Volume (2) No (1), 2026
Journal of Global Research in Multidisciplinary Studies (JGRMS)
Review Paper
Available online at https://saanvipublications.com/journals/index.php/jgrms/index

Survey on AI-Based Predictive Cooling in Data Centers and Edge Devices

Dr. Nilesh Jain
Associate Professor
Department of Computer Sciences and Applications
Mandsaur University Mandsaur
nileshjainmca@gmail.com

***Abstract*—Increasing energy consumption and thermal concentration in current data centers and thermal power systems necessitates a high level of cooling technologies and intelligent control measures in these systems to attain efficiency, reliability and sustainability. Modern data center cooling systems are covered in this paper. These systems include liquid, air, immersion, spray, and hybrid options. Optimization techniques offered include PID control, model predictive control, and reinforcement learning. Plus, it delves into how ML, DL, and RL (reinforcement learning) may revolutionize cooling prediction, real-time adaptability, and energy optimization. The article also delves into the topic of parameter control in thermochemical treatment processes, including gasification, combustion, and pyrolysis, covering topics like the impact of pressure, heating rate, residence time, and temperature on performance and energy output. The paper illuminates both AI-based and data-based solutions to the better thermal management, emissions cuts, stronger robustness, and sustainable operation.**

Keywords—Data center cooling, Intelligent thermal management, Liquid and immersion cooling, Predictive cooling, Optimization control strategies, Thermochemical treatments.

I. Introduction

The widespread adoption of smart devices and sensors has been accelerated by the digitisation of services and is now pervasive in many different industries, including transportation, healthcare, sports, and beyond. Internet of Things (IoT) [1][2] is the principal technology behind this change; it establishes a wireless network of diverse items, such as sensors, vehicles, and home appliances. Cloud computing and edge computing are two examples of third-party computers that receive and aggregate data; the latter allows for more complicated analysis to be performed remotely. By bringing the processing power closer to the user's location, edge computing [3] reduces latency and improves performance. An additional use for it is as an intermediate layer, which distributes resources by splitting up computationally intensive jobs among several nodes. One benefit of edge computing is the assurance of user privacy provided by processing sensitive information locally, independent of the cloud. In addition to lowering transmission bandwidth and operational expenses, the tiny data volume being transmitted also helps. The extremely demanding processing time requirements of real-time systems often necessitate the use of edge computing. These systems are commonly found in industrial and security applications. Topical and quick actions of rescue workers are required in the critical situations, when many people are involved, e.g. in the case of emergency in high-rise buildings [4]. A framework based on IoT [5] has been created to deal with monitoring various environmental parameters and informing the rescuers on whether thresholds are crossed and hence the usefulness of edge computing [6] to locally process the information and obtain real-time alerts and enhance responsiveness.

Data Centres are the basis of digital technologies in the energy sphere, which makes it possible to conduct advanced analytics, optimization, and automation [7]. The move of the traditional data centre design to the more dynamic and efficient project is becoming a critical issue in the current IT environment. The traditional data centre, which typically follows a predetermined paradigm for resource allocation and construction, ought to be better prepared to handle the ever-changing requirements of contemporary workloads. A great increase in the demand for data production and storage has resulted from the proliferation of digital technologies in many sectors, including the energy industry, e-commerce, cloud computing, telecommunications, and the Internet of Things (IoT) [8][9]. The need for data centres to support the expanding digital infrastructure has been driven up by this factor. However, data centres are power hogs since they require a steady stream of electricity to operate their equipment and maintain optimal storage conditions for data. There are growing worries about the impact of data centres on environmental sustainability because to their high emissions of greenhouse gases and negative impacts on water resources and air quality. Hence, data centres must be developed and operated in a sustainable manner so as to have minimal environmental effects and as much energy efficiency as possible.

Intelligent system design is required to tackle the difficult problem of reducing data centre energy usage [10] while preserving the requirements of compute resources. When dealing with a live data centre, the difficulty level rises even further. A temperature model that takes into account both internal and external factors, such as server energy consumption and the state of the cooling system, can estimate the temperature within the cooling system's hot corridor and help reach this goal without negatively impacting the compute resources' working conditions [11]. Optimising architectural design schemes and equipment control schemes are two main ways to increase the efficiency of data centre cooling systems. Coatings that improve indoor air quality are one example of how architectural design schemes are being optimised [12]. Data centre cooling efficiency and airflow homogeneity can be enhanced through duct network design optimisation, for instance. A new control mechanism is required for the data centre's cooling systems so that they may be fine-tuned to reduce energy consumption and increase operating efficiency.

Data Centre (DC) operations are constantly being optimised and automated with the help of models from AI and ML [13]. More and more DCs are turning to AI and ML apps to streamline and automate their processes. The use of AI and ML models is replacing traditional heuristics and technical solutions in DCs as they scale to meet ever-increasing demands. This has a major impact on plant performance modelling and efficiency improvement. Improving the accuracy of predictions is possible by utilising these powerful technologies across several layers, with a focus on information technology and cooling systems. Energy efficiency, cooling efficiency, resource allocation, problem detection, and many other operational optimisation and management tasks are all under their purview. Neural networks (both convolutional and recursive), LSTM, and GRUs (gated recurrent units) are the DL methods under scrutiny. Building energy performance prediction, simulation, control, and optimisation using machine learning and deep learning apps. Here is the energy distribution in data centres, as seen in Figure 1.

