iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** VI **Month of publication:** June 2025

DOI: https://doi.org/10.22214/ijraset.2025.72643

Resource Allocation and Load Balancing for Fog Computing Using Artificial Intelligence Techniques

Sadhna Yadav, Dr. Deepak Kumar Verma
Chhartrapati Shahu Ji Maharaj University, Kanpur

Abstract: Fog computing (FC) has emerged as a critical paradigm for enabling low-latency and efficient data processing in Internet of Things (IoT) environments. However, resource allocation and load balancing remain significant challenges due to the heterogeneous and dynamic nature of FC networks. This review systematically analyses recent advancements in artificial intelligence (AI)-driven approaches for optimizing resource allocation and load balancing in FC, with a focus on studies published between 2019 and 2024. The survey highlights the role of machine learning (ML), deep learning (DL), and reinforcement learning (RL) in intelligent resource allocation, ensuring efficient task offloading and execution. Additionally, it explores heuristic and nature-inspired meta-heuristic algorithms for dynamic load balancing, improving system throughput, energy efficiency, and Quality of Service (QoS). While these techniques have demonstrated significant improvements in FC performance, challenges such as real-world implementation complexity, scalability, and system heterogeneity persist. The review identifies future research directions, emphasizing the need for advanced AI-driven frameworks and deep reinforcement learning techniques to enhance resource management and load balancing in distributed FC environments.

Keywords: Fog Computing (FC), Internet of Things (IoT), Resource Allocation, Load Balancing, Artificial Intelligence (AI)

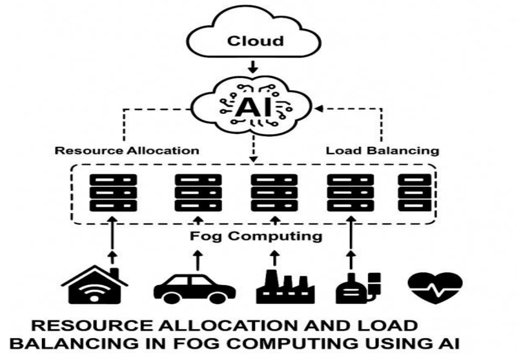
I. INTRODUCTION

Fog Computing (FC) has emerged as a crucial framework for enabling low-latency and effective data processing in Internet of Things (IoT) environments [1]. Even with its benefits, overseeing resources and equalizing computational loads continues to be challenging because of the naturally varied and changing traits of FC networks [2]. This review thoroughly analyzes recent advancements in AI-driven techniques for improving resource allocation and load balancing in FC environments, concentrating on publications from 2019 to 2025. It examines how machine learning (ML), deep learning (DL), and reinforcement learning (RL) contribute to smart resource allocation, efficient task offloading, and execution methods. The research additionally assesses nature-inspired heuristic and meta-heuristic algorithms for adaptive load balancing, enhancing throughput, energy efficiency, and overall Quality of Service (QoS). Despite the notable performance improvements from these AI-driven approaches, numerous problems persist, such as challenges in real-world implementation, constraints on scalability, and variations in network characteristics. The document discusses possible directions for upcoming studies, promoting the creation of advanced AI systems and the use of deep reinforcement learning to enhance resource and load management in decentralized FC environments

The Internet of Things (IoT) has emerged as a revolutionary influence in numerous industries, including healthcare, city infrastructure, transportation, and agriculture. It includes a wide range of networks of interconnected devices equipped with actuators, sensors, and processors, consistently generating and sharing large amount of data. These devices operates independently or interact directly with users, making IoT an crucial catalyst for smart technologies and real-time decision-making [3][4].

However, the rapid increase in the number of connected IoT devices introduces several challenges. These include limited storage, increased demands for data processing, the need to maintain consistent Quality of Service (QoS), constrained network bandwidth, and higher latency. Conventional cloud-based architectures, which depend on transmitting data to remote centralized servers for processing, are becoming increasingly insufficient. They frequently suffer from network congestion, high latency, and elevated energy usage, render them inappropriate for latency-sensitive applications such as autonomous vehicles, essential healthcare systems, and industrial automation task [5][6].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VI June 2025- Available at www.ijraset.com



To mitigate these challenges, Fog Computing (FC) has emerged as an intermediate layer connecting IoT devices and cloud infrastructure. FC decentralizes computing by locating processing resources near to the data source — at or closes the edge of the network through fog nodes. These nodes can execute local data processing, significantly reducing the dependency on cloud data centers. This model leads to more efficient bandwidth use, reduced latency, and improved responsiveness of IoT systems [7]. Additionally, FC improves privacy and security by locally processing sensitive information before transmitting only the necessary data to the cloud.

