IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 3, December 2023

Green Computing with Deep Learning for Data Centers

Doni Kavya

GMR Institute of Technology, Rajam, India

donikavya098@gmail.com

Abstract: *Due to development in cloud services, lots of data is transferred between users and servers of the cloud. This transmission of data consumes huge amounts of energy. This energy consumption occurs during the operation of network infrastructure, the conversion of electrical to optical signals to travel long distances, and signal amplification. As Green computing is the use of computing devices in an environmentally friendly way, i.e., using electrical energy efficiently as Data centers require a significant amount of electricity to operate and cool the servers, leading to carbon emissions from the burning of fossil fuels. Green computing in cloud services is about optimizing energy consumption and by incorporating deep learning algorithms, we can enhance the energy efficiency of cloud infrastructure. These algorithms can analyze real-time data from sensors, optimize resource allocation, and dynamically adjust power usage. Through intelligent workload scheduling, server consolidation, and power management, deep learning enables the reduction of energy waste and carbon emissions. The integration of deep learning in cloud services not only improves energy efficiency but also enhances performance and cost-effectiveness. Here we are using a deep learning model which can be used for workload prediction and resource provisioning. By analyzing historical workload patterns and user behavior, deep learning algorithmscan predict future resource demands and allocate resources accordingly, leading to more efficient resource utilization and energy savings.*

Keywords: Deep Learning, Green computing, Cloud Computing, Intelligent Workload Scheduling, Server Consolidation, Power Management

I. INTRODUCTION

In today's world, data centers play a crucial role in our digital lives, storing and processing vast amounts of information that power everything from our emails to the latest social media trends. However, these data centers also consume enormous amounts of energy, leading to concerns about their environmental impact.

The good news is that technology is evolving, and one promising solution on the horizon is green computing. Green computing aims to make data centers more environmentally friendly by reducing their energy consumption and carbon footprint. And one of the exciting tools within green computing's toolkit is deep learning.

Deep learning is a powerful type of artificial intelligence that has already revolutionized fields like image recognition and natural language processing. Now, it's making its way into data centers to help them operate more efficiently.

The idea is simple, by using deep learning algorithms, data centers can optimize their energy usage, making sure that servers and cooling systems run at their most efficient levels. This not only saves electricity but also reduces the heat generated, cutting down on the need for energy-intensive cooling systems.

In this paper, we'll explore how deep learning is being used to make data centers greener and more environmentally responsible. We'll see the innovative ways in which these algorithms are being applied to reduce energy waste, lower operational costs, and ultimately, make our digital world a little more eco-friendly.

II. LITERATURE SURVEY:

[1]Forecasting time series of workloads with complex periodic and long temporal dependencies is done by original workload forecasting framework based on representation learning, i.e., we learn multiscale representations from historical workload TimeSeries as external memory, characterizing long temporal dependencies. Maintaining stable

CPU utilization and characterizing uncertainty in the decision-making process is done by taskconditioned Bayesian neural model to learn the relationship between CPU utilization and workload

[2]. The rising usage and complexity of Deep Learning (DL) models lead to heightened energy consumption, a critical environmental concern. The paper applies FECoM to TensorFlow, showing how parameter size and execution time affect energy consumption. It discusses design and implementation considerations and challenges for fine-grained energy measurement tools. This is solved by using static instrumentation, accounting for factors like computational load and temperature stability.

[3]. The proposed approach uses Deep Reinforcement Learning with PID Lagrangian methods to efficiently schedule jobs in green datacenters, maximizing revenue while minimizing job delays. The system employs feedback control and constraint-controlled reinforcement learning to adapt to dynamic conditions, ensuring stable performance in green datacenters. The CoCoRL scheduler simultaneously achieves higher job value, high system utilization, and a high job completion ratio, making it effective in diverse scenarios.

[4]. The paper's main objective is to enhance resource utilization and reduce energy consumption in data centers through dynamic VM consolidation.the proposed method takes a multivariate approach, considering both current and future resource demands for more accurate predictions.It employs clustering-based stacked bidirectional LSTM deep learning networks to predict memory and CPU usage, resulting in accurate resource forecasts. Extensive simulations using Google's cluster workload traces show substantial improvements in energy efficiency compared to benchmark approaches

[5]. The paper highlights the limitations of current AI algorithms, which prioritize accuracy but result in unsustainable computational demands with negative economic, social, and environmental consequences. It introduces tensor networks (TNs) as a sustainable AI tool capable of improving efficiency while maintaining competitive accuracy. The paper advocates for evaluating AI algorithms based on both accuracy and efficiency, emphasizing the potential advantages of TNs, such as cost reduction, inclusivity enhancement, and reduced environmental impact. TNs are seen as a promising solution that can positively impact sustainability in AI research by reducing computational requirements, lowering costs, enhancing inclusivity, and mitigating environmental emissions.

[6]This paper address,Certainly, here are the points made in a more concise form: Minimize carbon emissions in FL by optimizing parameters while maintaining performance. FL ensures privacy by training models on decentralized devices, without sharing user data. Real-world profiling quantifies FL's carbon footprint, revealing its environmental impact. FL's growing adoption emphasizes the need for sustainability, addressing scalability and energy source diversity.

[7]This paper address,The paper addresses the limitations of traditional AI methods, such as high energy consumption and privacy concerns associated with centralized learning in data centers. It introduces the concept of Federated Learning (FL) as an energy-efficient alternative, distributing learning tasks across low-power edge devices and reducing the need for large centralized infrastructure. The paper proposes a framework to assess energy and carbon footprints in FL, emphasizing the importance of communication efficiency and learner population size for sustainability.Case studies in emerging 5G industry applications highlight FL's potential to trade off energy consumption and accuracy for improved environmental performanceliable fraud detection systems in the evolving landscape of electronic financial transactions.

[8]This paper address,The text focuses on a hybrid GA-PSO FLNN model for cloud resource prediction, targeting challenges like over/under-provisioning and energy consumption.It acknowledges limitations including overlooking resource correlations and the complexity of neural network training. Cloud computing is highlighted for its advantages such as cost reduction, flexibility, and efficient resource management, with a focus on machine learning for accuracy. Performance metrics encompass various aspects like accuracy, resource optimization, cost reduction, QoS enhancement, energy efficiency, scalability, comparative analysis, and future research prospects.