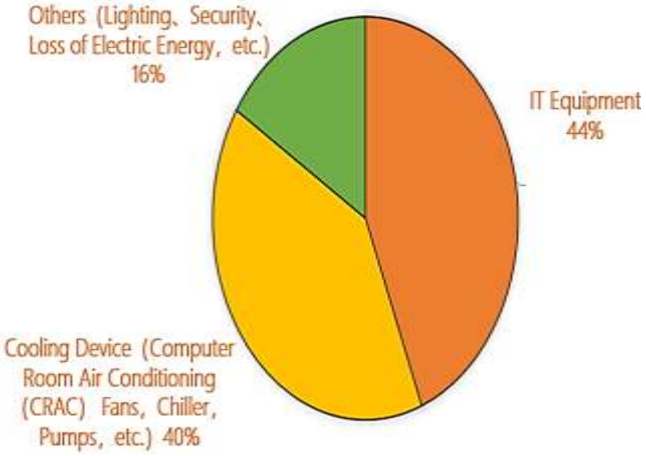

Fig. 1. Schematic Diagram of Data Center Energy Distribution

A. Structure of the Paper

The paper is based on the following structure, Section II: reviews data center cooling technologies and optimization strategies. Section III refers to AI methods of predictive cooling. Section IV looks at the parameter control in thermal treatments in thermochemical. The literature evaluation is presented in Section V, and the paper is concluded with Section VI, which includes the main findings and areas for future study.

II. Cooling Technologies and Optimization Control Strategies in Data Centers

The cooling of electronic components in base station antennas and last-generation mobile telecommunication networks, energy consumption, and intelligent thermal management are all important areas for future study and development [14]. Over the past few years, thermal management in high-power integrated circuits (ICs) has been an important field of research, especially with increasing demands on industries needing more efficient cooling systems, including the telecommunications and data centers industries. The data center components are illustrated in Figure 2.

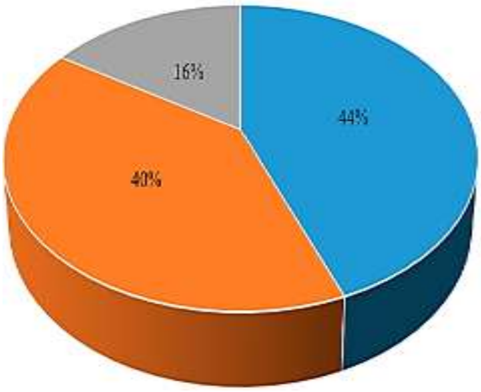

Fig. 2. Components of Data Center Energy Consumption

B. Liquid Cooling Technologies in Data Centers

The high thermal conductivity and specific heat capacity of liquids are used in liquid cooling technology to effectively get rid of heat and keep the equipment within a safe working temperature range, as shown in Figure 3. Using a sealed system to circulate a coolant, liquid cooling technology efficiently manages the heat produced by data centre equipment. First, there's the cooling water system, which uses towers to lower the water's temperature and dissipate the excess heat into the surrounding environment. The third stage involves transferring the cooled water to the central distribution unit (CDU), which supplies further cooling systems directly connected to the equipment. The CDU serves as a hub for the coolant distribution process. Specialised cooling systems in the data center's server cabinets distribute the coolant evenly throughout the room. By soaking up the heat that the servers produce, the systems keep them at the ideal working temperatures. Pumping the heated coolant back to the CDU and then recirculating it to the cooling towers repeats the cycle.

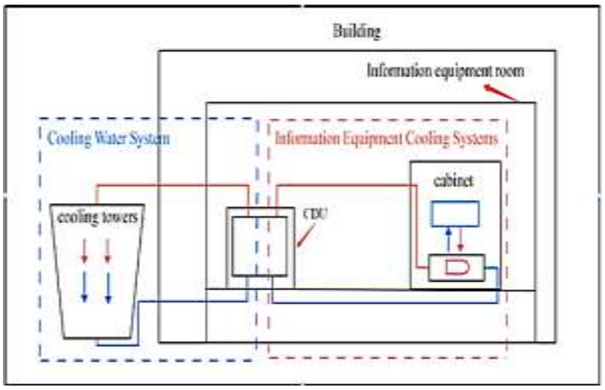
Building
Information equipment room
Cooling Water System
Information Equipment Cooling Systems
cabinet
cooling towers
CDU

Fig. 3. Basic Mechanism of Liquid Cooling Technology

1) Cold Plate Liquid Cooling

Liquid cooling at the chip level is an indirect approach where it follows the heat dissipation process of components producing a lot of heat by mounting cold plates on server CPUs and GPUs. An eco-friendly characteristic is its ability to use warm water as a coolant for direct-to-chip cooling, which is one of its remarkable qualities. At 45 °C or higher, this method can produce waste heat from water. Front and centre, show a waste heat recovery tubes on the majority of commercial liquid cooling products [14]. Keep in mind that the CDU's primary and secondary sides are both capable of recovering waste heat.

2) Immersion Liquid Cooling

The heat-emitting electronic components of an immersion liquid cooling system are submerged in a circulating coolant, resulting in a rapid heat exchange rate. The entire system is covered with a non-conductive coolant, such as mineral oil, silicone oil, or fluorinated fluids, while the IT equipment is in operation. The coolant's ability to undergo a phase transition during heat exchange is the defining characteristic of single-phase versus two-phase immersion cooling.

