooling is truly necessary? The answer is rarely addressed using quantitative methods. This gap leads to overcooling becoming a safe and inefficient default strategy.

This study frames overcooling as a measurable energy inefficiency and proposes a machine-learning-based approach for detecting it under real operational conditions. Using actual power and thermal measurements from the Frontier supercomputer, the analysis estimates thermal baselines and identifies instances of excessive cooling. By quantifying the energy implications of overcooling, the study highlights opportunities for improving efficiency while maintaining hardware safety.

2.1. Cooling Technologies

Cooling technologies play a crucial role in ensuring the efficient operation of data centers, with significant implications for energy efficiency and sustainability. This section provides a comprehensive overview of the various cooling systems employed in data centers. These technologies can be categorized based on the state of the cooling fluid, the heat transfer methods used, and the overall system design. Each technology offers unique advantages, limitations, and areas of application. The classification in **Figure 3** outlines the organization of these technologies and highlights the topics covered in this section. Following the figure, each category will be discussed in detail.

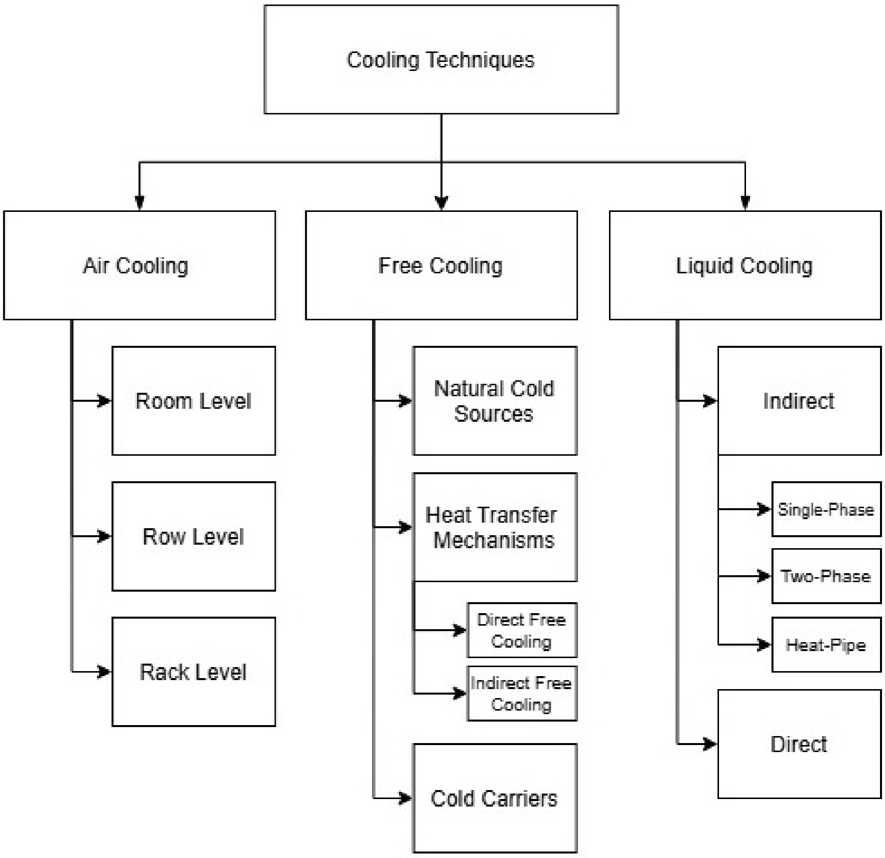

Figure 3. Classification of cooling techniques.

2.2. Air-Cooling

In most of data centers, racks are arranged back-to-back to form cold and hot aisles. This design aims to avoid mixing cold and hot air. Air from the hot aisle is pulled into the cooling equipment or an air handling unit, then redistributed to the cold aisle. Two options of this design is shown in **Figure 4** [4].

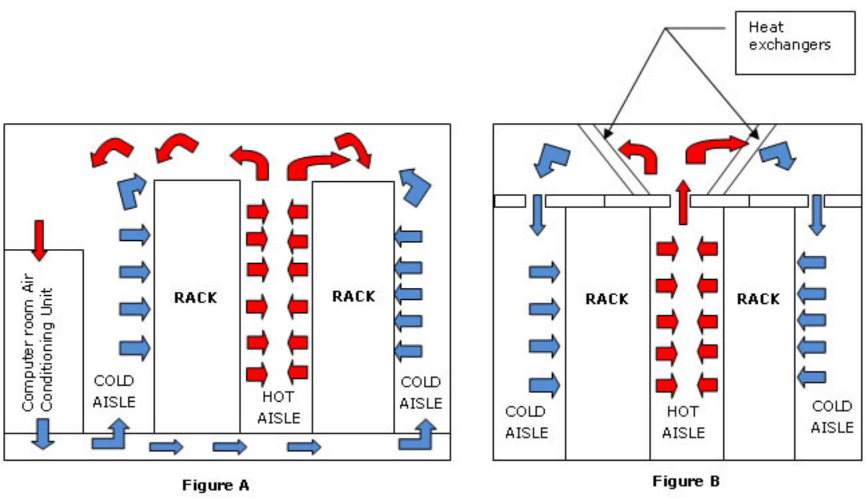

Figure 4. Back-to-back cooling system design.

Red arrows indicate the hot aisle, and blue arrows mark the cold aisle. The cold aisle delivers chilled air straight to the front of the racks, thus to the servers, while the hot aisle collects the warm air released by the servers after it has served its cooling function. In the basic hot aisle containment, cooled air is supplied through a space beneath the raised floor, creating positive pressure, while the enhanced hot aisle employs heat exchangers. Air cooling relies on the circulation of cool air through server racks to dissipate heat. It is one of the most widely used and cost-effective methods, though it can be less efficient in high-density data centers. Its primary benefits include ease of maintenance and reasonable operating costs [5]. The fundamental working principle of air cooling technology is depicted in **Figure 5**.

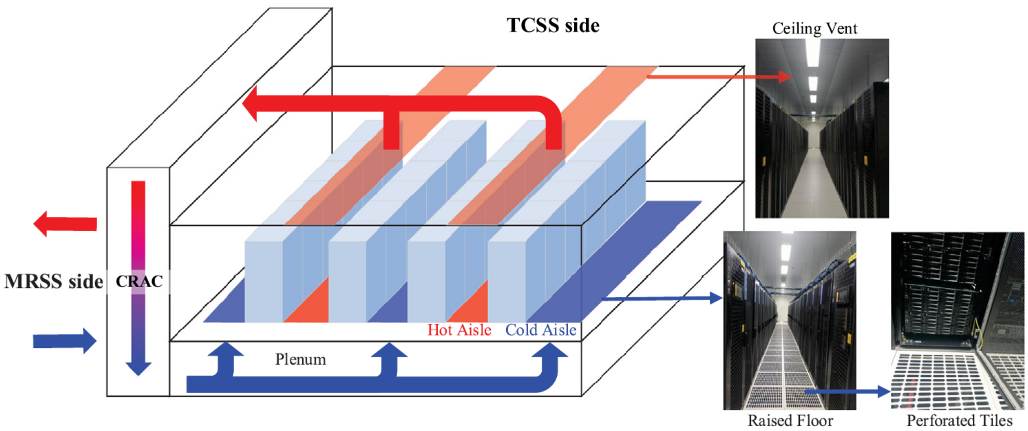

Figure 5. The basic mechanism of the room-level air-cooling technology under the raised-floor design [6].

Data center cooling can be categorized into three primary levels: room, row, and rack. Each level offers a unique approach to managing heat, tailored to different densities and configurations within the facility. These methods collectively ensure optimal temperature control and energy efficiency in data centers.