[9]Cloud data centers are essential for cost-effective computing, but energy efficiency is crucial due to environmental concerns. The paper explores clustering, optimization, and machine learning methods for energy-efficient resource allocation. Clustering methods like k-means and hybrid approaches address energy consumption and SLA violations. Optimization methods, including NSGA-II, enhance multi-objective optimization, while machine learning models like DNN and SVM predict energy consumption but have limitations.

[10]The paper highlights the importance of energy efficiency in data centers and focuses on thermal management. It introduces a machine learning-based approach to analyze thermal characteristics, identify overheated areas, and categorize computing nodes.The research aims to offer practical recommendations for enhancing data center thermal profiles and overall energy efficiency. It underscores the potential applicability of these methods to other data center environments.

[11]This paper address, Cloud data centers are essential for cost-effective computing, but energy efficiency is crucial due to environmental concerns. The paper explores clustering, optimization, and machine learning methods for energy-efficient resource allocation. Clustering methods like k-means and hybrid approaches address energy consumption and SLA violations. Optimization methods, including NSGA-II, enhance multi-objective optimization, while machine learning models like DNN and SVM predict energy consumption but have limitations.

[12]This paper address,Cloud data centers are essential for cost-effective computing, but energy efficiency is crucial due to environmental concerns.The paper explores clustering, optimization, and machine learning methods for energy-efficient resource allocation. Clustering methods like k-means and hybrid approaches address energy consumption and SLA violations.Optimization methods, including NSGA-II, enhance multi-objective optimization, while machine learning models like DNN and SVM predict energy consumption but have limitations.

[13]The paper presents a dynamic VM consolidation approach for energy-efficient cloud data centers, addressing the challenge of high energy consumption. This approach categorizes physical machines (PMs) using Markov chain models, introduces a novel VM selection policy, and employs a multi-objective algorithm for optimal VM placement. The goals are to reduce energy consumption, optimize resource utilization, enhance system reliability, and meet SLAs and QoS requirements. The paper's contributions include improved PM categorization, reduced unnecessary VM migrations, and a balanced approach to energy efficiency and reliability.

[14]This paper discussa novel approach that combines fuel cells and deep reinforcement learning to optimize data center energy management. Real-world data center traces show significant improvements, including a 16% reduction in power consumption and a 5% decrease in energy gaps compared to existing methods. The paper acknowledges challenges such as fuel cell limitations and parameter tuning requirements. This approach not only enhances data center efficiency but also promotes greener energy usage, making it a promising solution for data center sustainability.

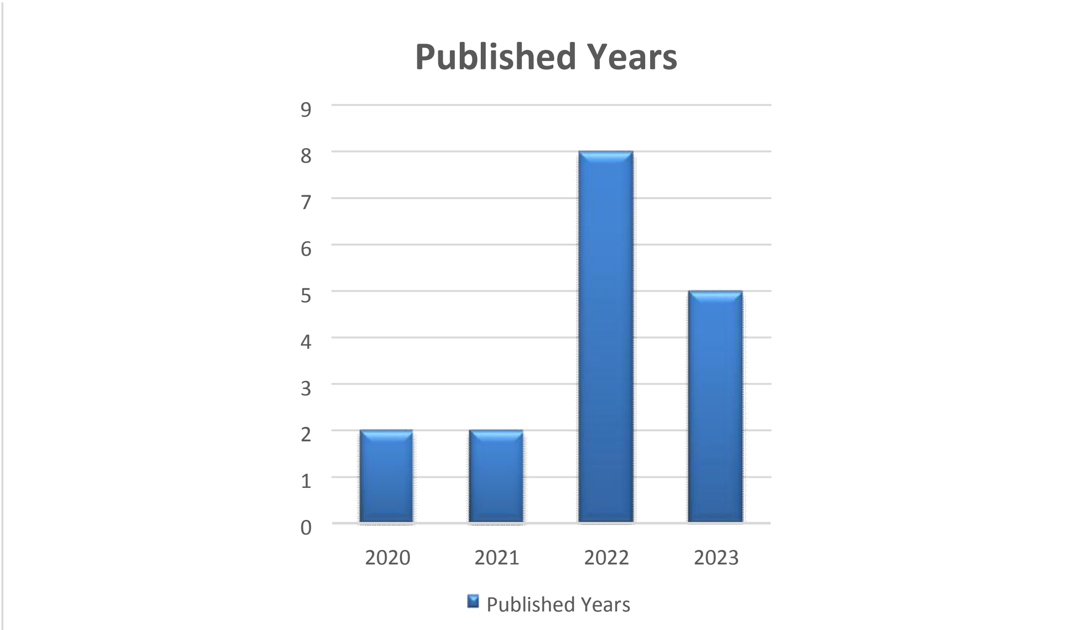
[15]This paper explores ways to make cloud computing use less energy and to protect environment. It does by describing how cloud computing affects the environment is a busy area. The goal of this study is to help people in the industry understand the research and find where more research is needed. Energy-efficient strategies stated here are VM-based, consolidation, bio-based, thermal control, and non-technical approaches.

[16]The paper introduces a novel approach that combines fuel cells and deep reinforcement learning to optimize data center energy management. Real-world data center traces show significant improvements, including a 16% reduction in power consumption and a 5% decrease in energy gaps compared to existing methods.The paper acknowledges challenges such as fuel cell limitations and parameter tuning requirements.This approach not only enhances data center efficiency but also promotes greener energy usage, making it a promising solution for data center sustainability.

[17]Ensemble deep learning refers to the technique of combining models to improve performance and accuracy.Ensemble learning in the context of deep learning refers to the idea of combining multiple individual models to create a more robust and accurate model. The basic concept is that a group of models, when working together, can often outperform any single model on its own. The key to a successful ensemble is diversity among the models. If all models are the same, they might make similar mistakes. Diversity is achieved by training models differently, using different algorithms or subsets of the data.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

III. GRAPHICAL REPRESENTATION



IV. METHODOLODY

Resource allocation is crucial as it directly addresses high energy consumption issues in data centers, ensuring a sustainable and effective operational model.