3) Spray Liquid Cooling

In contrast to the two above systems, heat exchange is attained by directly spraying the coolant over electronic equipment using specially designed nozzles with the spray liquid cooling technology. The coolant is applied onto the electronic equipment or other heat-conducting material surface directly during spraying. The hot coolant is recovered by the return pipeline of the system and pumped back to the CDU to cool. The cooling tower, control distribution unit (CDU), liquid cooling pipeline, spray liquid cooling cabinet, and pipeline system are the typical components of this mechanism, which is known for its precise and efficient cooling process. Spray cooling can be used in many other applications with promising future prospects in the aerospace, biomedicine, and battery safety areas [15]. Ongoing developments are likely to make it more efficient and applicable, defeating the current technological limitations and broadening the range of its activities to more developed and compact electronics.

C. Air Cooling Technologies in Data Centers

Air conditioning methods utilise fans to facilitate the cooling of refrigerant within the condenser, with heat dissipation occurring directly into the ambient air, as depicted in Figure 4. By eliminating the need for cooling towers, pumps, and pipes, this approach may guarantee proper cooling operation in 24 out of 42 water-scarce settings, in comparison to water-cooled chiller systems. Because of its simplicity, dependability, and ease of maintenance, air-cooled chiller systems are extensively utilised in medium to large data centres.

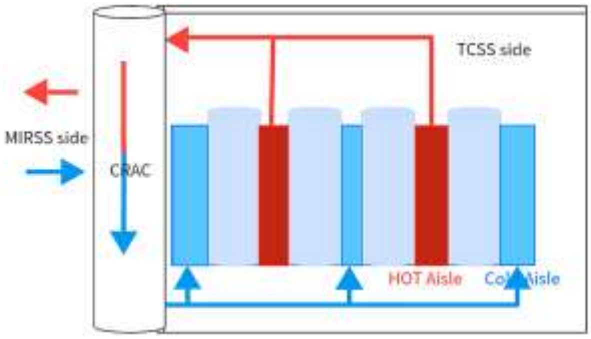

Fig. 4. Basic mechanism of air-cooling technology

1) Direct Air Cooling

Direct air cooling is simple and does not cost much especially in areas where the quality and the temperature of the ambient air are within the acceptable range of IT equipment operation [9]. Nonetheless, this approach is limited. It performs poorly in hot or polluted situations because its efficacy is significantly impacted by ambient air conditions. One other issue is the reliance on high-speed fans, which can generate a lot of noise—not ideal for data centres with a lot of users. Additionally, the system's cooling capability is limited compared to liquid cooling solutions due to air heat dissipation, making it less appropriate for extremely high-density configurations or high-performance computing (HPC) applications with a large heat load.

2) Indirect Air Cooling

Indirect air-cooling is a type of technology that removes heat by exchanging one medium with another by using a heat exchanger where typically the heat is removed by exchanging the hot equipment with water or coolant, which subsequently cools by exchange of heat through the air. Heat exchangers can be found in the data center in indirect air cooling systems [16], Indirect air-cooling is a type of technology that removes heat by exchanging one medium with another by using a heat exchanger where typically the heat is removed by exchanging the hot equipment with water or coolant, which subsequently cools by exchange of heat through the air. Heat exchangers can be found in the data center in indirect air cooling systems.

3) Evaporate Cooling

Cold air is created by absorbing heat as water evaporates; this process is known as evaporative cooling. By reducing the air temperature and increasing humidity, it achieves better thermal management. Evaporative cooling systems in environmentally conscious buildings use a combination of air cooling and natural evaporation to keep indoor temperatures tolerable while reducing the energy consumption of HVAC systems [17]. Even in extremely hot environments, this technology can significantly improve the system's efficiency and performance in terms of output. In extremely humid environments, evaporative cooling is nearly useless. The efficiency of this cooling method is highly dependent on the relative humidity of the surrounding air. The system also needs water supply that is constant and this may be a constraint in the scarcity of water in a region. The system should be regularly maintained to eliminate the possibility of mild and bacteria development and thereby impact the quality of air and efficiency of the system.

D. Optimization Control Strategies for Cooling Systems in Data Centers

Modern data centres rely heavily on the process of managing cooling systems to function. In addition to bolstering overall performance and dependability, this calls for the implementation of cutting-edge technology and methodologies to increase efficiency and decrease energy usage. Due to the inadequacy of the previous systems based on experience-based approaches, automation control strategies have emerged as a crucial instrument to deal with the increasing scale and complexity of data centres' operations [18]. Intelligent control system approaches utilise additional monitoring and automatic control to allow cooling equipment to be modified in real-time according to the actual thermal load. Not only can these techniques improve the system's intelligence and automation, but they also maximise operational efficiency and cost-effectiveness.

1) PID Control

The optimisation control of cooling systems in data centres often makes use of PID control. Stabilising operation levels at predetermined points is the primary goal of this widely used method, which is compatible with the majority of conventional cooling systems. Unfortunately, PID control has its limits when confronted with complicated and dynamic situations. One hundred years ago, PID technology was first created for processes with just one input and one output.

2) *Model Predictive Control (MPC)*

MPC's unparalleled predictive and optimisation capabilities have garnered it widespread interest in data centre control and established it as a foundational technology for enhancing system performance in terms of responsiveness and efficiency. Creating predictive optimisation strategies to improve chilled water systems overall is a prime example of how MPC is expected to regulate and mediate the connection between system performance and energy efficiency.

3) *Reinforcement Learning*

This is an advanced intelligent control strategy, which is known as reinforcement learning (RL) and has shown a lot of potential in terms of adaptive adaptation and performance optimization. The RL methods have intensively been used to make data center and cooling systems more energy efficient and responsive, and have demonstrated significant potential in responding to complex and dynamic environments.