2.2.1. Room-Level Cooling

In data centers focuses on dissipating heat within the entire room, often utilizing raised floors for cold air supply. This system typically includes a plenum beneath an elevated floor and ceiling vents for warm air exhaust. Cold air is distributed through perforated tiles on the floor and directed into cold aisles, while warm air, heated by server racks, is expelled through the hot aisles and circulated back to the ceiling vents [7]. However, inefficiencies arise due to issues such as hot air recirculation, where warm air mixes with cold air, and cold air bypass, where excessive cooling air escapes unused [8]. These inefficiencies lead to uneven cooling and energy waste. To address these challenges, airflow containment systems, such as hot aisle containment and cold aisle containment, have been developed. These systems use physical barriers to prevent the mixing of cold and hot air, significantly improving cooling efficiency [9]. While these approaches mitigate some inefficiencies, room-level cooling remains limited by its static components, high implementation costs, and uncertainties in airflow distribution. Nonetheless, its flexibility and ease of maintenance make it suitable for data centers with lower equipment density.

2.2.2. Row-Level-Cooling

Places Computer Room Air Conditioning (CRAC) units close to ICT devices, significantly shortening airflow paths compared to room-level cooling [10]. Two common configurations are inter-row cooling, where CRAC units are positioned between racks, and overhead cooling, where CRAC units are mounted above rows. Inter-row cooling provides efficient rear-to-front airflow, allowing flexible control of cooling capacity based on server IT load. However, it requires substantial space, which can be challenging to manage. Overhead cooling addresses space constraints by creating an up flow design, although it may face cold air bypass issues due to vertical temperature differences [11].

This method enhances cooling efficiency and allows for easier expansion by attaching cooling equipment to each row of racks. The shorter airflow paths improve airflow predictions and facilitate better rack placement for optimized cooling. However, row-level cooling has higher installation costs than room-level cooling and requires maintenance within the computer room, potentially interrupting operations near ICT devices [12].

2.2.3. Rack-Level Cooling

Involves installing rack coolers directly within the racks, significantly shortening the airflow cycle. Typically, an additional enclosure with a rack cooler is integrated into the rack, separating the internal space into hot and cold aisles through partitioning systems [5]. This setup minimizes the mixing of exhaust and supply air, leading to a more efficient and controllable airflow path [13].

An advantage of rack-level cooling is its adaptability to the varying cooling requirements of ICT devices. For example, blade servers demand higher airflow compared to communication enclosures, and rack-level cooling allows for airflow adjustments based on these specific needs [6]. This dynamic cooling approach optimizes power consumption by distributing cooling capacity according to the actual demand of individual racks, significantly reducing fan power usage. Additionally, the modular nature of in-rack coolers offers efficient use of space, making this method ideal for data centers with space constraints [14].

However, rack-level cooling comes with certain drawbacks. Its installation cost is relatively high, partly due to over-designed cooling capacities, and maintenance requires substantial effort. Despite these challenges, its ability to provide adaptable and localized cooling makes it an efficient solution for modern data centers, particularly in dynamic environments where server workloads and cooling demands fluctuate in real-time [15]. The three types of air-based systems [16] are shown in **Figure 6**.

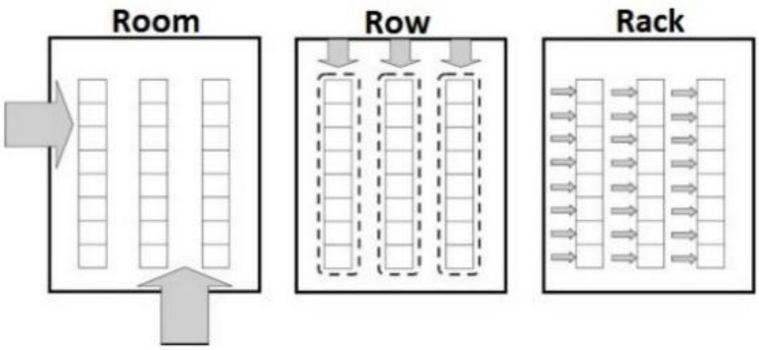

Figure 6. Room Row Rack based design [16].

Both air and water-based cooling systems advocate for cooling the hot equipment directly instead of the entire room. Consequently, newer, more energy-efficient cooling methods have been proposed, such as liquid cooling, nano-fluid cooling systems, and cooling solutions that are integrated directly into the server, rack, or row.

2.3. Liquid-Cooling

Liquid cooling has emerged as a highly efficient alternative to traditional air-cooling methods, particularly in data centers and high-performance computing environments. By leveraging the superior thermal conductivity of liquids compared to air, this technique enables faster and more effective heat dissipation from critical components like CPUs, GPUs, and memory units. Liquid cooling systems are capable of managing higher heat densities, making them ideal for modern data centers that require optimized energy efficiency and performance. As the demand for sustainable and energy-efficient cooling solutions grows, liquid cooling continues to play a pivotal role in reducing energy consumption and ensuring the reliability of IT infrastructure.

2.3.1. Indirect Liquid-Cooling (Mechanical Refrigeration)

Single-Phase

Liquid cooling involves the use of a coolant, typically water or a water-based mixture, that remains in the liquid phase throughout the cooling process. Heat is transferred from the components to the coolant via conduction, and the heated coolant is then circulated to a mechanical chiller or heat exchanger, where it is cooled before being recirculated. This method is widely used due to its simplicity, stability, and effectiveness in managing moderate heat loads. Single-phase systems are particularly suitable for cold plate cooling, where the coolant absorbs heat directly from electronic components [17].

Two-Phase

Liquid cooling leverages the phase change properties of a coolant to achieve highly efficient heat transfer. In this system, the coolant absorbs heat from components and undergoes a phase transition from liquid to vapor. The vaporized coolant is transported to a condenser or mechanical refrigeration unit, where it releases heat and reverts to its liquid state. This cycle enables significant heat dissipation with minimal energy input, making two-phase systems ideal for high-density computing environments. The ability to exploit latent heat during phase transition allows these systems to handle larger thermal loads compared to single-phase methods. The primary focus of improving two-phase cooling lies in the design of efficient porous media and micro-channel radiators at the chip level [17]. However, these systems are not without challenges. Key disadvantages include flow instability issues such as liquid reversal, as well as temperature and pressure fluctuations, which can lead to overheating and potential burnout of surfaces [18]. Addressing these limitations is critical to advancing the reliability and scalability of two-phase cooling methods for widespread deployment in data centers.

Heat-Pipe Cooling

The cooling process uses sealed tubes containing a working fluid that undergoes phase transitions to transport heat from hot to cold regions. Mechanical refrigeration enhances this process by actively cooling the condensation end of the heat pipe, improving its thermal performance. The heat absorbed at the evaporation section turns the fluid into vapor, which moves to the condenser end, where mechanical cooling facilitates rapid condensation. This method combines the efficiency of passive heat pipes with the power of active mechanical systems, making it a versatile solution for a range of applications, from individual server racks to entire data center systems. This type of cooling is shown in **Figure 7**.

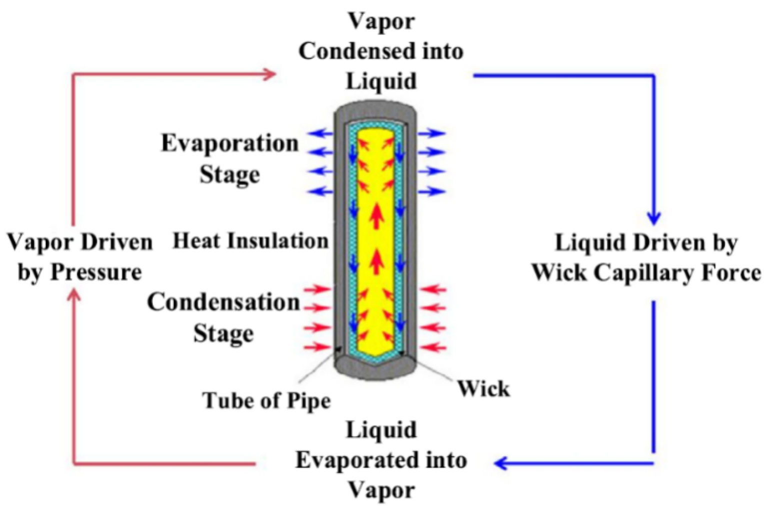

Figure 7. Heat-pipe cooling mechanism [6].