Resource allocation is crucial in addressing high energy consumption issues because:

- **Optimizing Efficiency:** Proper allocation prevents overutilization or underutilization, ensuring resources are used efficiently, reducing overall energy consumption.
- **Heat Management:** Adequate allocation helps maintain optimal operating levels, minimizing excess heat generation and the subsequent need for energy-intensive cooling.
- **Extended Hardware Lifespan**: Effective allocation reduces strain on components, extending their lifespan. This longevity reduces the need for frequent hardware replacements, which is both energyintensive and resource-consuming.
- **Scalability and Adaptability:** Strategic resource allocation allows for seamless scalability and adaptability, enabling the data center to respond efficiently to changing workloads without unnecessary energy overhead.
- **Environmental Impact:** Green data centers focus on minimizing their environmental footprint. Proper resource allocation directly contributes to reduced energy consumption, lowering greenhouse gas emissions and lessening the data center's environmental impact.
- **Cost Efficiency:** Efficient resource allocation leads to cost savings by minimizing energy consumption, reducing cooling needs, and lowering operational costs associated with maintenance and replacements.

What is Resource allocation in Data centers?

Resource allocation in data centers involves distributing computing resources, such as processing power, memory, storage, and network bandwidth, among various tasks and applications to optimize performance and efficiency. This allocation is typically managed by a combination of hardware and software solutions. Here are key aspects:

Virtualization: Technologies like virtual machines (VMs) or containers enable the creation of multiple isolated instances on a single physical server. Resource allocation involves assigning specific amounts of CPU, RAM, and storage to each virtual instance based on its requirements.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

Load Balancing: Distributing workloads evenly across servers helps prevent resource bottlenecks and ensures that no single server is overloaded. Load balancing algorithms dynamically allocate tasks to servers based on their current capacity and performance.

Dynamic Resource Scaling: Automated systems monitor the data center's workload in real-time and adjust resource allocation dynamically. For example, during periods of high demand, additional resources may be allocated to ensure optimal performance, and vice versa during periods of low demand.

Energy Management: Resource allocation also includes managing power consumption. Techniques such as dynamic voltage and frequency scaling (DVFS) adjust the power supplied to components based on their current workload, optimizing energy usage without compromising performance.

Quality of Service (QoS): Resource allocation considers the QoS requirements of different applications. Critical applications may be allocated higher priority and receive more resources to ensure responsiveness and reliability.

Policy-based Allocation: Administrators can define policies that dictate how resources are allocated based on specific criteria, such as workload type, user priority, or business rules. This helps ensure that resource allocation aligns with organizational goals.

Storage Tiering: In the context of storage, resource allocation involves tiering data based on access frequency and performance requirements. Frequently accessed data is stored on highperformance storage, while less frequently accessed data may be moved to lower-cost, lowerperformance storage.

Network Bandwidth Allocation: Efficient data center operation requires managing network resources. Bandwidth allocation involves prioritizing and distributing network resources to ensure smooth communication between servers, storage, and other components.

Here we are addressing Dynamic Resource Scaling using the following ideas:

1. Full Scaling Automation(FSA)model

2. Resource allocation in cognitive radio system

Base Paper :Wang, S., Sun, Y., Shi, X., Zhu, S., Ma, L. T., Zhang, J., ... & Liu, J. (2023). Full Scaling Automation for Sustainable Development of Green Data Centers. ***arXiv preprint arXiv:2305.00706*****.**

This paper involves 2 main goals,

1:With the help of time series database we can forecast the workloads, Based on long term observation of workloads by finding the trends associated and how long the trend continued i.e period this is achieved using Multi scale Time series Representation. Fusion module is then used to integrate the present state of the workload using nearby observation and the pattern identified in historical data using Multi scale Time series using Representation enhanced deep auto-aggressive model there by on decoding produces future workload prediction.

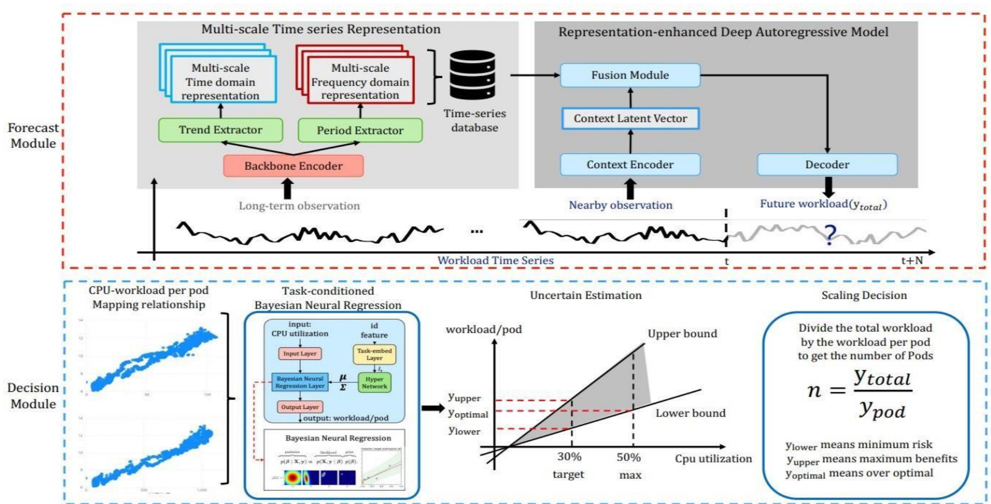


Fig: Architecture of proposed FSA Model

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

2:Decrease in workload leads to low usage of CPU resources this leads to wasting of energy. so estimate the uncertainty Task Conditioned Bayesian Neural Regression is usedto predict workload per pod thereby we can find the minimum, maximum, optimal cpu utilization for optimal usage of resources for dynamic workloads

Methods:

The following methods are used to built the model,

1. Time Series Representation Method:

It is a method to compress long term historical time series with various periodicities into compressed representation vectors.