However, for fog computing to function effectively, it must address two key functional aspects:

Resource Allocation – Proper distribution of processing resources such as memory, CPU, and storage between fog nodes to enable effective task performance and balanced utilization.

Load Balancing – The balanced distribution of processing workloads across multiple fog nodes to reducing delays, prevent bottlenecks, and ensuring high system efficiency.

II. CHALLENGES IN RESOURCE ALLOCATION AND LOAD BALANCING IN FOG COMPUTING

Fog computing eco system is heterogeneous and dynamic in nature, which consist of a combination of devices with varying computing abilities, connectivity aspects, and energy limitations. However, Resource management in Conventional cloud computing is centralized whereas decentralized and smart resource allocation strategies are required in fog environments.

A. *Main Challenges in fog computing consists of:*

1) *Variable Workload Dynamics:* Dynamic workload is generated by IOT applications and thus workload distribution at some fog nodes are over loaded while others remains under loaded.
2) *Latency Requirements:* Applications such as real-time medical diagnosis and self-driving vehicles need extremely low latency and thus smart workload management at the fog level becomes necessary.
3) *Energy Efficiency:* Fog nodes which operate on battery-powered edge devices need energy-efficient resource distribution methods to extend their operational duration.
4) *Scalability challenges:* Due to immense increase in number of IOT devices and fog nodes, a scalable load balancing methods are required that can accommodate large-scale deployments.
5) *Security and Privacy Issues:* Since IOT Devices tasks are distributed among fog nodes presents possible vulnerabilities requires secure and reliable resource allocation methods.

B. *Resource Allocation and Load Balancing Using AI-Driven Methods:*

The integration of Artificial Intelligence (AI) into fog computing eco system is considered as a efficient approach to address issues of resource allocation and load balancing. Machine Learning (ML), Reinforcement Learning (RL), and Deep Learning (DL) provides smart decision-making, and thus enable fog nodes to dynamically adjust to varying workload, provide promising resource allocation strategies, and improve overall efficiency of the system.[8][9]

C. *Resource allocation using Artificial Intelligence:*

Effective resource allocation guarantees that computational resources are efficiently assigned across several fog nodes to increase throughput, reduce processing delays, and elevate Quality of Services.

AT-based solutions for resource allocation in Fog computing environment consist of:

1) *Task Scheduling Using Machine Learning*: Supervised learning algorithms Predict workload requirements using previous data, facilitating proactive task scheduling[10].
2) *Predictive Resource Management using Deep Learning*: Neural networks evaluate real-time workload load to dynamically assign resources and avoid congestion [10].
3) *Adaptive Allocation through Reinforcement Learning (RL)***:** Reinforcement learning models enable self-directed decision-making, enabling fog nodes to discover unique allocation methods according to the conditions of the surrounding[11].
4) *Heuristic and Meta-Heuristic Algorithms*: Techniques like Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO) are used for resource allocation issues to find efficient and near-optimal solutions. Load Distribution with AI Load balancing methods seek to allocate tasks effectively among various fog nodes, guaranteeing balanced resource allocation and avoiding network overload. [12]

D. AI-Powered Load Balancing Techniques Comprise:
1) *Static vs. Dynamic Load Balancing:* Static approaches uses fixed criteria for distributing tasks, while dynamic AI-based methods modify task assignments in real-time according to system status.
2) *Load Balancing Using Meta-Heuristics:* Ant Colony Optimization (ACO) and Genetic Algorithms (GA) are employed to enhance load distribution, reducing response times and avoiding overload on nodes
3) *Reinforcement Learning for Efficient Load Distribution:* RL-driven models uses the best strategies for workload distribution, continually enhancing load balancing strategies
4) *Load Balancing through Software-Defined Networking (SDN):* SDN-powered FC systems leverage AI to enhance resource distribution across the network and adjust workload routing in response to current traffic scenarios