III. AI Techniques for Predictive Cooling

The AI technologies layer processes, improves, and digs deeper into the data and information received from the information analysis layer by using computers, tweaking hardware, and AI models. This aids in real-time dynamic control and makes judgements with better precision. For the data centre, this means smart management of the cooling system that makes the most efficient use of the available energy. By combining several AI models such as knowledge graphs, deep reinforcement learning, and generative AI, gain deep insights into the complex system's operational status and optimize it.

E. *Machine Learning Algorithms*

The algorithms of artificial intelligence began to thrive in the early 1950s, and scientists already theorized that it was possible to give machines the ability to reason logically, and they would become intelligent. This stage's notable accomplishments include, among others, the Logic Theorist and General Problem-Solving programs [19]. Advances in research, however, have shown that AI cannot be achieved through the application of logic alone. Following this, a plethora of expert systems were developed through the process of imparting information to computers [20]. But expert systems are confined in their application spectrum because of their complexity. A number of connectionisms based on neural networks and inductive learning systems based on logic have emerged in the last several decades, ushering in the era of learning machines in the field of artificial intelligence. The circumstances under which distinct ml algorithms perform optimally in diverse application contexts vary greatly, reflecting the recent growth of machine learning as a substantial academic area.

1) *Support Vector Machine (SVM)*

SVM is a ML model that excels in high-dimensional, nonlinear phenomena, small sample sizes, and adheres to the structural risk minimisation principle and the Statistical Learning Theory's dimension approach. Pattern recognition, regression modelling, and many more fields can benefit from it. This method is expressed as a problem of limited quadratic programming [21]. Traditional optimisation techniques effective for small-scale QP. However, this methodology suffers when the size of the training corpus increases, leading to sluggish training speed, complicated algorithm design, and decreased efficiency. Presently, training entails breaking down a big QP problem into smaller ones, solving each of those subproblems in turn until reaching a solution that is close to the original.

2) *Decision Tree*

The decision tree algorithm provides a framework for understanding hierarchical data structures, decision rules, and categorisation outputs. This algorithm is an example of inductive learning; it takes raw data and sorts it into trees, which can then represent unknown data in a predictive way. Where each internal node stands for a feature attribute test [22], each outward branch for the test's conclusion, and each final node for a category or option outcome. The DT method has the benefit of making the decision path from the root node to the terminal node very obvious to the user. It is also possible to interpret the model. Using a decision tree approach is a breeze whether data is numerical or categorical, and it even works with missing values to a certain degree. What's more, it requires very little data preparation. But when the decision trees are complicated, the algorithm tends to overfit.

3) *Random Forest*

Random forest (RF) is a method that is based on statistics learning theory. Bootstrap resampling is a novel technique that takes a range of samples from the original data and uses them to build a decision tree model. This makes the model stronger and more accurate. The randomness of the RF approach is its distinguishing feature for avoiding overfitting; while training each tree [23], a random number of features and samples are selected, lowering the model's variability. One of the most active subfields of bioinformatics and data mining right now is random forest.

F. *Deep Learning Algorithms*

Recent advances in artificial intelligence, known as deep learning, have made classical DNN obsolete in several domains, such as picture recognition, data analysis [24], and processing of time series data. Batteries can benefit from these deep learning techniques for thermal management in order to overcome the shortcomings of traditional methods for defect diagnosis, numerical modelling of thermal behaviour, and state prediction.

1) *Convolutional Neural Network (CNN)*

Recent advances in artificial intelligence, known as deep learning, have outperformed the more traditional DNN in a number of areas, including image identification, data analysis, and processing of time series data [25]. Integrating these recently created deep learning algorithms into battery thermal management helps overcome the shortcomings of older approaches to predicting battery states, defect diagnosis, and numerical modelling of thermal behaviour. Research and applications of CNNs have made it feasible to estimate the spatial thermal parameters of batteries and battery states. CNNs are highly effective at processing photos and multidimensional data.

2) *Recurrent Neural Network (RNN)*

RNN is a DL model [26] to work with sequential data capacity, placing the associations between the previous time offers and the upcoming ones within a neural system. The three main components of a regular RNN are the input, hidden, and output layers. One sequence element at a time is fed into the hidden layer of an RNN, which processes it and then utilises its output as extra input for the next element in the sequence. The RNN takes into account both the present and past aspects of the sequence while making predictions.

Despite its great abilities to solve time series issues, RNN tends to lose feature of the previous sequences as more time steps are added.

3) Residual Neural Network (ResNet)

ResNet is an additional deep learning method that builds upon CNN. The CNN's signals go straight from input to output after incorporating residual modules, which causes the network to go through one or more layers [27]. This can be used to address frequent problems like vanishing and exploding gradients, and also the decline in performance that is often experienced when training deep network (the effect of adding more layers to the model reduces its effectiveness). As a result, ResNet makes DL models more efficient and stable. Nowadays, ResNet is employed for deeper network and feature recognition tasks that are more complicated.

IV. Parameter Control of Thermal Treatments

Through gifted parameters of Thermal Treatments, the harmonious operation conditions: temperature, heating rate, residence time, pressure, and oxidizing atmosphere of thermochemical operation, like pyrolysis, gasification, and combustion, are operated in such a way as to guarantee optimal performance and product quality. Proper management of these parameters defines reaction routes, reaction efficiency, reaction energy output, and emission properties. High-level monitoring and control measures allow operating the processes stably, achieve energy savings, minimize the formation of pollutants, and flexible adjustment to the changes in the properties of the feedstock, making the control of parameters a key to the safe and efficient application of thermal treatment methods.