2.3.2. Direct Liquid-Cooling (Terminal Cooling)

Pool Boiling

It is a passive, fully liquid-based cooling method where the electronic components are entirely submerged in a dielectric coolant [19]. When the heat source surface temperature exceeds the coolant's saturation temperature, the liquid begins to boil. Steam bubbles generated at the heat source rise to the cooling tank, where they condense via a water-cooled heat exchanger, effectively dissipating heat from the system. The primary advantage of this method is its simplified design, eliminating the need for sealed pipes, enclosures, and fluid connectors at the server level. This makes pool boiling a flexible and cost-effective cooling solution for server components.

Spray Boiling

In this method, the liquid coolant is atomized into fine droplets by the differential pressure generated at the nozzle plate before reaching the heat source surface [20]. Spray cooling can be implemented in two ways: direct or indirect. The direct method effectively cools server components but comes with the drawback of requiring high-cost maintenance, as servers are typically enclosed in steam chambers with two-phase coolant flow. In contrast, the indirect method relies on heat transfer through a cold plate, providing an alternative approach with potentially simpler maintenance requirements. **Table 1** is a comparative table highlighting the features of Pool Boiling and Spray Boiling cooling methods.

Table 1. Comparison of pool boiling and spray boiling methods.

Feature	Pool Boiling	Spray Boiling
Boiling Mechanism	Surface-to-pool contact	Spray-to-surface contact
Heat Transfer	Natural Convection and boiling	Uniform spray and boiling
Equipment	Simple (pool design is sufficient)	Requires mechanical spray system
Application Area	Low-to-moderate density cooling	High-density cooling

2.4. Free-Cooling

Free cooling leverages natural environmental conditions, such as outside air or water, to cool data centers without relying heavily on mechanical refrigeration. It is highly energy-efficient and helps reduce operating costs in suitable climates. By utilizing natural cold sources, free cooling significantly reduces the overall energy consumption in data centers. As highlighted in recent searches [21], free cooling technologies are generally categorized into air-side, water-side, and heat pipe-based systems. This classification is derived from the traditional cooling technologies namely air-cooling and liquid-cooling. Free cooling technologies based on three key aspects: natural cold sources, cooling carriers, and heat transfer mechanisms. The comprehensive heat transfer framework between natural cold sources and heat sources within a data center is illustrated in **Figure 8**.

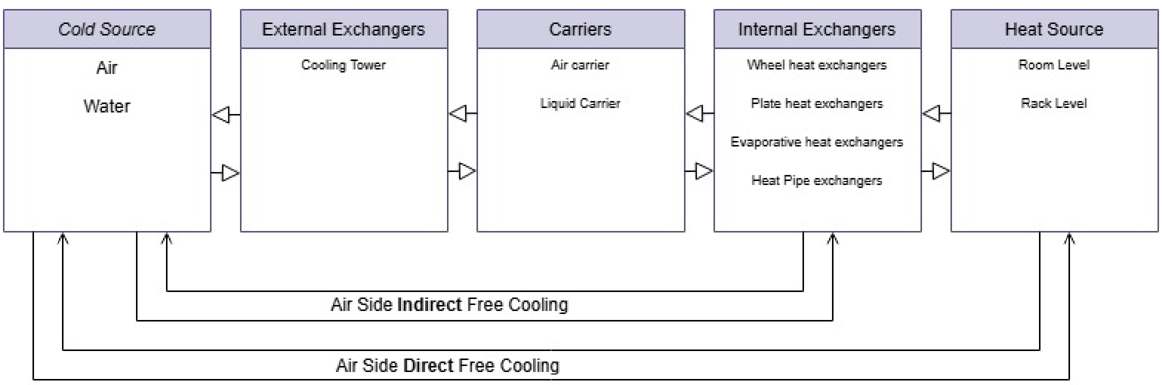

Figure 8. Overview of the free cooling process and examples.

3. Brief Evaluation of Previous Studies

Efficiency of a datacenter is based on some factors [22], as listed below:

(i) The cooling load, or heat output, of equipment
(ii) The heat output of lighting
(iii) The heat output of personnel
(iv) The cooling load of building, if necessary
(v) Any oversizing required due to humidification effects
(vi) Oversizing for redundancy
(vii) Oversizing for potential future growth

Heat generated in a datacenter due to these factors is decreased by datacenter cooling system. Continuously, new techniques and methods are searched to design efficient cooling systems. In scenarios where cooling systems operate with low efficiency, the cost of cooling can surge to as much as 40% of a data center's total energy consumption. The unique operational demands of data centers require robust security measures and substantial cooling capacity. To balance energy savings, cost reduction, and enhanced security, innovative cooling systems that incorporate thermal energy storage technologies have been suggested [23].

A study reviews thermal management in data centers, detailing various cooling strategies [24]. It highlights the use of the open aisle configuration technique, which, while not impacting energy efficiency, increases the electricity consumed by air conditioning. The paper identifies and discusses cooling systems that can absorb energy flux, reduce electricity usage, and enhance energy efficiency. Among these systems are free cooling, liquid cooling, two-phase technologies, and optimizing the building envelope.

A new index, termed the EUED (Energy Usage Effectiveness Design), is proposed for assessing efficiency at the design stage, particularly for systems employing "free cooling" and adiabatic systems [25]. The approach involves comparing equipment performance in the most challenging conditions, utilizing the thermodynamic parameter of enthalpy for calculations. This methodology helps discern differences in energy efficiency across three Brazilian cities. According to the study, the EUED values were 1.245 kW/kW for Curitiba, 1.260 kW/kW for São Paulo, and 1.377 kW/kW for Rio de Janeiro.

In another study focusing on a real-time monitoring system for analyzing thermal efficiency in data centers, 3D visualization technology is introduced for intuitive monitoring of IT equipment [26]. The system employs wireless sensors to gather detailed and comprehensive environmental data from various locations within the data center. It provides both basic and extended real-time metrics to demonstrate the system's effectiveness.

Recent research has shifted focus from relying on device setpoints to analyzing precise temperatures within an actual data center (DC) cluster [27]. This involves using server-level sensors to collect thermal data, which is then examined using comprehensive statistical methods. The study conducts an evaluation of both global and local thermal metrics, aiding in the isolation of risks associated with potential adverse cooling-related factors. Such an approach significantly contributes to the improvement of thermal management in data centers.

In a particular study, a review of prevalent evaluation metrics is conducted, examining their characteristics and areas of application [28]. The data center is segmented into room, row, rack, and server levels to analyze how these metrics apply at each stage. The study proposes a conceptual diagram of thermal performance for these levels as a foundation for future research. It concludes that evaluation metrics are crucial for providing real-time feedback on air management, enabling scalability across various levels from racks to rooms.

A new computational facility being designed for the University of California in Berkeley, CA, aims to serve as both a model for high-performance computing and a showcase for energy efficiency. A paper details its unique design features, which include the flexibility to accommodate both air and liquid cooling [29]. Berkeley's mild climate presents an excellent opportunity for minimizing energy use through free cooling. However, conventional data center methods don't fully exploit this climate for energy savings. The facility's design describes using outside air for cooling most of the year, with provisions for shifting to various liquid cooling configurations in the future.

Efficiency in data centers begins with the efficiency of the machines themselves. Numerous studies suggest that most of the heat in a data center emanates from electronic chips. Therefore, reducing the heat emitted by these chips can significantly decrease the energy required for air conditioning in data centers. One study introduces a Y-shaped liquid cooling heat sink with microchannels, allowing water to diffuse through and cool the periphery effectively [30]. This design is compared to conventional S-shaped liquid cooling heat sinks, revealing a 15.2 °C reduction in peak surface temperature and a 6.3 °C decrease in average temperature, thereby enhancing the overall cooling efficiency.

4. Proposed Innovative Datacenter Cooling Solutions

4.1. Adding Renewable Energy for Data Center Cooling Energy Needs

The carbon footprint of data centers is on the rise because of the substantial energy consumed to power their IT and cooling systems [16]. In contrast, having low emission is another important parameter for efficiency of a datacenter. The lower the greenhouse gases emission, the more efficient a datacenter is. For efficiency a datacenter should not only work within a certain temperature scale, but also should generate less carbon gas. Further and naturally less emission is required during cooling of a datacenter, which can be obtained by special techniques.