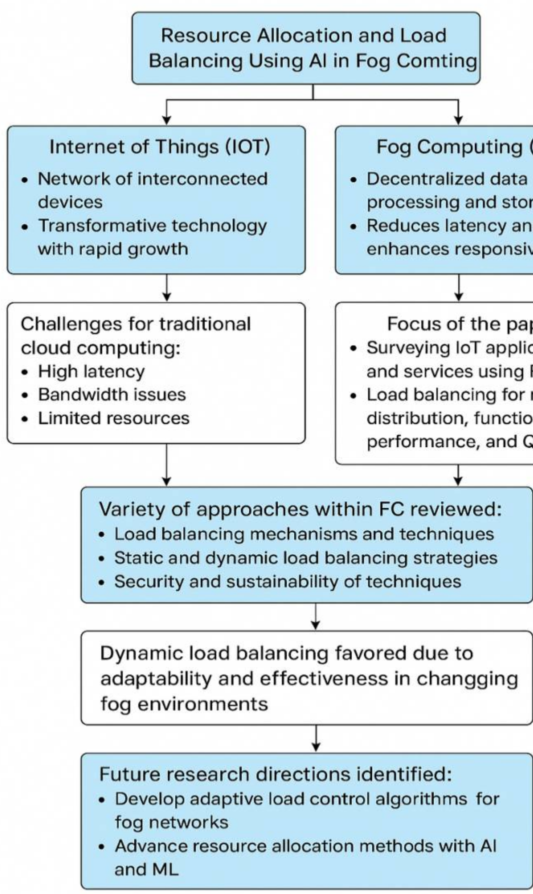


Figure: Conceptual Frame work for resource allocation and load balancing Using AI in FOG computing environment

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VI June 2025- Available at www.ijraset.com

III. COMPARATIVE ANALYSIS OF AI-BASED TECHNIQUES

Various AI based methods have been proposed to improve **resource allocation and load balancing** in fog computing environment. An analysis comparing these methods shows that:

AI Technique	**Application in FC**	**Advantages**	**Limitations**
Machine Learning (ML)	Task prediction & scheduling	Fast decision-making, adaptive	Requires large training datasets
Deep Learning (DL)	Predictive resource allocation	High accuracy in pattern recognition	Computationally intensive
Reinforcement Learning (RL)	Adaptive workload balancing	Self-learning, real-time optimization	Slow convergence in complex environments
Genetic Algorithms (GA)	Load balancing optimization	Efficient in large-scale environments	May require high computational time
Particle Swarm Optimization (PSO)	Task scheduling & allocation	Converges quickly, efficient resource utilization	May get trapped in local optima
Ant Colony Optimization (ACO)	Dynamic load balancing	Self-adaptive, scalable	Slow for large-scale networks

IV. STATE-OF-THE-ART SOLUTIONS AND FUTURE DIRECTIONS

Recent studies have shown considerable progress in AI baesd resource allocation and load balancing within Fog Computing. Some **state-of-the-art solutions** includes:

- Secure and sustainable load balancing strategies for Edge Data Centers (EDCs), ensuring efficient task allocation in distributed FC environments.
- Integration of Deep Reinforcement Learning (DRL) for self-learning and adaptive resource allocation in real-time applications.
- Development of AI-powered SDN frameworks for intelligent load balancing, reducing latency and improving scalability.
- Energy-aware resource allocation strategies using AI to enhance the power efficiency of fog nodes.

Despite these advancements, **challenges remain**, including:

- Scalability of AI models for large-scale IoT-FC environments.
- Security and trust management in AI-driven fog architectures.
- Reducing computational overhead in AI-based solutions to ensure real-time performance.

V. LITERATURE SURVEY

Anees Ur Rehman (2020) [13] Proposed the Dynamic Energy Efficient Resource Allocation (DEER) strategy for optimizing resource utilization in fog computing. It clusters tasks based on workload and dynamically adjusts power states, reducing energy consumption by 8.67% and computational costs by 16.77% compared to DRAM. Fatma M. Talaat (2020) [14] Developed the Load Balancing and Optimization Strategy (LBOS) using reinforcement learning and genetic algorithms for healthcare IoT-fog environments. It improved load balancing by 85.71%, enhancing resource utilization and reducing allocation costs.

D. Baburao et al. (2021) [15] Introduced Enhanced Dynamic Resource Allocation Method (EDRAM), leveraging Particle Swarm Optimization (PSO) to optimize fog resource distribution. It improved latency, bandwidth usage, and task scheduling, ensuring better Quality of Experience (QoE). Ahmad Raza Hameed (2021) [16] Proposed a cluster-enabled, capacity-based load balancing approach for vehicular fog computing. It dynamically clusters moving vehicles based on position, speed, and direction, reducing latency and improving energy efficiency. Ebrahim and A. Hafid (2023) [17] Implemented an ELECTRE multi-criteria decision analysis method for real-time workload distribution in fog computing. Achieved a 67% performance improvement over conventional methods by optimizing resource allocation across service replicas.