G. Thermochemical Treatment Technologies

There are a number of thermal processes that can be employed as substitutes for more conventional methods of dealing with biomass and municipal solid waste. These processes guarantee the production of heat, fuels, and electricity, but they also come with their fair share of drawbacks. The main distinction between these systems is the concentration of oxygen at the input. It enters the power plants through the reactors, generates separate thermal pathways, and ultimately impacts the fuels and harmful gaseous emissions that come out of them.

1) Pyrolysis

Thermal decomposition of sewage sludge in pyrolysis reactions occurs in an oxygen-free environment, which yields economically viable products such as biochar, bio-oil, and synthesis gas (the proportion of which varies with the pyrolysis route selected), while reducing carbon dioxide emissions [28]. The percentage composition of products created during pyrolysis is primarily affected by three primary process parameters: heating rate, temperature, and residence duration. In addition to playing a role in the reactor's geometry and supply system, they are depending on the physical and chemical interactions that make up this intricate process. The assessment of secondary factors, like particle size and pressure, aids in the prevention of equipment corrosion, which reduces its useful life.

2) Gasification

Gasification offers a number of different modes of operation; it takes renewable inputs and converts them into fossil fuels by chemical reactions with low degrees of oxidation in different reactor building configurations. The integration of biological pathways is the subject of extensive investigation into the validation of technology-enabled concurrent processes with the goals of increased yields and decreased pollution emissions. By using trash as a raw material in thermochemical processes, the aforementioned technologies achieve an average energy efficiency of 30%. Catalysts can be used to produce power and a wide range of various products with different economic values, including fuels, renewable gases, and chemicals [29]. Accordingly, these processes can enhance the global yield rate and provide additional chances for scale benefits in the long term, depending on the demand for each product created in the plant and its demand in the national or worldwide market.

3) Combustion

A key component of combustion is the automation of the flow, which minimises the mechanical use of parts while connecting electrical power from sensors, the temperature of the exit gas, and the volumetric concentration of each gas. The chemical percentage of hydrogen is typically measured by these sensors, which provide us the data needed to show the relationship between thermal activities such as electricity generation, gas cooling, and gas separation [30].

H. Important Features for Thermal Power Generation

The successful operation of thermal power production, which involves transforming heat power into electrical energy, depends on several characteristics: A heat source is an energy source that is consistent and dependable, such as coal, natural gas, or renewable energy sources like solar or geothermal power; System cooling: efficient system cooling prevents overheating and guarantees operational efficacy; control systems: these systems ensure that all the critical parameters are set up to run securely and efficiently through accurate monitoring and control.

1) Temperature influence on thermal generation

Catalyst conversion of carbon from inserted biomasses can be enhanced by utilizing catalysts such as calcined dolomite, which can reduce tar content while increasing synthesis gas concentration. Improved efficiency in thermal cracking allows for the increased utilization of solids and liquids in two-stage pyrolysis reactors [31]. Starting at 500 °C, the hydrogen and carbon monoxide contents, with the dry gas yield rising to the 850 °C range, are all possible.

2) Constant Pressure Importance

The natural state of the gases and residues, determined by their average molecular weight, is stabilized by keeping the pressure constant or linear in gasification and pyrolysis [32]. This facilitates condensation and cools the generated gas to an ideal temperature for usage in internal combustion engines, which, with the help of increased load constants and a partial loss of heat during combustion, transform mechanical rotation into electrical energy.

3) Heating Rate on Thermal Generation

The concentration of carbon monoxide increases at temperatures exceeding 550 °C when the heating rate is maintained at 10 °C/min and carbon dioxide serves as the reactive medium gas. Reduced tar production is a result of the breakdown of more volatile chemicals at higher temperatures. Raising the heating rate, usually between 15 and 30 degrees Celsius per minute, is a frequent approach in industrial scale gas product and bio-oil extraction to maximise production. This makes sense because the temperature changes quickly and greatly, hitting its best range between 500 and 800 degrees

Celsius [33]. It also helps to raise the energy density of the synthesis gas in models that use standard and slow pyrolysis technology.

4) Time Importance in Thermal Power Plants

The processes for starting and stopping a thermal power plant are distinct and time-consuming. The plant's ability to respond quickly to changes in demand may be affected by its start-up timings, whereas shutdown durations can be important for maintenance and other operational reasons. To maintain a steady and dependable power supply, it is critical to handle these times effectively [34]. Regular maintenance is essential for thermal power plants to ensure optimal operation and prevent failure. When parts wear out, when to replace them, and how reliable the plant is as a whole are all factors in the planning of maintenance schedules. Planning such repair tasks during periods of low demand can help mitigate the effects of power outages.

V. Literature Review

This review summarizes the latest developments in data center cooling and intelligent optimization with a focus on the digital twin synchronization, the control using machine learning and reinforcement learning, and liquid/immersion/hybrid cooling to enhance efficiency, reliability, scalability, and sustainability in a wide range of deployment environments.

Rinaldi et al. (2025) provided a thorough analysis of the challenges associated with keeping digital twins of data centres connected to the edge in smart city settings in sync with one another. The study examines essential factors affecting synchronization accuracy and reliability, identifies significant bottlenecks, and suggests strategies to alleviate their impact. The results aid in the creation of effective synchronization mechanisms, facilitating the implementation of dependable digital twin systems [35].

Chen et al. (2025) introduced a novel method to predict energy efficiency in data center cooling systems, combining feature selection with a deep learning model. The approach uses a three-step feature selection process—mRMR, XGBoost, and NSGA-II—to optimize input features and hyperparameters, enhancing accuracy while minimizing sensor data requirements. The resulting deep neural network (DNN) processes time-series data without steady-state assumptions [36].