Significance: The representation method is crucial for handling non stationary and high frequency time series data. The compressed representations capture complex long term dependencies, allowing the model to learn and make predictions efficiently.

2. CNF based Deep Autoregressive Fusing Model:

This model uses a Continuous Normalizing Flow (CNF) to capture the temporal dependencies in time series data. It fuses long term historical time series representations with short term observations using autoregressive modeling.

Significance: CNFs are powerful for modeling complex distributions, and the autoregressive fusing model allows the framework to combine information from historical data and recent observations, enabling accurate time series predictions.

3. Task Conditioned Bayesian Neural Regression: This model utilizes Bayesian neural networks to estimate uncertainty and includes a task conditioned hypernetwork. It involves variational Bayes and Stochastic Gradient Variational Bayes (SGVB) for training**.**

Significance: Bayesian Neural Regression is valuable for estimating uncertainty in predictions. The task conditioned approach allows the model to adapt to different scenarios efficiently. The uncertainty estimation is crucial for decision making, providing upper and lower bounds for per Pod workload.

Multi Scale Time Series Representation: The method involves random cropping and timestamp masking for data augmentation. It uses a Convolutional Transformer (ConvTrans) for contextual embedding, extracting trends in the time domain and periodicities in the frequency domain.

Significance: Multi scale representation allows the model to capture different temporal scales in the time series data. Data augmentation techniques like random cropping and timestamp masking enhance the model's ability to generalize and learn complex dependencies.

Frequency Domain Contrastive Learning and Time Domain Contrastive Learning:

These involve using Fast Fourier Transforms (FFT) for frequency domain representation and 1d causal convolution layers for time domain representation. Contrastive learning is used for discriminative learning in both domains.

Significance: Contrastive learning helps the model to learn representations that are discriminative in both the time and frequency domains. FFT captures periodic patterns, and causal convolution layers extract trends, contributing to a comprehensive understanding of the time series data.

Task Projector Network and Hypernetwork:

The task projector network (MLP) is used to create task embeddings, and the hypernetwork generates the parameters for the Bayesian model.

Significance: Task embeddings help the model adapt to different scenarios, and the hypernetwork provides a task conditioned approach to Bayesian modeling, improving the model's efficiency and effectiveness across various scenarios.

Forecast, Scale, and Adapt (FSA) Method:

The FSA method involves running the trained model multiple times to generate samples, calculating mean and variance for decision making. It uses the Task Conditioned Bayesian Neural Regression to provide upper and lower bounds for scaling Pods.

Significance: The FSA method effectively balances benefits and risks in the decision making process. By incorporating uncertainty estimation, it provides a decision range for scaling Pods, ensuring stable Service Level Objectives (SLOs).

Reference-1:[19] Lee, W., & Chung, B. C. (2023). Ensemble deep learning based resource allocation for multi-channel underlay cognitive radio system. *ICT Express*, *9*(4), 642-647.

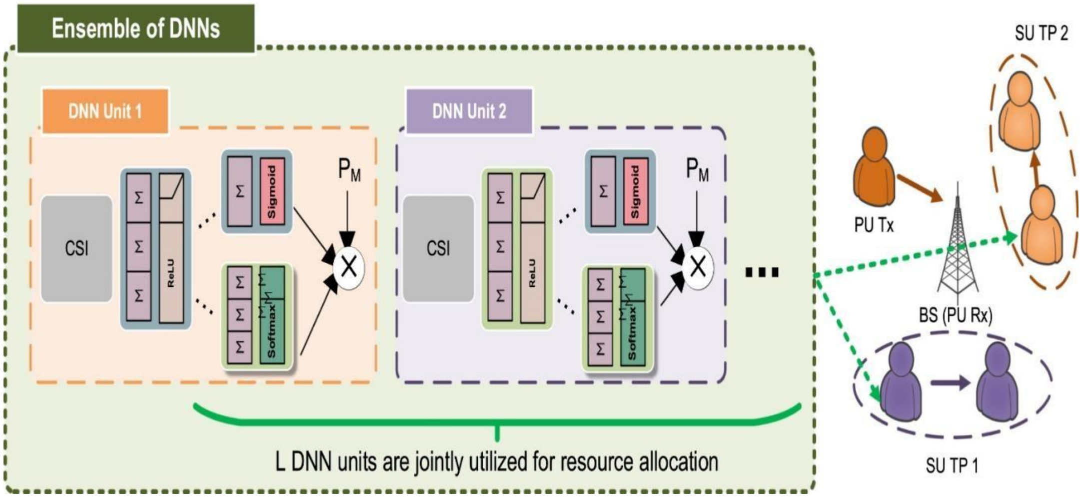


Fig: Architecture of cognitive radio system:

Ensemble deep learning refers to the technique of combining models to improve performance and accuracy

Ensemble learning in the context of deep learning refers to the idea of combining multiple individual models to create a more robust and accurate model. The basic concept is that a group of models, when working together, can often outperform any single model on its own.

The key to a successful ensemble is diversity among the models. If all models are the same, they might make similar mistakes. Diversity is achieved by training models differently, using different algorithms or subsets of the data.

when you want a prediction, each model gives its opinion. This can be in the form of voting (classification problems) or averaging (regression problems).

Voting Example (Classification):to go with the majority vote and predict

Averaging Example (Regression): If we are predicting a numerical value ,each model gives its prediction, and you take the average to get the final prediction.

Methods:

In this architecture, methods used are multiple deep learning models are combined to make effective resource allocation decisions.

The architecture consists of three main components:

Data Collection: The system collects data related to the available channels, spectrum availability, and quality parameters. This data is used to train the deep learning models.

gathers information about the channels, such as their occupancy and interference levels, as well as the quality of the received signals. This data forms the training dataset for the deep learning models

Ensemble Deep Learning Models: Multiple deep learning models, FNN, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), are trained using the collected data. Each model focuses on different aspects of the resource allocation problem, capturing various features and patterns.

Weighted Averaging:

Idea: Assign different weights to the predictions of the different models and calculate the weighted average.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 3, December 2023

Implementation: Weights can be assigned based on the performance of each model on a validation set. Optimal weights can be found through experimentation.