Sun et al.[18] Proposed a three-layer cooperative framework for Vehicular Fog Computing (VFC) using deep learning (DL) for traffic prediction and deep reinforcement learning (DRL) for resource allocation. Improved allocation success by 1.2×, but lacked details on algorithm implementation and baseline selection. Abdulazeez and Askar [19] Reviewed reinforcement learning (RL) and deep reinforcement learning (DRL) approaches for task offloading in fog computing. Categorized methods into value-based, policy-based, and hybrid techniques, analyzing performance metrics and evaluation tools. Provided a structured comparison but did not propose a new solution. Patil and Sharma [20] Developed a machine learning-based system for dynamic resource allocation and memory management in cloud and mobile computing. Improved device performance by automatically removing unused data using Python and TensorFlow. Required further model enhancements for efficiency.

Lakhan et al. [21] Introduced a Deep Q-Network-based Reinforcement Learning method (DQBRA) for resource allocation in Fog Computing (FC) networks using Software-Defined Networking (SDN). Achieved 30% cost reduction, but lacked security considerations and node validation. Shruthi et al. [22] Explored heuristic scheduling techniques for IoT applications in fog-cloud computing. Analyzed priority-based, greedy, and hybrid methods, highlighting efficiency and adaptability, but also noted challenges with dynamic environments and security. Kumar et al. [23] Proposed a deep autoencoder (DAE) model for workload prediction in Industrial IoT (IIoT) fog systems, combined with crow search algorithm (CSA) for optimal node selection. Outperformed existing models in cost, throughput, and response time, though scalability concerns remain. Mishra et al. [24] Applied Analytic Hierarchy Process (AHP) for resource allocation in fog-cloud hybrid systems, optimizing computational and network loads. Achieved task delay reduction and better system efficiency compared to existing strategies.

Tsokov and Kostadinov [25] Introduced a Mixed-Integer Linear Programming (MILP) model for dynamic container allocation in Cloud/Fog networks. Focused on reducing end-to-end network latency and improving QoS, but faced execution complexity issues. Peng et al. [26] Developed a resource allocation model for Fog Radio Access Networks (F-RANs) using convex optimization and game theory. Proposed Economical Energy Efficiency (E3) as a metric, improving throughput and energy efficiency despite network interference challenges. Yakubu and Murali [27] Proposed a meta-heuristic resource allocation method with Modified Harris-Hawks Optimization (MHHO) for IoT-fog-cloud computing. Achieved improvements in execution

S.no.	Reference	Algorithm/ AI Technology Used	Main Idea	Limitations	Advantages
1.	Anees Ur Rehman [13]	Dynamic Energy Efficient Resource Allocation (DEER)	Task grouping and resource allocation based on utilization to enhance load balancing and energy efficiency	Require additional computational overhead for real-time power state adjustments	-Reduces energy consumption by 8.67% -lowers computational costs by 16.77% - optimizes resource allocation
2.	Fatma M. Talaat [14]	Load Balancing and Optimization Strategy (LBOS)	Dynamic resource allocation across servers to balance load efficiently	Computational complexity due to hybrid algorithms, potential overhead in real-time processing	Achieves 85.71% load balancing, reduces response time and allocation costs, optimizes resources for healthcare IoT-fog systems
3.	D. Baburao et al. [15]	Particle Swarm Optimization (PSO)	Resource allocation and load balancing by particle swarm optimization (PSO)in fog environment ,optimizing the distribution of computational tasks among fog nodes in a fog computing environment to improve overall system performance	1-Parameter Sensitivity 2-high complexity 3-Optimization Challenges	1-Dynamic Adaptation 2-high resource utilization

4.	Ahmad Raza Hameed [16]	Dynamic clustering and predictive load balancing	Forms dynamic clusters based on vehicle mobility patterns and predicts departure times to optimize load distribution	Performance may vary in highly dynamic traffic environments, dependency on accurate vehicle mobility predictions	Balanced energy utilization, reduced delays , improved resource utilization, and leverages vehicular movement for fog computing environment
5.	Ebrahim and A. Hafid [17]	ELECTRE multi-criteria decision analysis	Distributes workloads based on various goals, such as processing and network loads, for optimized fog resource allocation	Complexity in implementing the ELECTRE methodology, potential computational overhead in real-time decision-making	Improves system performance by 67%, enhances QoS for delay-sensitive applications, optimizes workload distribution, and efficiently utilizes fog resources
6.	Prabhdeep Singh [28]	A Fog-Cluster Based Load-Balancing Technique	A multi layered framework with fog , user, and cloud, fog subsystems for cost-effective resource management	Requires careful tuning of the refresh period for optimal performance, potential overhead in maintaining cluster-based load balancing	Reduces energy consumption, minimizes VM migrations and host shutdowns, optimizes fog resource utilization, and balances dynamic loads efficiently
7.	F. Banaie et al. [29]	Stochastic heuristic	Load balancing strategy based on analytic hierarchy method for multiple gateways in a fog computing environment	This algorithm can not handle the situations where the network has different traffic classes of task request which are linked to the gateways based to their Quality od service requirements	Effectiveness of given solution in the quick and reliable acquisition of big data from the IoT domain, which is expected to drive IoT advancement.
8.	S. S. Karthik and A. Kavithamani [30]	Stochastic Metaheuristics	fog computing-based deep learning model for optimizing a micro grid connected Wireless Sensor Network (WSN) with a focus on load balancing network considering load balancing	1-high complexity 2-Data Requirements	1-Optimized Micro grid Performance 2-Adaptability
9.	S. P. Singh et al. [31]	Probabilistic\ static Fuzzy logic	Design and exploration of load balancers for fog computing using fuzzy logic. The focus is on leveraging fuzzy logic principles to create load-balancing mechanisms tailored for fog computing environments	1-Complexity 2-Interpretability 3-Optimization Challenges	1-Adaptability 2-Handling Uncertainty (Fuzzy logic enables robust fog load balancing amid unpredictability.