Green Data Center term is used for data centers which give no or less harm to environment. In this type of datacenters all IT equipment, building systems and everything is planned due to the requirements of being "green". Consuming renewable energy sources is the main principle with green data centers. Furthermore, the carbon footprint of these facilities can be lessened by implementing energy conservation measures and utilizing energy from renewable sources [16].

4.2. Using Heat Pump for Reuse of Data Center's Waste Heat

The reuse of waste heat is advocated across the entire industry and related sectors. The European Union (EU) has set targets to achieve at least a 32% share of renewable energy and a minimum of 32.5% improvement in energy efficiency by 2030 [31]. In case of datacenters, waste heat recovery and its potential benefits become a recent area of concern also. Waste heat is an ideal source for district heating due to its stable and consistent supply. It can be employed in a range of applications, including heating nearby facilities, regulating temperatures in greenhouses and aquaculture facilities, drying biomass, and district heating. The aim is to enhance energy efficiency and recover as much waste heat as possible.

Waste heat recovery in data centers has two major difficulties: One of them is the scale of the installation and the other is thermal isolation of the data center to facilities which are aimed to utilize the rejected heat. The design approach is to have the heat utilized by a facility which have high heating demand. The scale of which can be a laboratory or central campus heating system [32]. Waste heat can be recovered from hot aisles explained before and directed to air conditioner.

Generally, source temperature wasted by datacenter will be low (between 25–60 °C) for direct utilization.The temperatures utilized in district heating typically range from 70–110 °C, varying with the seasons. [31]. Heat pumps play an important role at this point. A heat pump absorbs thermal energy from lower temperature sources and transfer it to a higher temperature system by a technical process. **Figure 9** shows the network that uses heat pump and utilize waste heat for residents and commercial areas.

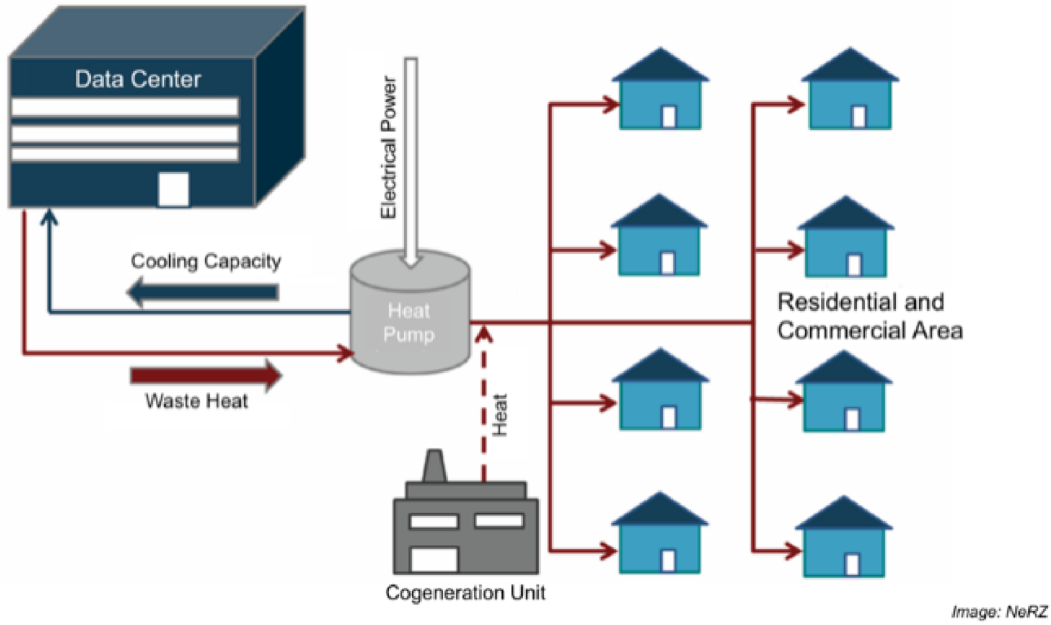

Figure 9. Waste heat utilization for residents.

By the help of heat pump waste heat of a datacenter can be brought to a level of 60 °C or higher. Multistage pumps can be used also which can increase temperature to a significant level. 150 °C or higher temperatures are in possible ranges in this case.

As declared before heat pump solutions have great contribution to the subject. Nonetheless, it necessitates significant investment, thereby requiring long-term planning and economic incentives. To this end, simulations for economic optimization under various conditions are conducted [33]. In this study, the Coefficient of Performance (COP) is determined using estimated figures for electricity costs and district heat prices across different scenarios. The analysis identifies an economically viable alternative that operates without a heat pump.

The efficient cooling of data centers remains a critical area of technological research, especially as digitalization efforts intensify. Key domains identified as influential in the digitization process of cooling systems include the application of Artificial Intelligence (AI), Augmented Reality (AR), and Virtual Reality (VR) within data center cooling infrastructures [34], with the use of AI, it has been proposed that power density could be boosted and data centers could be in proximity with the end users, according to the EU-funded ECO-Qube project specifications, a zonal energy management system will be developed, it will make use of Computational Fluid Dynamics (CFDs) to ensure cooling systems achieve high performance, at the same time have minimum energy consumption.

4.3. Using of Artificial Intelligence in Data Center Cooling Systems

The use of AI has found its application in data centers cooling systems, which is evidenced by numerous studies [35], through the use of Computational Fluid Dynamics (CFD) data sets, a Recurrent Neural Network (RNN) can be used to forecast data centers eminent conditions, such as temperature variability and airflow within the cooling systems [35], it was realized that AI driven approach minimized energy wastage, and generalized that AI can achieve efficiency, low operational cost and facilitate sustainable environment. In other studies, Artificial Neural Network (ANN) has been used in characterization of the dampers for thermal management [36], whereby an automated management technique has been used to deliver cold air to individual aisles, which is in reference to the information gathered from the Information Technology Equipment (ITE). For example, the air flow demanded at various aisles as shown in **Figure 10**, the layout shows Computer Room Air Handlers (CRAH) and the optimal location of aisles.

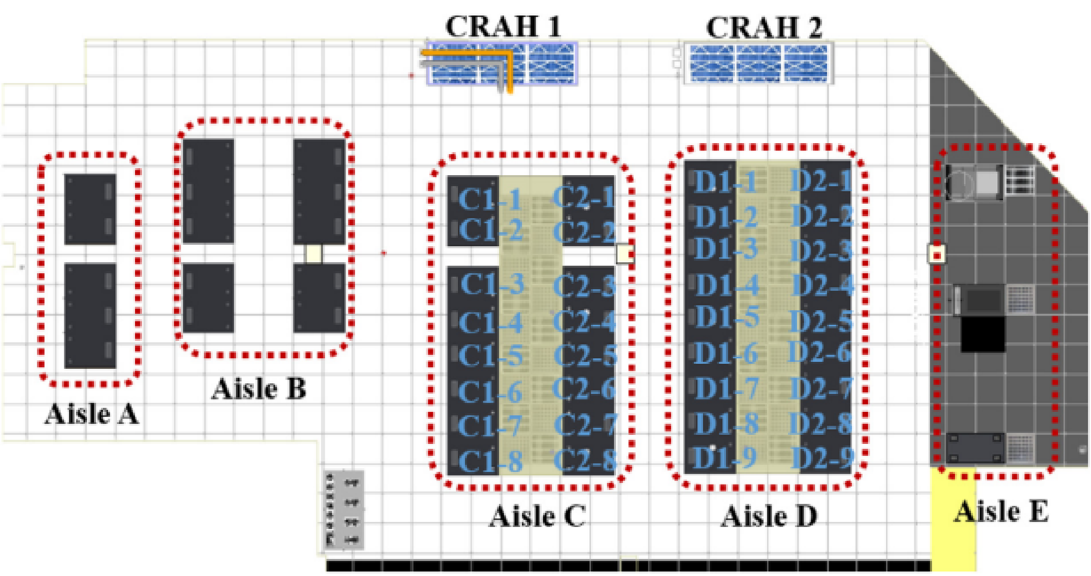

Figure 10. Layout of the data center laboratory at the University Binghamton [36].