Voting:

Idea: Allow each model to vote on the predicted outcome, and the final prediction is determined by majority voting.

Implementation: For binary classification, if a majority of models predict '1', the ensemble prediction is '1'; otherwise, it's '0'. For regression, you can take the average of the predictions.

The architecture utilizes ensemble deep learning models to improve resource allocation decisions. These models work together to capture different aspects of the resource allocation problem.

Feed forward neural network (FNN) is a type of neural network that processes information in a one directional manner, from input to output. In the context of deep learning based resource allocation for multi channel underlay cognitive radio systems, an FNN can learn patterns and relationships from data to make optimal decisions on resource allocation. It considers factors like channel quality, user demands, and system constraints to allocate frequency channels efficiently and minimize interference. This approach enhances the overall performance of the system

The outputs of these models are then combined using ensemble techniques such as majority voting or weighted averaging to generate the final resource allocation decision.

The goal of utilizing ensemble deep learning models is to enhance the robustness and reliability of the decision making process. By leveraging multiple models, the architecture can adapt to dynamic changes in the spectrum and make more accurate resource allocation decisions.

Decision Making: The outputs of the individual deep learning models are combined using an ensemble technique, such as majority voting or weighted averaging, to make the final resource allocation decision. This ensemble approach helps to leverage the strengths of each model and improve the overall accuracy and robustness of the resource allocation process.

By utilizing ensemble deep learning, this architecture aims to effectively allocate resources in a multi channel underlay cognitive radio system, optimizing spectrum utilization and ensuring efficient communication. The ensemble approach allows for more reliable decision making and enhances the system's adaptability to dynamic and changing spectrum conditions

Paper 2: Joloudari, J. H., Mojrian, S., Saadatfar, H., Nodehi, I., Fazl, F., Alizadehsani, R., ... & Acharya, U. R. (2022). The state-of-the-art review on resource allocation problem using artificial intelligence methods on various computing paradigms. *arXiv preprint arXiv:2203.12315*.

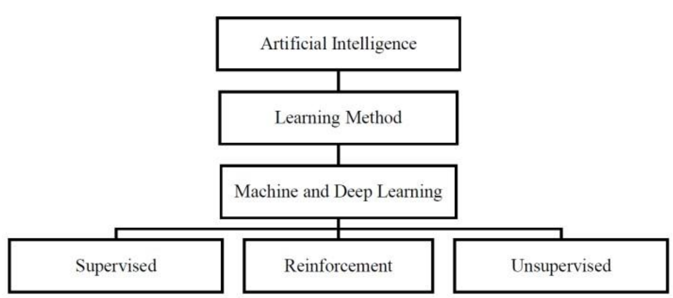


Fig: Machine learning methods for resource allocation problems in different computational environments.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

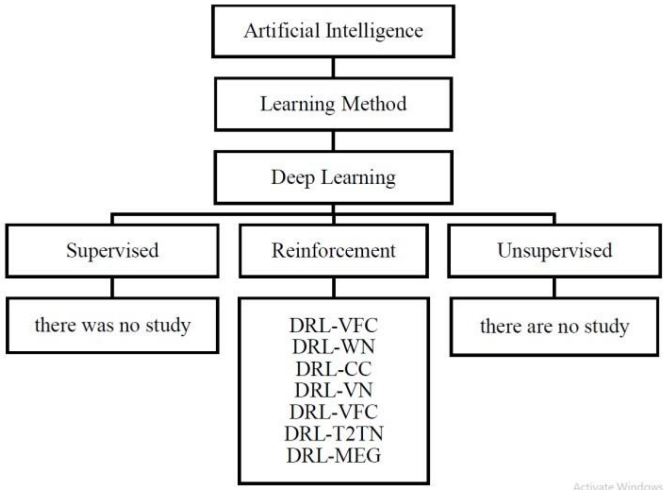


Fig: DL methods for the resource allocation problem in various computational paradigms.

3 Steps involved here are

1. Data Collection: The authors conducted a comprehensive literature study (CLS) on resource allocation problems, reviewing articles conducted up to 2020, and two taxonomies underlying machine learning and deep learning methods .

2. Data Processing: The comparison between artificial intelligence methods in resource allocation is described in the third section of the paper .

3. Data Analysis: The paper presents open research challenges in resource allocation on multilayer computing environments and concludes with future research work .

Methods:

The stated methods and their effectiveness in different fields:

Cloud Computing: The paper proposes the use of Markov Decision Process (MDP) combined with Bayesian learning to optimize the cost of allocating dynamic resources in cloud computing . This method is effective in optimizing resource allocation in cloud computing environments.

Edge Computing: Reinforcement and heuristic learning methods are categorized for public safety communications on 5G networks, which are suitable for addressing resource allocation challenges in edge computing environments .

Fog Computing and Vehicular Networks: The paper provides an overview of applying deep learning technology to wireless resource allocation with a focus on vehicular networks . Deep reinforcement learning (DRL) is highlighted as a method for improving resource allocation in vehicular networks.

Wireless Networks: The paper discusses the use of machine learning and deep learning methods for resource allocation in wireless networks. Specifically, the Evolutionary RL/GA algorithm and the Valuebased RL/Q-learning algorithm are mentioned as effective methods for resource allocation in wireless networks .

Machine-to-Machine Communication: The use of Q-learning algorithm with K-means clustering is proposed for gap allocation for machine-type communication devices in machine-to-machine communication .

Mobile Edge Computing: Reinforced deep learning, specifically the Deep Reinforcement Learning based Resource Allocation (DRLRA) method, is proposed for resource allocation in mobile edge computing. The DRLRA method improves the average service time as the request aggregation districts number increases compared to the OSPF method .

Vehicular Fog Computing: Distributed deep reinforcement learning is proposed for computational resources management, resource allocation, and system complexity reduction in vehicular fog computing environments. A contract-based incentive mechanism for the allocation of resources in the vehicular fog network is also suggested .