10.	R. Beraldi et al. (2020) [32]	Probabilistic\ static Random walk	Fog load balancing algorithm based on random walk (loosely) the algorithm may account for varying and possibly partially correlated states of the fog nodes or the system	1-Deterministic Challenges 2-Optimization Challenges 3-Resource Overheads	1-Adaptability 2-Scalability
11.	K. Cui et al. (2020) [33]	Graph theory Dijkstra algorithm	Load balancing of (USV)an unmanned surface vehicles in a fog system	1-Environmental Challenges 2-Limited Resources 3-Complexity	1- Efficient Task Distribution 2- Enhanced Resource Utilization
12.	Q. Fan and N. Ansari (2020) [34]	Gradient based Gradient algorithm	Workload balancing algorithm in IOT\ fog mode	1-IoT Device Heterogeneity 2-Communication Overheads. 3-Security and Privacy Concerns	1-Efficient Workload Distribution 2-Improved System Performance
13.	F. Alqahtani et al.(2021) [35]	LBSSA	Service scheduling requests and load balancing in fog\cloud environment	1-Algorithm Complexity 2-Dependency on Network Conditions 3-Resource Overheads	1-high reliability 2-Efficient Load Balancing
14.	A. Asghar et al. (2021) [36]	LBS	fog-based architecture and load balancing methodology for health monitoring systems	1-Algorithm Complexity 2-Data Security and Privacy 3-Scalability	1-Real-Time Health Monitoring 2-Resource Optimization
15.	N. Mazumdar et al.(2021) [37]	LFA	Trust-aware offloading and load balancing in a fog network	1-Algorithm Complexity 2-Dependency on Trust Models 3-Overhead in Trust Establishment	1-low latency: by making informed decisions about task offloading to reliable fog nodes. 2-Enhanced Reliability
16.	F. M. Talaat et al. (2020) [14]	Weighted RR,Q learning GA	Load balancing and optimization strategy (LBOS) using reinforcement learning in a fog computing environment. A load balancing strategy based on genetic algorithm and Q-learning in healthcare system	1-high complexity 2-Training Overhead 3-Dependency on Training Data	1-high resource utilization 2-low response time

17.	F. M. Talaat et al. (2019) [38]	FUZZY logic probabilistic neural networks	Effective load balancing strategy (ELBS) specifically designed for real-time applications in a fog computing environment the system based on neural network and fuzzy logic	1-Complexity 2-Training Overhead 3- Dependency on Network Characteristics	1- Real-Time Optimization 2- Handling Uncertainty 3- Probabilistic Modeling: Probabilistic neural networks may enable the algorithm to make decisions based on probabilistic information, enhancing adaptability to varying conditions.
18.	J. Yan et al. (2021) [39]	Greedy and coalitional game-based algorithm	task offloading algorithms designed for achieving novel load balancing in a homogeneous fog network	1-Algorithm Specificity 2- Real-world Implementation Challenge 3-Scalability	1-low energy 2-low delay
19.	Sun at al. [40]	DL, Deep Reinforcement Learning	Work focused on resource allocation issues in vehicular FC (VFC). This study suggested a three-tier VFC framework to address dynamic resource allocation among VFC	Lack of algorithm and baseline method details	Dynamic resource allocation, DL and DRL application
20.	Abdulazeez and Askar [41]	Reinforcement Learning, Deep RL	RL and deep reinforcement learning (DRL) for task offloading decisions in FC, divided into value-based, policy-based, and hybrid approaches.	Lack of Specific RL/DRL solution proposal	Taxonomy of offloading mechanism, performance evaluation
21.	Patil and Sharma [42]	Machine Learning (ML) with feature extraction and optimization algorithms.	Dynamic resource allocation and memory management leveraging ML techniques to optimize performance by removing unnecessary data blocks.	Need for further improvements in models and structures for enhanced performance.	- Enhanced device performance and energy management. - Automatic release of unused cache files. - Improved memory optimization and data processing time
22.	Lakhan et al. [43]	Deep Q-Network-based Reinforcement Learning (DQBRA).	Novel container-based architecture with DQBRA for resource allocation optimization in FC networks, addressing application mobility and dynamic network conditions.	- Lack of data security measures during application mobility. - Absence of node validation in the fog cloud network.	- 30% reduction in application costs compared to existing methods. - Efficient handling of dynamic network conditions and application mobility. - Improved energy consumption, latency, and resource utilization.