Machine learning or Reinforcement Learning (RL) [37] was used to train a model to obtain the characteristics of Data Center (DC) room. In ideal cases, RL is usually used to control a system using trial and error strategy, but in regard to DCs an high-fidelity model was considered, and this was accomplished through CFD that contemplated Lattice Boltzmann Method (LBM) Bhathagar-Gross-Krook (BGK) algorithm, which factored in all the transient conditions of the DC room, and was followed by the training of the RL agent that in turn controlled the cooling equipment, it is noted that an RL agent can adapt to environmental changes and can flag the system when an equipment breaks down. Multi-Agent Reinforcement Learning [38] has been formulated using multiple Computer Room Air conditioners (CRACs) as agents, every single agent learns minimal local cooling policy which is adopted by the Counterfactual Multi-Agent (COMA), the results showed reduction of 6.2% power consumption of CRACs when compared to the Independent Q-Learning (IQL) method.

5. Case Study: AI-Assisted Cooling Optimization in High-Performance Computing

In the previous sections, we discussed the foundations of data center cooling systems and explored several innovative cooling strategies aimed at improving energy efficiency and sustainability. Building on these insights, this section presents a case study designed to demonstrate the practical application of such strategies. By examining real-world data from Frontier, currently the world's most powerful publicly disclosed supercomputer hosted at Oak Ridge National Laboratory, we aim to evaluate the potential of AI-assisted cooling optimization [39]. Frontier's infrastructure offers a unique opportunity for data-driven analysis, as it provides high-resolution measurements of power demand, waste heat, coolant temperatures, and flow rates [39]. This case study not only quantifies avoidable energy waste through overcooling but also provides a data-driven perspective on how predictive models can contribute to smarter, more efficient cooling management in high-performance computing environments.

5.1. Supercomputers and Frontier

Supercomputers are at the pinnacle of high-performance computing, designed to solve complex problems at unprecedented speeds. These machines are not only computationally powerful but also energy-intensive, requiring advanced cooling technologies to manage their operations efficiently. Among them, The Hewlett Packard Enterprise–Cray EX Frontier stands out as the world's first and fastest exascale supercomputer, ushering in a new era of computational power [40] (**Figure 11**). Frontier is located at the Oak Ridge National Laboratory's Oak Ridge Leadership Computing Facility (OLCF). It comprises 74 compute rack cabinets, each containing 64 blades, with each blade housing two nodes (**Figure 12**). Each node is equipped with four GPUs, one CPU, 4 TB of flash memory, and its own cooling loops. The system's high-temperature water cooling infrastructure is integral to its operation, featuring coolant distribution units and an innovative tertiary cooling system to ensure efficiency and sustainability [39].

Figure 11. A general exterior view of Frontier located at Oak Ridge National Laboratory's Oak Ridge Leadership Computing Facility.

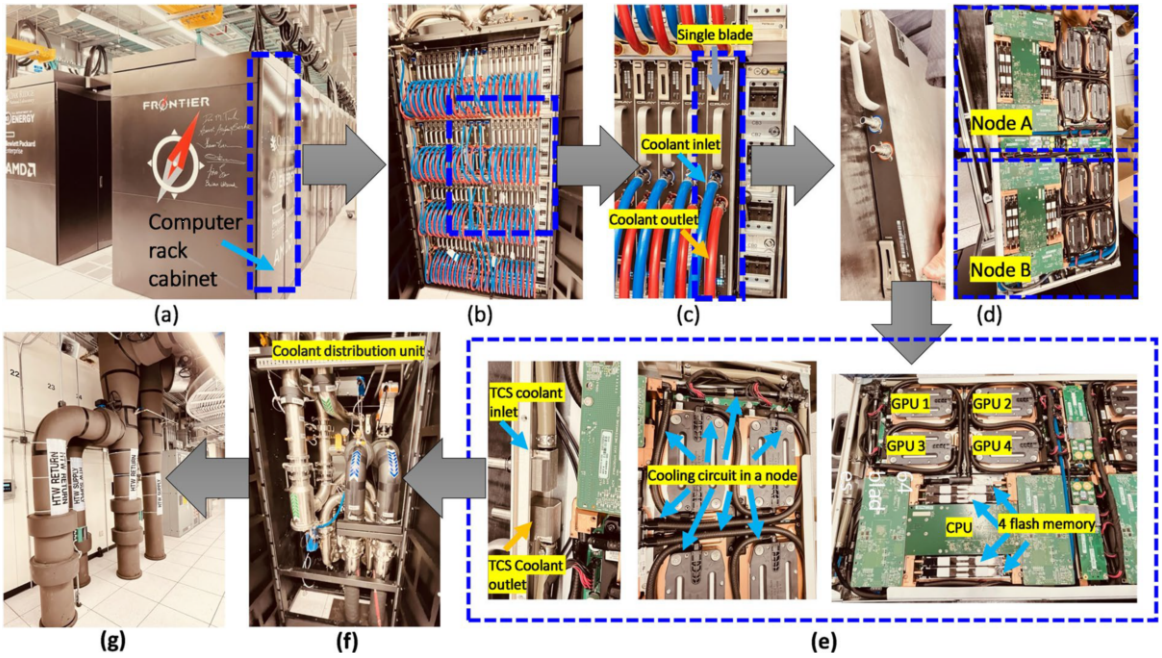

Figure 12. Components and their cooling hardware for the Frontier high-performance computing system: (**a**) 74 compute rack cabinets; (**b**) 64 blades in each rack cabinet; (**c**) coolant inlet and outlet in each blade; (**d**) two nodes in an open blade; (**e**) four GPUs, one CPU, 4 TB of flash memory, and their cooling loops in each node; (**f**) coolant distribution unit; and (**g**) high-temperature water return and supply.

Note: (CPU = central processing unit, GPU = graphics processing unit, TCS = tertiary cooling system).

5.2. Data-Driven Overcooling Detection

The analysis presented in this section utilizes real-world operational data obtained from the Frontier supercomputer, one of the most powerful high-performance computing (HPC) systems currently in operation. The publicly available dataset includes time-series measurements of compute power, total waste heat, coolant flow rates, and return temperatures. These parameters are recorded at 10-minute intervals and reflect the dynamic behavior of Frontier's liquid-cooling infrastructure under varying workloads [39].

Building on this comprehensive dataset, an AI-assisted model was developed to evaluate the degree of overcooling observed in the Frontier cooling infrastructure. The model was trained using the full dataset comprising approximately 52,000 time-series records spanning an entire year. For illustrative purposes, a subset of 200 consecutive time intervals (~ 33 hours) was selected to demonstrate the modeling framework and present representative results. The model was designed to predict the expected average return temperature of the coolant based on two input features: compute power and total waste heat as indicated by Equation (1). The AI learns the relationship:

$$\hat{T}_{return} = f\left(P_{compute}, Q_{waste}\right) \tag{1}$$

A linear regression model was used to estimate the expected coolant return temperature based on compute power and total waste heat. Overcooling was defined as any instance where the actual return temperature T_return was more than 1.5 °C lower than the predicted value T_return to avoid misclassifying natural thermal fluctuations as shown by Equation (2):

$$\Delta T = \hat{T}_{return} - T_{return} > \theta \tag{2}$$

This approach enabled the identification of potential energy inefficiencies resulting from excessive cooling. The complete structure of the overcooling detection framework including the data inputs, AI model, and decision logic is illustrated in **Figure 13**.

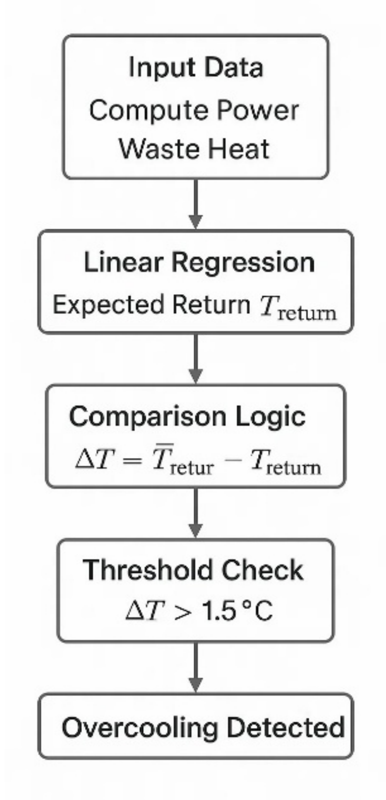

Figure 13. Overcooling detection framework based on ai regression model.