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 3, December 2023

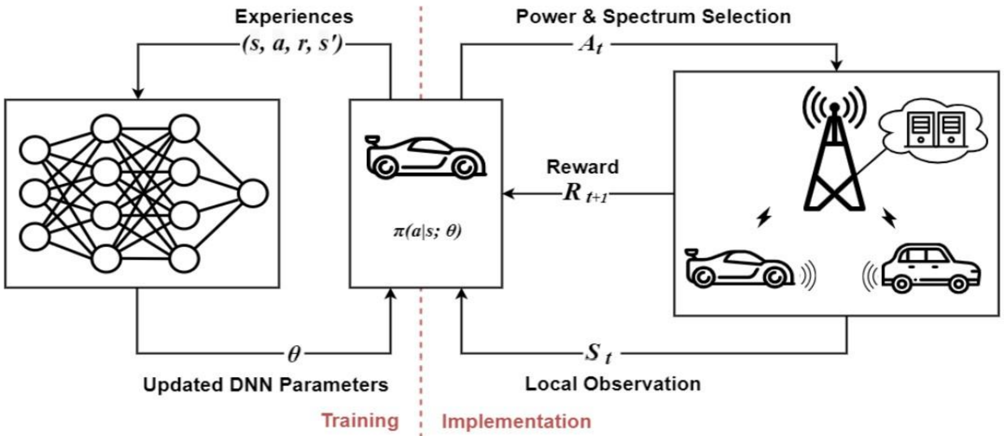
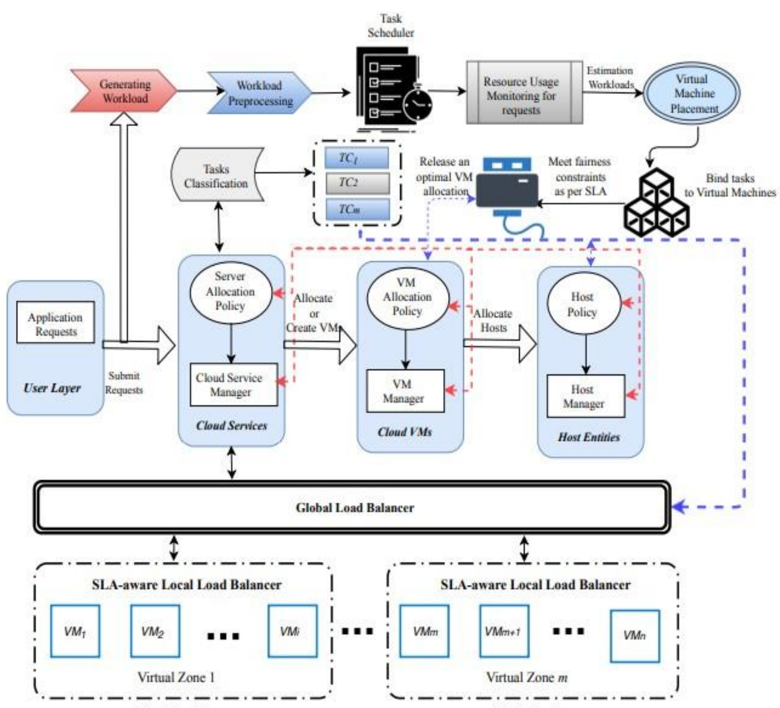


Fig: Resource Allocation in Vehicular Fog Computing

Train-to-Train Communication: Multi-Agent Deep Reinforcement Learning (MADRL) is used to reduce co-channel interference, prevent collisions, and increase system power in the proposed smart resource allocation method in train-to-train communication. The multi-agent deep Q-network method is also used to improve the throughput of any train-to-train connection and the system throughput .

These methods show promise in optimizing resource allocation in various computing paradigms, addressing the specific challenges and requirements of each environment.

V. CASE STUDY



IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 3, December 2023

Whenever a User places a request in cloud, User requests are received by the cloud data center network. The requests are then processed by the load manager, which analyzes the resource requirements of each application, such as CPU, Memory, Energy, and Bandwidth usage. Based on the resource requirements, an appropriate number of virtual machines (VMs) are allocated for each application. The resource information is collected and updated, and the resources are sorted into four queues based on their loads: CPU intensive, Memory intensive, Energy intensive, and Bandwidth intensive. The scheduling method, known as Dynamic Resource Allocation for Load Balancing (DRALB), ensures SLA-aware scheduling by considering the specific requirements and constraints outlined in the Service Level Agreements (SLAs).The scheduling method aims to maximize cloud profits by minimizing resource utilization and avoiding SLA violation penalties.Once the requests are processed and the resources are allocated, the cloud data center network sends the replies back to the users

The two parameters mainly focused here are: response time (RT) and Resource Utilization criteria (RUC). Dynamic Resource Allocation for Load Balancing (DRALB) method is used to address load imbalances in cloud data center networks is used. The method involves analyzing resource requirements and allocating an appropriate number of virtual machines (VMs) for users as traffic of users entering can be varying with time.Resources are then sorted into four queues based on their loads: CPU intensive, Memory intensive, Energy intensive, and Bandwidth intensive. The method aims to improve scheduling efficiency, resource utilization, and SLA-aware scheduling for cloud consumers.

Parameters	Value
VM Setup of Data Center	
CPU Computing ability	1860 MIPs, 2660 MIPs
Disk I/O	8 GB
RAM	4096 MB
Bandwidth	100 M/s
Storage	10 G
Task Setup of Data Center	
Length (CPU)	[250-1000] MIPs
File Size	[100-2000] MB
Output size (Memory)	[20-40] MB

Methods Involved here in the DRAM for Load Balance Scheduling Over Cloud Data Center Networks:

Load Manager Analysis : The load manager analyzes the resource requirements of each application, such as CPU, Memory, Energy, and Bandwidth usage, to determine the appropriate number of virtual machines (VMs) needed for each application .

Resource Allocation: Based on the analysis, the load manager allocates the required number of VMs for each application, ensuring that the resources are efficiently tilized .

Resource Information Collection and Update: The resource information is collected and updated, and the resources are sorted into four queues based on their loads: CPU intensive, Memory intensive, Energy intensive, and Bandwidth intensive .