23.	Shruthi et al. [44]	Heuristic scheduling methods, including priority-based, greedy, metaheuristics, learning-based, hybrid, and nature-inspired techniques.	Investigation of heuristic scheduling techniques to improve energy efficiency and optimize resource utilization in fog-cloud environments for IoT applications.	- Challenges in dynamic environments. - Heterogeneity in fog-cloud systems. - Security concerns in task scheduling.	- Enhanced adaptability and efficiency in task scheduling. - Focus on energy efficiency and resource optimization. - Applicability to diverse IoT scenarios.
24.	Kumar et al. [45]	Deep Autoencoder (DAE) model for workload prediction and Crow Search Algorithm (CSA) for fog node selection.	Framework for workload prediction and resource allocation in fog-enabled Industrial IoT systems, aiming to minimize cost and delay.	- Potential challenges in real-world applicability and scalability. - Evaluation and implementation details not provided.	-Outperformed existing models in execution cost, request rejection ratio, throughput, and response time. - Autonomic nature and efficient dynamic workload management.
25.	Mishra et al. [46]	Analytic Hierarchy Process (AHP) for resource allocation, considering compute and network loads.	Resource allocation for delay-sensitive IoT applications in fog-cloud hybrid systems, using AHP to optimize resource distribution.	Not explicitly mentioned, but potential challenges could include scalability and adaptation to different network conditions.	- Reduced task delays. - Improved system efficiency. - Innovative application of AHP for resource allocation. - Relevant for IoT applications.
26.	Tsokov and Kostadinov [47]	Mixed-Integer Linear Programming (MILP) for optimizing microservice allocations in dynamic, mobile environments	Dynamic network-aware container allocation for Cloud/Fog environments with mobile nodes to minimize latency and optimize resource use.	- Potential issues with execution time and complexity, especially in environments with frequent node movements.	- Reduced end-to-end network latency. - Improved QoS and reduced network costs. - Practical applicability and adaptability to mobile infrastructures.
27.	Peng et al. [48]	Convex optimization, mixed-integer nonlinear programming, cooperative game theory, deep reinforcement learning.	Centralized and distributed resource allocation in F-RANs, utilizing joint optimization and deep reinforcement learning for dynamic environments.	- Complex communication mode selection. - Challenges in edge caching dynamics. - Deep reinforcement learning as a potential solution.	- Improved transmission efficiency and QoS. - New Economical Energy Efficiency (E3) metric. - Solutions for interference and limited spectrum in C-RANs.