The model takes compute power and waste heat as input features to estimate the expected return temperature of the coolant using a linear regression model. This predicted temperature is then compared with the actual return temperature measured by sensors. When the temperature difference (ΔT) exceeds a predefined threshold (1.5 °C), an overcooling event is flagged. This approach enables the identification of unnecessary cooling efforts, which may lead to energy inefficiency.

5.3. Dataset Description

The dataset used in this study was obtained from the Oak Ridge National Laboratory (ORNL) and reflects one year of operational measurements from the Frontier supercomputer—the world's first exascale HPC system [39]. The data includes time-series records sampled at 10-minute intervals over the course of 2023. Each record captures key parameters from the liquid-cooling infrastructure and system power metrics.

The dataset consists of 18 parameters including: coolant return and supply temperatures for three cooling subloops, flow rates (in gallons per minute), subloop-specific and total waste heat (in megawatts), and compute power demand. Additional parameters include total power consumption, accessory power, and Power Usage Effectiveness (PUE). These variables provide a detailed thermal and energy profile of the Frontier HPC system under real-world workload conditions.

This dataset serves as the foundation for training the AI-based temperature prediction model, enabling data-driven evaluation of overcooling patterns and their impact on thermal efficiency.

Figure 14 illustrates how compute power and total waste heat correlate with coolant return temperature over a 33-hour period (200 consecutive records). As compute power increases, both waste heat and return temperature generally follow upward trends, indicating a clear thermal response to workload intensity. The alignment of these parameters justifies the selection of compute power and waste heat as the two input features for the AI-based prediction model. Their combined influence captures both the processing load and the resulting thermal output, making them effective predictors of coolant behavior under varying operational conditions.

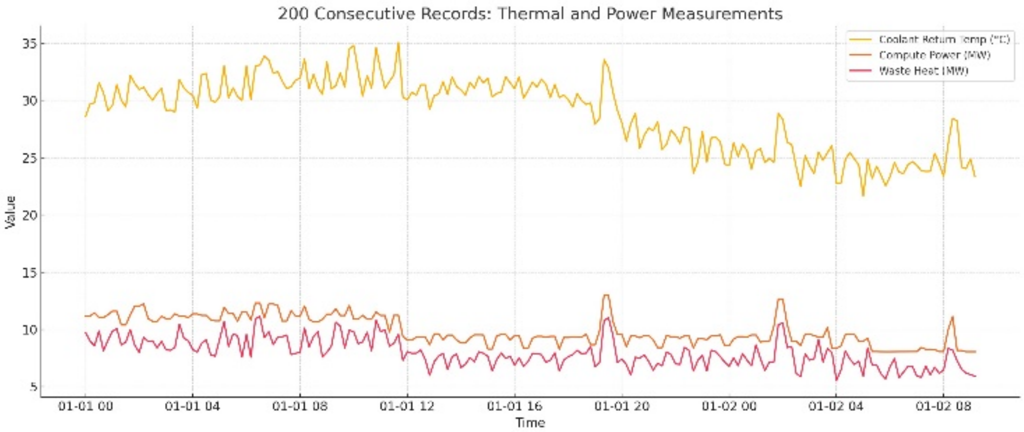

Figure 14. Compute power and waste heat vs. return temperature (200 consecutive records).

5.4. Model Training

To estimate the expected coolant return temperature under varying workload conditions, a supervised learning model was trained using the cleaned Frontier dataset. After removing missing and inconsistent records, the final dataset contained approximately 49,000 valid samples. The input feature vector is defined by Equation (3):

$$X = [P_{compute}, Q_{waste}] \tag{3}$$

where:

- $P_{compute}$ denotes the system-wide compute power (MW)
- Q_{waste} represents a total measured waste heat (MW)

The model output is identified by using Equation (4):

$$\hat{T}_{return} \in R^{+} \tag{4}$$

A linear regression model was selected due to its simplicity, interpretability, and suitability for system-level estimation where real-time implementation and transparency are important [41]. The model was trained on the full dataset without partitioning into training and test sets, as the primary goal was not forecasting accuracy, but identifying thermal baselines for overcooling detection.

This approach was adopted because the study is framed as a diagnostic analysis of historical operational data, not as a forecasting task. Using the full dataset ensures that all observed workload and thermal variations are included in the baseline estimation, which is essential for accurately identifying systemic overcooling trends.

Since the objective of this study was not predictive forecasting but the establishment of a thermal baseline for overcooling detection, the full dataset was utilized to capture the complete range of operating conditions and workload variations. Nevertheless, to ensure the model's generalizability, a time-based validation was conducted by dividing the dataset chronologically into training and testing segments. The validation results were consistent with the full-data model, confirming that the regression captured stable thermal trends without significant overfitting.

The loss function used for model fitting and can be identified using Equation (5) as the Mean Squared Error (MSE):

$$\mathcal{L}_{MSE} = \frac{1}{n}\sum_{i=i}^{n}\left(T_{return,i} - \hat{T}_{return,i}\right)^2 \tag{5}$$

Model performance was evaluated using several standard regression metrics:

- MSE = 11.96 °C^2

- RMSE = 3.46 °C
- MAE = 2.76 °C
- R^2 score = 0.357

Although the R^2 value indicates only moderate explanatory power, the model successfully captures the general trend between workload and return temperature. Given its lightweight structure and consistent trend alignment, this model was deemed appropriate for identifying overcooling behavior relative to predicted thermal baselines [42].

Although the R^2 value is moderate, this is acceptable because the model's purpose is not to achieve precise predictive accuracy but to identify systemic thermal trends associated with potential overcooling behavior. The focus is thus on trend consistency and interpretability, rather than maximizing prediction accuracy.

While more advanced machine-learning techniques such as random forests or artificial neural networks could potentially capture nonlinear interactions, a simple linear regression model was deliberately selected to ensure key requirements for real-time implementation in datacenter environments like interpretability, low computational overhead, and transparency. Preliminary correlation analysis of Frontier's dataset indicated approximately linear trends between compute power, waste heat, and return temperature, justifying the linear assumption. Future work will explore nonlinear extensions to assess whether marginal performance gains justify the added complexity.

The plot at **Figure 15** compares model predictions with actual return temperatures over a 33-hour period. The alignment between the two curves confirms the model's ability to capture general thermal behavior under varying workloads.

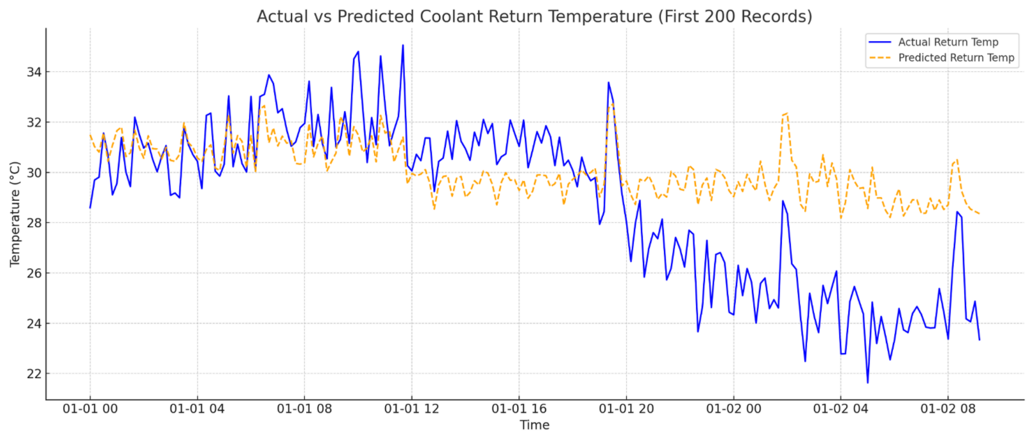

Figure 15. Comparison of actual and predicted coolant return temperatures over 200 consecutive records.