See et al. (2024) aimed to bring liquid immersion cooling technology into client/edge desktop segments, with the intent of miniaturization without trading off the performance. A self-contained liquid immersion cooling approach has been developed for desktop/IOT systems. The system chassis is designed with compartmentalization as a standalone form factor system. The heat source distribution, liquid flow, and the liquid cooling heat sink have been optimized. The prototyped solution has demonstrated a cooling capability of 650W, showing a path for Intel® Xeon® Scalable Processors family CPUs to get into a Small Form Factor (SFF) PC [37].

Mebratu et al. (2023) developed a framework that utilises Reinforcement Learning (RL) to aid in decision-making. This methodology is based on the Markov Decision Process (MDP) and the contextual bandits approach. Agents, states, rewards, actions, and environments make up the RL algorithm. In this scenario, the agent (learner) keeps tabs on the state of the liquid cooling system, makes a decision, and then watches the outcome of that decision. To learn how to make judgements based on the current status of the system, the agent trains on a variety of physical and virtual data that depicts the environment of the liquid cooling system. For example, it learns to recognize when a leak has occurred [38].

Chen et al. (2023) highlighted the design of Intel's Open IP immersion cooling reference system, which incorporates a modular layout for the server node, coolant distribution unit, and immersion tank. A data centre that is better equipped to support itself can be created using this layout. Analysis and evaluation of system architectural originality, design optimization, experimental verification and validation outcomes, and reductions in both operational and embodied carbon footprints are conducted. For the design of cloud and edge scalable immersion cooling systems, key lessons learnt in engineering practice are a great resource [39].

Guo et al. (2022) unveiled a state-of-the-art Xeon Scalable Processor edge server for outdoor use, utilizing a hybrid cooling method that combines refrigeration with inner and outer circulation mechanisms to cool two layers of air. This hybrid cooling solution was created to support the redeployment of IT devices or components that have a working temperature restriction of 5~35℃ in a data center or 5~45℃ in an HTA data center. In terms of energy efficiency and the dependability of edge servers, it is perfect [40].

Table I summarizes recent studies in data center optimization and cooling from a variety of disciplines; it describes the advantages of digital twins, machine learning, and reinforcement learning in terms of reliability, energy efficiency, scalability, and sustainability; and it describes the notion of liquid, immersion, and hybrid cooling systems; it also identifies the main obstacles in the areas of synchronization, real-time control, deployment limits, and carbon reduction.

TABLE I. SUMMARY OF RECENT STUDIES ON PREDICTIVE COOLING IN DATA CENTERS

Reference	Domain	Cooling System	Optimization Strategy	Deployment Context	Challenges
Rinaldi et al. (2025)	Smart cities, edge-enabled data centers	Digital twin–assisted cooling systems	Time synchronization analysis and mitigation strategies	Edge data centers in smart city environments	Synchronization accuracy, latency, reliability bottlenecks
Chen et al. (2025)	Data center energy management	Data center cooling systems	Feature selection (mRMR, XGBoost, NSGA-II) with DNN-based prediction	Large-scale data centers	High sensor dependency, dynamic operating conditions
See et al. (2024)	Edge computing, desktop/IoT systems	Liquid immersion cooling	Thermal design optimization (heat source layout, liquid flow, heat sink design)	Client/edge desktops and SFF systems	Miniaturization, thermal density, form factor constraints
Mebratu et al. (2023)	Intelligent cooling control	Liquid cooling systems	Reinforcement learning (MDP, contextual bandits) for anomaly detection	Simulated and physical liquid-cooled environments	Real-time state estimation, leak detection accuracy

Chen et al. (2023)	Sustainable data centers	Immersion liquid cooling	Modular system architecture and design optimization	Cloud and edge data centers	Carbon footprint reduction, scalability, system integration
Guo et al. (2022)	Edge computing infrastructure	Hybrid air cooling with refrigeration	Two-layer air circulation and hybrid cooling optimization	Outdoor edge servers, extended temperature environments	Thermal reliability, energy efficiency in HTA conditions

VI. Conclusion and Future Work

Hyperscale and edge data centers, which use immersion cooling, have been mushrooming recently, with the latter seeing particularly quick expansion. The location, footprint, design cost, PUE objective, and other considerations of an immersion cooling data center's system architecture can differ. At the same time, in this age of sustainability, computers are once again measured by their ability to generate more power and better performance in a way that is orientated towards the circular economy, with the end objective of producing no net carbon emissions. This paper has discussed the most modern cooling technologies and intelligent control measures that are intended to solve the increasing thermal and energy issues in the present-day data center and thermal power systems. Cooling solutions of liquid, air, immersion, spray, and hybrid were reviewed and optimization tools of PID, model predictive control, and RL were distinguished with their respective advantages and shortcomings in operating in various conditions. The combination of AI, such as ML and DL, proves a great potential of predictive cooling, real-time adaptation of the system, and increased energy efficiency. Besides, the discussion of parameter control in thermochemical treatment processes: pyrolysis, gasification, and combustion indicate the importance of high accuracy of the temperature, pressure, heating rate, and residence time in the maintenance of steady operation and maximum energy production.

The aim of future work should be to create the unified AI-based control frameworks like introducing digital twins, real-time sensing, and adaptive learning to cooling and thermochemical systems. The focus on scalability, carbon-conscious optimization, and validation of operations in practice will additionally increase sustainability and resilience in operations.

References

[1] S. Garg, "Next-Gen Smart City Operations with AIOps & IoT : A Comprehensive look at Optimizing Urban Infrastructure," *J. Adv. Dev. Res.*, vol. 12, no. 1, 2021, doi: 10.5281/zenodo.15364012.