28.	Yakubu and Murali [49]	Modified Harris-Hawks Optimization (MHHO) for resource allocation, with a layer fit algorithm for task distribution between fog and cloud layers.	Meta-heuristic resource allocation with load balancing in IoT-fog-cloud environments, aiming to optimize task distribution and resource assignment.	- Relied on synthetic data for evaluation. - Did not address task scheduling or application module placement on fog devices.	- Improved system performance in terms of make span time, execution cost, and energy consumption. - Efficient resource allocation with MHHO.
29.	Esmat and Lorenzo [50]	Q-learning (Q-EFNS), Deep Reinforcement Learning (DQ-EFNS), and Deep Dueling (Dueling DQ-EFNS) algorithms.	Dynamic edge/fog network slicing scheme (EFNS) for 6G networks, optimizing resource management with tenant leasing and idle subscriber terminals as fog nodes.	Evaluation based on simulations, with potential real-world application complexities not fully addressed.	- Significant improvements in performance (20% to 60%) over fixed network slicing scenarios. - Enhanced long-term revenue maximization for the Infrastructure Provider (InP).
30.	Javaheri et al. [51]	Clipped Double Deep Q-Learning (CDDQL) and Particle Swarm Optimization (PSO).	Autonomous resource allocation system in cloud computing using PSO for task prioritization and CDDQL for virtual machine resource allocation within the Fog layer.	- Challenges in managing the increasing number of IoT and FC devices. - Dependence on reliable internet connectivity for cloud services.	- Significant improvements in MakeSpan, response time, task completion rates, resource utilization, and energy consumption. - Autonomous operation with optimized task prioritization.
31.	Baburao et al. [52]	Particle Swarm Optimization (PSO) for efficient resource distribution among fog nodes.	Enhanced Dynamic Resource Allocation Method (EDRAM) to optimize load balancing, task waiting time, and network bandwidth usage, improving QoE.	- Scalability challenges. - Variability issues. - Potential need for enhanced fog node security in real-time applications.	- Improved latency, task waiting time, and load balancing. - Efficient energy utilization and reduced power consumption. - Seamless migration across fog systems using Docker.
32.	Liang et al. [53]	Reinforcement learning, framed as a semi-Markov decision process.	Dynamic resource allocation for the Internet of Vehicles (IoV), integrating resource reservation and secondary allocation mechanisms.		- Improved resource utilization and user QoE. - Adaptive learning capabilities enhancing system performance. - Optimization over traditional greedy algorithms.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VI June 2025- Available at www.ijraset.com

33.	Bhandari et al. [54]	Deep Neural Networks (DNN) for cache resource allocation prediction.	DNN-based framework for optimal cache resource allocation in Fog Radio Access Networks (F-RANs) to maximize delivered data and improve real-time performance.	- Challenges in managing large numbers of User Equipments (UEs) and Fog Access Points (F-APs). - High computational cost of iterative algorithms.	-Potential to reduce system complexity. - Improved real-time performance. - Closely approximates traditional iterative methods.
34.	Khan et al. [55]	Modified Particle Swarm Optimization (MPSO) for resource allocation and task scheduling.	Framework for load balancing and task scheduling to optimize performance in delay-sensitive IoMT applications, using MPSO to minimize latency overhead.	No specific limitations mentioned, but challenges may arise in real-world deployment with highly dynamic IoMT environments.	- Up to 80% improvement in resource utilization. - Significant reductions in execution time delay, cost, energy consumption, and network bandwidth usage.

VI. DISCUSSION AND COMPARATIVE ANALYSIS

The literature review indicates that hybrid AI-driven approaches offer superior performance in dynamic fog computing environments. Stochastic and heuristic methods provide efficient resource distribution but struggle with scalability and computational overhead. Below is a comparative analysis of different techniques:

Approach	Main Idea	Advantages	Limitations
ELECTRE multi-criteria decision analysis	Workload distribution based on compute and network loads	67% performance improvement, optimized real-time allocation	Complexity in implementation, computational overhead
Deep Q-Network-based RL	Optimized resource allocation for dynamic networks	Reduces application cost by 30%, improved energy efficiency	Lack of data security measures, no node validation
Particle Swarm Optimization (PSO)	Load balancing via efficient resource distribution	Reduces latency and bandwidth consumption, enhances QoE	Scalability challenges, continuous optimization overhead
Cluster-Based Load Balancing	Dynamic clustering for efficient task scheduling	Reduced energy consumption, better resource utilization	High dependency on accurate mobility predictions
Hybrid Reinforcement Learning & GA (LBOS)	AI-based dynamic resource allocation	85.71% load balancing efficiency, low response time	Computational complexity, overhead in real-time processing
Gradient Algorithm	Workload balancing in IoT/Fog mode	Improved system performance, efficient workload distribution	IoT device heterogeneity, communication overhead
Graph Theory - Dijkstra Algorithm	Load balancing of unmanned surface vehicles (USV) in a fog system	Efficient task distribution, enhanced resource utilization	Environmental challenges, complexity
Fuzzy Logic-Based Probabilistic Approach	Load balancer design for fog computing	Adaptability, handling uncertainty	Complexity, optimization challenges
Modified Particle Swarm Optimization (MPSO)	Resource allocation and task scheduling in IoMT	80% improvement in resource utilization, lower execution time and latency	Scalability challenges, real-world deployment difficulties
Deep Reinforcement Learning-Based Methods	Dynamic resource allocation with deep learning	High adaptability, energy-efficient, reduced latency	Algorithm complexity, high data dependency
Mixed-Integer Linear Programming (MILP)	Optimizing microservice allocations	Reduced network latency, improved QoS	Execution time and complexity, mobile infrastructure challenges

VII. CONCLUSION

Fog computing is involved into a large amount of IOT enabled system. This Provides load sharing on cloud, computing resources and enhanced system throughput. This paper investigates rapidly growing field of Fog computing and its impactful contribution to overcome the limitations of conventional cloud based architecture, especially for latency sensitive IoT applications. with the evolution of Fog, processing of data is possible at the network edge and thus improves responsiveness and efficiency. This study investigates core research areas such as resource management, load balancing, task scheduling, and security. Integration of AI based techniques in Fog Computing issues improves decision-making and resource optimization. This work presents an analysis of essential areas with respect of their effectiveness in enhancing essential performance matrix such as energy efficiency, latency, and quality of services (QoS). Comparison tables are provided to highlight scope, techniques, strength and weaknesses to deliver valuable insight for researchers.