SLA-Aware Scheduling: The scheduling method, known as Dynamic Resource Allocation for Load Balancing (DRALB), ensures SLA-aware scheduling by considering the specific requirements and constraints outlined in the Service Level Agreements (SLAs) .

Optimal Resource Search: The method is based on the diversity of client's applications and involves searching for the optimal resources for the particular deployment, considering factors such as average response time, resource utilization, SLA violation rate, and load balancing .

Traffic Reduction: The experimental results demonstrate that this method can reduce the wastage of resources and reduce traffic by up to 44.89% and 58.49% in the network [1].

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

These methods work together to dynamically allocate resources, ensure SLA-aware scheduling, and optimize resource utilization in cloud data center networks.

Allocation Policies	When, T < R				Average Wastage	When, T > R				Average Wastage
	When, T < R					When, T > R				
RND	66.42	69.61	89.62	70.33	49.53%	69.67	78.72	70.99	85.68	40.73%
SEQ	71.14	78.41	75.61	80.52	33.24%	73.42	76.14	79.52	80.42	30.51%
DHLB	73.62	77.52	80.01	76.21	27.98%	76.42	79.32	83.67	79.42	22.61%
DRALB	70.33	79.39	75.52	77.67	20.67%	75.11	80.12	82.55	81.52	18.31%

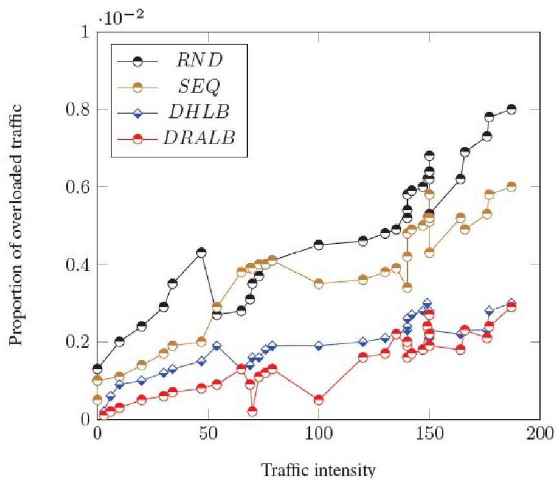


Fig: Improved Resource Utilization on using DRALB

DRALB improves resource utilization by analyzing resource requirements such as CPU, Memory, Energy, and Bandwidth usage and allocating an appropriate number of virtual machines (VMs) for each application .By sorting resources into four queues based on their loads (CPU intensive, Memory intensive, Energy intensive, and Bandwidth intensive), DRALB ensures that resources are allocated efficiently and effectively.This method aims to reduce resource wastage by matching the specific resource needs of each application with the available resources, thereby maximizing resource utilization .The experimental results demonstrate that DRALB can reduce the wastage of resources and improve resource utilization in cloud data center networks .By optimizing resource allocation and load balancing, DRALB minimizes the traffic in the network, reducing it by up to 44.89% and 58.49% .Overall, DRALB's dynamic resource allocation approach ensures that resources are utilized optimally, leading to improved efficiency and reduced resource wastage in cloud data center networks

VI. RESULTS

The paper Full Scaling Automation for Sustainable Development of Green Data Centers, presents an in-depth exploration of advanced automation methodologies to promote the sustainable evolution of green data centers. It’s contributions include novel strategies for resource allocation optimization and efficiency improvement within data centers. It involves comprehensive, full-scale automation approach, incorporating state-of-the-art technologies to address environmental sustainability challenges .Incorporating findings of paper Ensemble deep learning based resource allocation for multi channel underlay cognitive radio system, introduces an innovative ensemble deep learning-based framework for resource allocation specifically tailored for cognitive radio systems. This approach leverages multiple learning models to enhance decision-making in dynamically changing environments, providing a technically sophisticated solution to resource optimization challenges. The Reference the state-of-the-art review on resource allocation problem using artificial intelligence methods on various computing paradigms, adds a layer of technical depth by discussing the latest artificial intelligence methods applied to the resource allocation problem across

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

diverse computing paradigms. This includes advancements in machine learning, optimization algorithms, and intelligent decision-making systems, further enriching the technical discourse on sustainable data center development.

In summary, this paper not only emphasizes the importance of full-scale automation for green data centers but also incorporates technically sophisticated approaches, such as ensemble deep learning and artificial intelligence methods, to address resource allocation challenges in a sustainable and eco-friendly manner.

VII. DISCUSSION

The integration of ensemble deep learning techniques for resource allocation in cognitive radio systems, as cited in the main paper, brings a significant advancement to the field. The use of multiple models working in concert enables more robust decision-making in the face of complex and dynamic data center environments. This approach aligns with the goals of the main paper by enhancing resource optimization and efficiency, providing a promising avenue for addressing the challenges of sustainable data center development.

The incorporation of Joloudari et al.'s review on the state-of-the-art AI methods for resource allocation is crucial for understanding the broader landscape of technological solutions. The review brings to light the latest advancements in artificial intelligence, including machine learning algorithms and optimization strategies. This comprehensive overview enriches the main paper's discussion, emphasizing the importance of staying abreast of cutting-edge methodologies to address resource allocation issues across diverse computing paradigms.

The main paper advocates for full-scale automation as a holistic solution for sustainable development in green data centers. By combining the insights from the cited papers, the discussion emphasizes the technical depth brought by ensemble deep learning and state-of-the-art AI methods. The ensemble learning approach enhances adaptability, while the AI methods provide intelligent decision-making capabilities, aligning with the broader goals of resource optimization and efficiency improvement in the pursuit of sustainable data centers.

In conclusion, the combination of these papers underscores the importance of embracing advanced technologies such as ensemble deep learning and cutting-edge AI methods to achieve full-scale automation for sustainable green data centers. The synergy of these approaches contributes to a more comprehensive understanding of the technical landscape and paves the way for innovative solutions to address the evolving challenges in the field.