[2] S. Dodda, N. Kamuni, P. Nutalapati, and J. R. Vummadi, "Intelligent Data Processing for IoT Real-Time Analytics and Predictive Modeling," in *2025 International Conference on Data Science and Its Applications (ICoDSA)*, IEEE, Jul. 2025, pp. 649–654. doi: 10.1109/ICoDSA67155.2025.11157424.

[3] D. Patel and R. Tandon, "Cryptographic Trust Models and Zero-Knowledge Proofs for Secure Cloud Access Control and Authentication," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 2, no. 1, pp. 749–758, Dec. 2022, doi: 10.48175/IJARSCT-7744D.

[4] F. C. Andriulo, M. Fiore, M. Mongiello, E. Traversa, and V. Zizzo, "Edge Computing and Cloud Computing for Internet of Things: A Review," *Informatics*, vol. 11, no. 4, 2024, doi: 10.3390/informatics11040071.

[5] K. M. R. Seetharaman, "Internet of Things (IoT) Applications in SAP: A Survey of Trends, Challenges, and Opportunities," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 3, no. 2, Mar. 2021, doi: 10.48175/IJARSCT-6268B.

[6] G. Maddali, "An Efficient Bio-Inspired Optimization Framework for Scalable Task Scheduling in Cloud Computing Environments," *Int. J. Curr. Eng. Technol.*, vol. 15, no. 3, 2025.

[7] R. Patel and R. Tandon, "Advancements in Data Center Engineering: Optimizing Thermal Management, HVAC Systems, and Structural Reliability," *Int. J. Res. Anal. Rev.*, vol. 8, no. 2, 2021.

[8] R. Patel, "Optimizing Communication Protocols in Industrial IoT Edge Networks: A Review of State-of-the-Art Techniques," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 4, no. 19, pp. 503–514, May 2023, doi: 10.48175/IJARSCT-11979B.

[9] G. Sarraf, "Resilient Communication Protocols for Industrial IoT : Securing Cyber- Physical-Systems at Scale," *Int. J. Curr. Eng. Technol.*, vol. 11, no. 6, pp. 694–702, 2021, doi: 10.14741/ijcet/v.11.6.14.

[10] P. B. Patel, "Energy Consumption Forecasting and Optimization in Smart HVAC Systems Using Deep Learning," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 4, no. 3, pp. 780–788, 2024, doi: 10.48175/IJARSCT-18991.

[11] A. Kula, D. Dąbrowski, M. Blachnik, M. Sajkowski, A. Smalcerz, and Z. Kamiński, "Modelling the Temperature of a Data Centre Cooling System Using Machine Learning Methods," *Energies*, vol. 18, no. 10, 2025, doi: 10.3390/en18102581.

[12] T. Maggos *et al.*, "Improvement of Buildings' Air Quality and Energy Consumption Using Air Purifying Paints," *Appl. Sci.*, vol. 14, no. 14, 2024, doi: 10.3390/app14145997.

[13] Y. Gebreyesus, D. Dalton, D. De Chiara, M. Chinnici, and A. Chinnici, "AI for Automating Data Center Operations: Model Explainability in the Data Centre Context Using Shapley Additive Explanations (SHAP)," *Electronics*, vol. 13, no. 9, 2024, doi: 10.3390/electronics13091628.

[14] Q. Chang, Y. Huang, K. Liu, X. Xu, Y. Zhao, and S. Pan, "Optimization Control Strategies and Evaluation Metrics of Cooling Systems in Data Centers: A Review," *Sustainability*, vol. 16, no. 16, 2024, doi: 10.3390/su16167222.

[15] T. Zhang *et al.*, "Advanced Study of Spray Cooling: From Theories to Applications," *Energies*, vol. 15, no. 23, 2022, doi: 10.3390/en15239219.

[16] H. Chen and D. Li, "Current Status and Challenges for Liquid-Cooled Data Centers," *Front. Energy Res.*, vol. Volume 10, 2022, doi: 10.3389/fenrg.2022.952680.

[17] P. B. Patel, "Predictive Maintenance in HVAC Systems Using Machine Learning Algorithms: A Comparative Study," *Int. J. Eng. Sci. Math.*, vol. 13, no. 12, 2024.

[18] Y. Zhang, C. Fan, and G. Li, "Discussions of Cold Plate Liquid Cooling Technology and Its Applications in Data Center Thermal Management," *Front. Energy Res.*, vol. Volume 10-2022, 2022, doi: 10.3389/fenrg.2022.954718.

[19] L. Xu, S. Jin, W. Ye, Y. Li, and J. Gao, "A Review of Machine Learning Methods in Turbine Cooling Optimization," *Energies*, vol. 17, no. 13, 2024, doi: 10.3390/en17133177.

[20] M. R. R. Deva and N. Jain, "Utilizing Azure Automated Machine Learning and XGBoost for Predicting Cloud Resource Utilization in Enterprise Environments," in *2025 International Conference on Networks and Cryptology (NETCRYPT)*, IEEE, May 2025, pp. 535–540. doi: 10.1109/NETCRYPT65877.2025.11102235.

[21] Y. Macha and S. K. Pulichikkunnu, "A Survey of DevOps Practices for Machine Learning and Artificial Intelligence Workflows in Modern Software Development," *ESP J. Eng. Technol. Adv.*, vol. 4, no. 3, pp. 200–208, 2024, doi: 10.56472/25832646/JETA-V4I3P121.

[22] V. Pal, "Hybrid Quantum-Classical Machine Learning Architectures for Accelerated Drug Discover," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 6, no. 2, pp. 1641–1653, 2021, doi: 10.48175/IJARSCT-6582M.