5.5. Overcooling Detection and Energy Implication

To identify unnecessary cooling effort, we compared the actual return temperature T_{return} with the predicted baseline $\hat{T}_{return}$. Overcooling was defined as any instance where the actual return temperature was at least $\theta = 1.5^{\circ}$C below the predicted value as shown by Equation (6).

$$\Delta T = \hat{T}_{return} - T_{return} > \theta \tag{6}$$

Rather than directly estimating energy losses, we adopted a relative cooling contribution metric based on the product of coolant flow and temperature differential (ΔT × Q). This method allows for physically interpretable and unit-consistent comparison of total versus required cooling effort. The required cooling was recalculated using the threshold-adjusted formulation see Equation (7) below.

$$\Delta T_{required} = max(0, \hat{T}_{return} - T_{supply} - \theta) \tag{7}$$

The total and required cooling contributions were compared across all valid records, and the avoidable overcooling effort was quantified. Results showed that 6.9% of the total cooling load could be reduced without compromising system thermal targets. This represents a significant energy-saving opportunity for the Frontier system. To account for measurement variability and model uncertainty, a sensitivity analysis was conducted by adjusting the overcooling detection threshold (ΔT) within ±0.5 °C around the nominal 1.5 °C baseline. The estimated reduction in cooling effort ranged between 6.2% and 7.5%, indicating moderate sensitivity of the results to the chosen threshold. Although these variations do not substantially alter the overall conclusion, they provide a confidence range that strengthens the robustness of the findings.

5.6. Energy Impact of Avoidable Overcooling

To evaluate the broader implications of overcooling reduction, the estimated avoidable cooling effort was translated into energy consumption terms and contextualized using public-scale usage equivalents. According to operational data from the Frontier supercomputer, which sustains an average power demand of approximately 12.2 MW throughout the year [39], a substantial portion of this load is attributed to cooling infrastructure. Prior studies and industry assessments typically estimate that cooling systems account for roughly 30–40% of total data center energy consumption; thus, an assumed value of 35% was adopted for this analysis [43]. Based on this proportion, the total annual energy consumption attributed to cooling operations in the Frontier system is estimated to be approximately 37.4 million kilowatt-hours (kWh). This figure provides a baseline for assessing the potential energy savings that could be achieved by mitigating unnecessary cooling efforts.

This estimated cooling energy corresponds to approximately 37.4 million kilowatt-hours (kWh) annually. To contextualize this amount, it is equivalent to the annual electricity consumption of over 3100 average U.S. households, based on the U.S. Energy Information Administration's (EIA) 2022 estimate of 11,000 kWh per household per year [44]. Alternatively, this amount of energy could be generated by a utility-scale solar power plant with a capacity of approximately 5 MW, assuming a capacity factor of 20% and continuous annual operation. These comparisons highlight the significant impact that even modest improvements in datacenter cooling efficiency can have on energy conservation at scale, reinforcing the importance of predictive and adaptive cooling strategies (**Figure 16**).

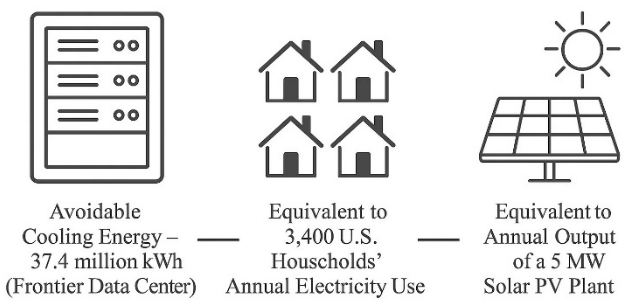

Figure 16. Public-scale energy equivalence of avoidable cooling in the Frontier data center.

Note: Estimated annual savings are approximate and may vary due to uncertainties in PUE and cooling energy assumptions.

When compared with previous AI-driven cooling optimization studies, the estimated 6.9% reduction in cooling effort is consistent with reported efficiency gains in the range of 5–10%. For instance, Senthilkumar et al. [35] demonstrated approximately a 5.8% reduction using a machine-learning-based thermal control framework, while Heimerson et al. [37] achieved about 7% energy savings through deep reinforcement learning for adaptive cooling. Similarly, Chen et al. [38] reported a 6.2% reduction using multi-agent reinforcement learning in data center environments. These benchmarks indicate that the performance of the proposed regression-based approach aligns well with more complex AI techniques, demonstrating that even lightweight and interpretable models can achieve comparable improvements in cooling efficiency.

6. Conclusions

Data centres have become the central hubs of the digital age. Consequently, advancements in information technologiessuch as data analytics, cloud computing, big data, and the rapid expansion of cloud-based applicationscontinue to increase the computational load and energy demands on these facilities.

Datacenters house numerous servers, storage devices, and networking equipment, all of which operate on electricity and inevitably generate heat, creating significant cooling challenges. Despite growing interest in "Green Data Center" concepts and ongoing research into eco-friendly designs, contemporary datacenters still convert approximately one-third of their total energy consumption into heat.

This study has explored the impact of temperature on datacenter efficiency, supported by a thermal analysis of a CPU heat sink to emphasize the necessity of effective cooling. A review of recent literature has also been provided, covering the state-of-the-art in cooling systems for efficient and low-carbon-footprint datacenters.

Furthermore, waste heat utilization has been discussed as a promising strategy to improve overall efficiency, reduce operational costs, and support environmental sustainability. These findings underscore the importance of integrated thermal management and energy reuse strategies in the pursuit of greener, more efficient datacenter operations.

Despite its promising findings, this study has certain limitations. The linear regression model, while interpretable and lightweight, does not capture potential nonlinear interactions among thermal parameters. In addition, the analysis is limited to historical data from a single HPC system, and real-time adaptability was not evaluated. Future work will focus on extending the proposed framework toward dynamic cooling control through real-time feedback mechanisms and integrating renewable-driven cooling systems, such as geothermal or heat-pump-assisted loops. Further testing on multiple data center environments will also help validate the model's scalability and generalizability.

Author Contributions

Conceptualization, O.Ş. and O.N.O.; methodology, O.Ş. and O.N.O.; software, O.Ş.; validation, O.Ş., O.N.O. and B.Y.; formal analysis, O.Ş.; investigation, O.N.O.; resources, B.Y.; data curation, O.Ş.; writing—original draft preparation, O.Ş. and O.N.O.; writing—review and editing, O.N.O.; visualization, O.Ş.; supervision, B.Y.; project administration, B.Y. All authors have read and agreed to the published version of the manuscript.

Funding

This work received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data used was obtained from Oak Ridge National Laboratory (ORNL) website: https://www.olcf.ornl.gov/frontier/.

Conflict of Interest

The authors declare that there are no conflict of interest.

References

1. Masanet, E.; Shehabi, A.; Lei, N.; et al. Recalibrating Global Data Center Energy-Use Estimates. *Science* **2020**, *367*, 984–986.
2. What You Should Know About Data Center Cooling Technologies? Available online: https://www.vxchnge.com/blog/data-center-cooling-technology (accessed on 12 May 2022).
3. El-Sayed, N.; Stefanovici, I.; Amvrosiadis, G.; et al. Temperature Management in Data Centers: Why Some (Might) Like It Hot. In Proceedings of the 12th ACM SIGMETRICS/PERFORMANCE Joint International Conference on Measurement and Modeling of Computer Systems, London, UK, 12–14 June 2012; pp. 163–174.