VIII. CONCLUSION

The concept of ensemble deep learning based resource allocation for multi channel underlay cognitive radio systems and full scaling automation for sustainable development of green data centers involves optimizing resource usage. While the specific challenges and environments are different, there are technical parallels. We can relate them and the methods that can be applied to achieve efficient resource allocation in green data centers, Cognitive Radio Systems. Cognitive Radio Systems Adjusts communication parameters dynamically for optimal spectrum use like Allocating computing resources dynamically based on workload demands and energy efficiency requirements in Data centers. Radio systems utilizes ensemble deep learning for predicting and optimizing channel allocation. Uses spectrum sensing to detect available channels and avoid interference where as data centers uses machine learning models to predict and optimize resource allocation, considering factors like energy consumption and performance. The idea of sustainability aligns with optimizing resource usage in both domains. Introducing DRL helps doing these tasks more efficiently, as the model performs a task then gets a reward which may be positive or negative depending on the action performed, it constantly updates its policy or value function, and repeats the cycle to improve its decision making over time. The neural network serves as a function approximator, capturing the mapping between states and actions or Qvalues. Techniques like experience replay, target networks, and exploration strategies are often employed to enhance stability and efficiency during training.

By drawing parallels between these concepts of Resource Allocation in Cognitive Radio systems and Green Data Centers, we can apply similar technical principles such as dynamic resource allocation, machine learning, monitoring, QoS considerations, adaptability, optimization algorithms, and automation to achieve sustainable resource allocation in green data centers. The goal of optimized resource usage while considering environmental impact and operational efficiency can be reached by using these techniques.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

REFERENCES

[1] Wang, S., Sun, Y., Shi, X., Zhu, S., Ma, L. T., Zhang, J., ... & Liu, J. (2023). Full Scaling Automation for Sustainable Development of Green Data Centers. arXiv preprint arXiv:2305.00706.(Base paper)

[2] Rajput, S., Widmayer, T., Shang, Z., Kechagia, M., Sarro, F., & Sharma, T. (2023). FECoM: A Step towards Fine-Grained Energy Measurement for Deep Learning. arXiv preprint arXiv:2308.12264.

[3] Venkataswamy, V. (2022). Job Scheduling in Datacenters using Constraint Controlled RL. arXiv preprint arXiv:2211.05338.). Deep learning-based multivariate resource utilization prediction for hotspots and coldspots mitigation in green cloud data centers. The Journal of Supercomputing

[4] Patel, Y. S., Jaiswal, R., &Misra, R. (2022). Deep learning-based multivariate resource utilization prediction for hotspots and coldspots mitigation in green cloud data centers. The Journal of Supercomputing, 78(4), 5806-5855.

[5] Memmel, E., Menzen, C., Schuurmans, J., Wesel, F., &Batselier, K. (2022). Towards Green AI with tensor networks--Sustainability and innovation enabled by efficient algorithms. arXiv preprint arXiv:2205.12961.

[6] Yousefpour, A., Guo, S., Shenoy, A., Ghosh, S., Stock, P., Maeng, K., ... &Mironov, I. (2023).Green Federated Learning. arXiv preprint arXiv:2303.14604.

[7] Savazzi, S., Rampa, V., Kianoush, S., & Bennis, M. (2022). An energy and carbon footprint analysis of distributed and federated learning. IEEE Transactions on Green Communications and Networking, 7(1), 248-264.

[8] Malik, S., Tahir, M., Sardaraz, M., &Alourani, A. (2022). A resource utilization prediction model for cloud data centers using evolutionary algorithms and machine learning techniques. Applied Sciences, 12(4), 2160.

[9] Jayaprakash, S., Nagarajan, M. D., Prado, R. P. D., Subramanian, S., &Divakarachari, P. B.(2021). A systematic review of energy management strategies for resource allocation in the cloud: Clustering, optimization and machine learning. Energies, 14(17), 5322.

[10] Grishina, A., Chinnici, M., Kor, A. L., Rondeau, E., & Georges, J. P. (2020). A machine learning solution for data center thermal characteristics analysis. Energies, 13(17), 4378.

[11] Jayaprakash, S., Nagarajan, M. D., Prado, R. P. D., Subramanian, S., &Divakarachari, P. B.(2021). A systematic review of energy management strategies for resource allocation in the cloud: Clustering, optimization and machine learning. Energies, 14(17), 5322.

[12] Zeng, J., Ding, D., Kang, K., Xie, H., & Yin, Q. (2022). Adaptive DRL-based virtual machine consolidation in energy-efficient cloud data center. IEEE Transactions on Parallel and Distributed Systems, 33(11), 2991-3002.

[13] Zhou, Q., Xu, M., Gill, S. S., Gao, C., Tian, W., Xu, C., &Buyya, R. (2020, May). Energy efficient algorithms based on VM consolidation for cloud computing: comparisons and evaluations.In2020 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID) (pp. 489-498). IEEE.

[14] Hu, X., & Sun, Y. (2020). A deep reinforcement learning-based power resource management for fuel cell powered data centers. Electronics, 9(12), 2054.

[15] Bharany, S., Sharma, S., Khalaf, O. I., Abdulsahib, G. M., Al Humaimeedy, A. S., Aldhyani, T. H., ... &Alkahtani, H. (2022). A systematic survey on energy-efficient techniques in sustainable cloud computing. Sustainability, 14(10), 6256.

[16] Hu, X., & Sun, Y. (2020). A deep reinforcement learning-based power resource management for fuel cell powered data centers. Electronics, 9(12), 2054.

[17] Lee, W., & Chung, B. C. (2023). Ensemble deep learning based resource allocation for multichannel underlay cognitive radio system. ICT Express, 9(4), 642-647.

[18] Salles, A. C., Lunardi, G. L., & Thompson, F. (2022). A Framework Proposal to Assess the Maturity of Green IT in Organizations. Sustainability, 14(19), 12348.

[19] Islam, R. U., Ruci, X., Hossain, M. S., Andersson, K., &Kor, A. L. (2019). Capacity management of hyperscale data centers using predictive modelling. Energies, 12(18), 3438.

[20] Joloudari, J. H., Mojrian, S., Saadatfar, H., Nodehi, I., Fazl, F., Alizadehsani, R., ... & Acharya, U. R. (2022). The state-of-the-art review on resource allocation problem using artificial intelligence methods on various computing paradigms. arXiv preprint arXiv:2203.12315.