[23] P. Chandrashekar, "Data-Driven Loan Default Prediction : Enhancing Business Process Workflows with Machine Learning," *Int. J. Emerg. Res. Eng. Technol.*, vol. 6, no. 4, pp. 18–26, 2025.

[24] V. Verma, "Deep Learning-Based Fraud Detection in Financial

Transactions : A Case Study Using Real-Time Data Streams," vol. 3, no. 4, pp. 149–157, 2023, doi: 10.56472/25832646/JETA-V3I8P117.

[25] S. Amrale, "Anomaly Identification in Real-Time for Predictive Analytics in IoT Sensor Networks using Deep," *Int. J. Curr. Eng. Technol.*, vol. 14, no. 6, pp. 526–532, 2024, doi: 10.14741/ijcet/v.14.6.15.

[26] R. Bagwe, J. Kachhia, A. Erdogan, and K. George, "Automated Radar Signal Analysis Based on Deep Learning," in *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, 2020, pp. 215–221. doi: 10.1109/CCWC47524.2020.9031240.

[27] R. P. Mahajan, "Optimizing Pneumonia Identification in Chest X-Rays Using Deep Learning Pre-Trained Architecture for Image Reconstruction in Medical Imaging," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 5, no. 1, pp. 52–63, Apr. 2025, doi: 10.48175/IJARSCT-24808.

[28] K. Mphahlele, R. H. Matjie, and P. O. Osifo, "Thermodynamics, kinetics and thermal decomposition characteristics of sewage sludge during slow pyrolysis," *J. Environ. Manage.*, vol. 284, p. 112006, Apr. 2021, doi: 10.1016/j.jenvman.2021.112006.

[29] M. Khamies, S. Kamel, M. H. Hassan, and M. F. Elnaggar, "A Developed Frequency Control Strategy for Hybrid Two-Area Power System with Renewable Energy Sources Based on an Improved Social Network Search Algorithm," *Mathematics*, vol. 10, no. 9, 2022, doi: 10.3390/math10091584.

[30] S. G. Nnabuife, J. Ugbeh-Johnson, N. E. Okeke, and C. Ogbonnaya, "Present and Projected Developments in Hydrogen Production: A Technological Review," *Carbon Capture Sci. Technol.*, vol. 3, p. 100042, Jun. 2022, doi: 10.1016/j.ccst.2022.100042.

[31] A. B. H. Trabelsi, K. Zaafouri, A. Friaa, S. Abidi, S. Naoui, and F. Jamaaoui, "Municipal sewage sludge energetic conversion as a tool for environmental sustainability: production of innovative biofuels and biochar," *Environ. Sci. Pollut. Res.*, vol. 28, no. 8, pp. 9777–9791, 2021, doi: 10.1007/s11356-020-11400-z.

[32] R. C. Brown, "The Role of Pyrolysis and Gasification in a Carbon Negative Economy," *Processes*, vol. 9, no. 5, 2021, doi: 10.3390/pr9050882.

[33] J.-H. Kim, J.-I. Oh, J. Lee, and E. E. Kwon, "Valorization of sewage sludge via a pyrolytic platform using carbon dioxide as a reactive gas medium," *Energy*, vol. 179, pp. 163–172, Jul. 2019, doi: 10.1016/j.energy.2019.05.020.

[34] H. P. Jagtap, A. K. Bewoor, R. Kumar, M. H. Ahmadi, and L. Chen, "Performance analysis and availability optimization to improve maintenance schedule for the turbo-generator subsystem of a thermal power plant using particle swarm optimization," *Reliab. Eng. Syst. Saf.*, vol. 204, p. 107130, Dec. 2020, doi: 10.1016/j.ress.2020.107130.

[35] S. Rinaldi, S. Dello Iacono, P. Bellagente, and M. Pasetti, "Analysis of Time Synchronization Challenges in Digital Twins for Edge-Enabled Data Centers in Smart Cities Scenario," in *2025 33rd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP)*, 2025, pp. 506–511. doi: 10.1109/PDP66500.2025.00078.

[36] L. Chen, X. Li, K. Wang, and X. Yu, "An Integrated Feature Selection and Deep Learning Method for Energy Efficiency Prediction in Data Center Cooling Systems," in *2025 IEEE 14th Data Driven Control and Learning Systems (DDCLS)*, 2025, pp. 159–164. doi: 10.1109/DDCLS66240.2025.11065957.

[37] K. E. See *et al.*, "Exploration on Miniaturization of Immersion Cooling Technology for Client Desktop and Edge Devices," in *2024 IEEE 40th International Electronics Manufacturing Technology (IEMT)*, 2024, pp. 1–5. doi: 10.1109/IEMT61324.2024.10875284.

[38] D. Mebratu, B. Wondimu, R. Desai, G. Chaudhary, C. Winkel, and M. Hossain, "Liquid Cooling Pipeline Leakage Prediction and Detection using Reinforcement Learning," in *2023 22nd IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, 2023, pp. 1–6. doi: 10.1109/ITherm55368.2023.10177565.

[39] C. Chen *et al.*, "A Novel Scalable Modular Immersion Cooling System Architecture for Sustainable Data Center," in *2023 22nd IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, 2023, pp. 1–7. doi: 10.1109/ITherm55368.2023.10177547.

[40] L. Guo *et al.*, "A Novel Outdoor Edge Server Design with Hybrid Air Cooling and Refrigeration," in *2022 21st IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (iTherm)*, 2022, pp. 1–5. doi: 10.1109/iTherm54085.2022.9899539.