4. Data Centres and Renewable Energy Project. Available online: https://www.esru.strath.ac.uk//EandE/Web_sites/09-10/Data_centres/Cooling_Approaches.html (accessed on 6 June 2023).
5. Khalaj, A.H.; Halgamuge, S.K. A Review on Efficient Thermal Management of Air- and Liquid-Cooled Data Centers: From Chip to the Cooling System. *Appl. Energy* **2017**, *205*, 1165–1188.
6. Zhang, Q.; Meng, Z.; Hong, X.; et al. A Survey on Data Center Cooling Systems: Technology, Power Consumption Modeling and Control Strategy Optimization. *J. Syst. Archit.* **2021**, *119*, 102253.
7. Taniguchi, Y.; Suganuma, K.; Deguchi, T.; et al. Tandem Equipment Arranged Architecture with Exhaust Heat Reuse System for Software-Defined Data Center Infrastructure. *IEEE Trans. Cloud Comput.* **2017**, *5*, 182–192.
8. Wan, J.; Gui, X.; Kasahara, S.; et al. Air Flow Measurement and Management for Improving Cooling and Energy Efficiency in Raised-Floor Data Centers: A Survey. *IEEE Access* **2018**, *6*, 48867–48901.
9. Ni, J.; Bai, X. A Review of Air Conditioning Energy Performance in Data Centers. *Renew. Sustain. Energy Rev.* **2017**, *67*, 625–640.
10. Wang, C.-H.; Tsui, Y.-Y.; Wang, C.-C. Airflow Management on the Efficiency Index of a Container Data Center Having Overhead Air Supply. *J. Electron. Packag.* **2017**, *139*.
11. Sorell, V. Raised Floor Versus Overhead Cooling in Data Centers. In *Data Center Handbook*; John Wiley and Sons: Hoboken, NJ, USA, 2014; pp. 429–439.
12. Lin, P.; Avelar, V. How Row-Based Data Center Cooling Works. *Schneider Electric White Paper*, **2014**, 1–10.
13. Kai, S.; Weijian, C.; Xuyan, Z; et al. Thermal Comfort Study in Aircraft Cabin Based on Human Thermal Regulation Model. In Proceedings of the CSAA/IET International Conference on Aircraft Utility Systems (AUS 2018), Guiyang, China, 19–22 June 2018.
14. You, R.; Chen, J.; Shi, Z; et al. Experimental and Numerical Study of Airflow Distribution in an Aircraft Cabin Mock-Up with a Gasper On. *J. Build. Perform. Simul.* **2016**, *9*, 555–566.
15. Lin, M.; Wierman, A.; Andrew, L.L.; et al. Dynamic Right-Sizing for Power-Proportional Data Centers. *IEEE/ACM Trans. Netw.* **2011**, 1098–1106.
16. Murugesans, S.; Gangadharan, G.R. Harnessing Green IT: Principles and Practices. Wiley Publishing: Hoboken, NJ, USA, 2012.
17. Li, Z.; Kandlikar, S.G. Current Status and Future Trends in Data-Center Cooling Technologies. *Heat Transf. Eng.* **2015**, *36*, 523–538.
18. Kadam, S.T.; Kumar, R. Twenty First Century Cooling Solution: Microchannel Heat Sinks. *Int. J. Therm. Sci.* **2014**, *85*, 73–92.
19. Bar-Cohen, A.; Arik, M.; Ohadi, M. Direct Liquid Cooling of High Flux Micro and Nano Electronic Components. *Proceedings of the IEEE*, **2006**, *94*, 1549–1570.
20. Kim, J. Spray Cooling Heat Transfer: The State of the Art. *Int. J. Heat Fluid Flow* **2007**, *28*, 753–767.
21. Zhang, H.; Shao, S.; Xu, H.; et al. Free Cooling of Data Centers: A Review. *Renew. Sustain. Energy Rev.* **2014**, *35*, 171–182.
22. How to Calculate Cooling Requirements for a Data Center. Available online: https://www.dataspan.com/blog/how-to-calculate-cooling-requirements-for-a-data-center/ (accessed on 1 December 2022).
23. Jin, C.; Bai, X.; Yang, C. Effects of Airflow on the Thermal Environment and Energy Efficiency in Raised-Floor Data Centers: A Review. *Sci. Total Environ.* **2019**, *695*, 133801.
24. Nadjahi, C.; Louahlia, H.; Lemasson, Z. A Review of Thermal Management and Innovative Cooling Strategies for Data Center. *Sustain. Comput. Inform. Syst.* **2018**, *19*, 14–28.
25. Santos, A.F.; Gaspar, P.D.; Souza, H.J.L. Evaluation of the Heat and Energy Performance of a Datacenter Using a New Efficiency Index: Energy Usage Effectiveness Design – EUED. *Braz. Arch. Biol. Technol.* **2019**, *62*.
26. Liu, T.; Wan, J.; Rasheed, Z.; et al. A Real-Time Monitoring System for Data Center Thermal Efficiency Analysis. In Proceedings of the 5th International Conference on Information Science and Control Engineering (ICISCE), Zhengzhou, China, 20–22 July 2018; pp. 84–88.
27. Grishina, A.; Chinnici, M.; Kor, A.L.; et al. Thermal Awareness to Enhance Data Center Energy Efficiency. *J. Cleaner Eng. Technol.* **2020**, *6*, 100409.
28. Gong, X.; Zhang, Z.; Gan, S.; et al. A Review on Evaluation Metrics of Thermal Performance in Data Centers. *Build. Environ.* **2020**, *177*.
29. Greenberg, S.; Khanna, A.; Tschudi, W. High Performance Computing with High Efficiency. *ASHRAE Trans.* **2009**, *115*.
30. Lu, Z.; Zhang, K. Study on the Performance of a Y-Shaped Liquid Cooling Heat Sink Based on Constructal Law for Electronic Chip Cooling. *J. Therm. Sci. Eng. Appl.* **2021**, *3*, 034501.

31. Leppänen, T.; Romka, R.; Tervonen, P. Utilization of Data Center Waste Heat in Northern Ostrobothnia. *Tehnički Glasnik* **2020**, *14*, 312–317.
32. Waste Heat Recovery in Data Centers. Available online: https://www.smithgroup.com/perspectives/2018/waste-heat-recovery-in-data-centers (accessed on 6 May 2023).
33. Sharfuddin, M.; Øi, L.E. Simulation of Heat Recovery from Data Centers Using Heat Pumps. In Proceedings of the 61st SIMS Conference on Simulation and Modelling SIMS, Online, 22–24 September 2020; pp. 71–76.
34. Artificial-Intelligence-Augmented Cooling System for Small Data Centres. ECO-Qube Project, Fact Sheet H2020. Available online: https://cordis.europa.eu/project/id/956059 (accessed on 31 October 2023).
35. Senthilkumar, G.; Rajendran, P.; Suresh, Y.; et al. Computational Engineering-Based Approach on Artificial Intelligence and Machine Learning-Driven Robust Data Centre for Safe Management. *J. Mach. Comput.* **2023**, 465–474.
36. Mohsenian, G; Khalili, S.; Tradat, M.; et al. A Novel Integrated Fuzzy Control System Toward Automated Local Airflow Management in Data Centers. *Control Eng. Pract.* **2021**, *112*, 104833.
37. Heimerson, A.; Sjolund, J.; Brannvall, R.; et al. Adaptive Control of Data Center Cooling Using Deep Reinforcement Learning. In Proceedings of the 3rd IEEE International Conference on Autonomic Computing and Self-Organizing Systems Companion (ACSOS-C), Online, 19–23 September 2022; pp. 1–6.
38. Chen, D.; Wan, J.; Li, L.; et al. Distributed Data Center Cooling Control Based on Multi-Agent Reinforcement Learning. In Proceedings of the 4th International Conference on Frontiers Technology of Information and Computer (ICFTIC), Qingdao, China, 2–4 December 2022; pp. 456–461.
39. Sun, J.; Gao, Z.; Grant, D.; et al. Energy Dataset of Frontier Supercomputer for Waste Heat Recovery. *Sci Data*, **2024**, *11*, 1077.
40. Frontier: The World's First Exascale Supercomputer. Available online: https://www.olcf.ornl.gov/frontier/ (accessed on 31 October 2023).
41. Draper, N.R.; Smith, H. Applied Regression Analysis, 3rd ed. Wiley-Interscience: Hoboken, NJ, USA, 1998.
42. Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning, 2nd ed. Springer: London, UK, 2009.
43. Annual Global Data Center Survey 2022. Available online: https://uptimeinstitute.com/resources/research-and-reports/uptime-institute-global-data-center-survey-results-2022 (accessed on 5 October 2025).
44. How Much Electricity Does an American Home Use? Available online: https://www.eia.gov/tools/faqs/faq.php?id=97&t=3 (accessed on 5 October 2025).