REFERENCES

[1] Z. Ma, M. Xiao, Y. Xiao, Z. Pang, H. V. Poor, and B. Vucetic, ‘‘High-reliability and low-latency wireless communication for Internet of Things: Challenges, fundamentals, and enabling technologies,’’ IEEE Internet Things J., vol. 6, no. 5, pp. 7946–7970, Oct. 2019.
[2] K. Lone and S. A. Sofi, ‘‘A review on offloading in fog-based Internet of Things: Architecture, machine learning approaches, and open issues,’’ High-Confidence Comput., vol. 3, no. 2, Jun. 2023, Art. no. 100124.
[3] H.K.Apat,R.Nayak,andB.Sahoo,‘‘Acomprehensivereview onInternet of Things application placement in fog computing environment,’’ Internet Things, vol. 23, Oct. 2023, Art. no. 100866.
[4] G. K. Walia, M. Kumar, and S. S. Gill, ‘‘AI-empowered fog/edge resource management for IoT applications: A comprehensive review, research challenges, and future perspectives,’’ IEEE Commun. Surveys Tuts., vol. 26, no. 1, pp. 619–669, 1st Quart., 2024.
[5] M. Aqib, D. Kumar, and S. Tripathi, ‘‘Machine learning for fog computing: Review, opportunities and a fog application classifier and scheduler,’’ Wireless Pers. Commun., vol. 129, no. 2, pp. 853–880, Mar. 2023.
[6] S. Askar, ‘‘Deep learning and fog computing: A review,’’ Available SSRN, 2021.
[7] Z. Ma, M. Xiao, Y. Xiao, Z. Pang, H. V. Poor, and B. Vucetic, ‘‘High-reliability and low-latency wireless communication for Internet of Things: Challenges, fundamentals, and enabling technologies,’’ IEEE Internet Things J., vol. 6, no. 5, pp. 7946–7970, Oct. 2019
[8] K. H. Abdulkareem, M. A. Mohammed, S. S. Gunasekaran, M. N. Al-Mhiqani, A. A. Mutlag, S. A. Mostafa, N. S. Ali, and D. A. Ibrahim, ‘‘A review of fog computing and machine learning: Concepts, applications, challenges, and open issues,’’ IEEE Access, vol. 7, pp. 153123–153140, 2019.
[9] G. Kumar and A. Altalbe, ‘‘Artificial intelligence (AI) advancements for transportation security: In-depth insights into electric and aerial vehicle systems,’’ Environ., Develop. Sustainability, pp. 1–51, Apr. 2024.
[10] H. Djigal, J. Xu, L. Liu, and Y. Zhang, ‘‘Machine and deep leaming for resource allocation in multi-access edge computing: A survey,’’ IEEE Commun. Surveys Tuts., vol. 24, no. 4, pp. 2449–2494, 4th Quart., 2022.
[11] H. Tran-Dang, S. Bhardwaj, T. Rahim, A. Musaddiq, and D.-S. Kim, ‘‘Reinforcement learning based resource management for fog computing environment: Literature review, challenges, and open issues,’’ J. Commun. Netw., vol. 24, no. 1, pp. 83–98, Feb. 2022.
[12] V. Kashyap, R. Ahuja, and A. Kumar, ‘‘Nature inspired meta-heuristic algorithms based load balancing in fog computing environment,’’ in Proc. 7th Int. Conf. Parallel, Distrib. Grid Comput. (PDGC), Nov. 2022, pp. 396–401.
[13] A. U. Rehman et al., “Dynamic energy efficient resource allocation strategy for load balancing in fog environment,” IEEE Access, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3035181.
[14] F. M. Talaat, M. S. Saraya, A. I. Saleh, H. A. Ali, and S. H. Ali, “A load balancing and optimization strategy (LBOS) using reinforcement learning in fog computing environment,” J Ambient Intell Humaniz Comput, vol. 11, no. 11, 2020, doi: 10.1007/s12652-020-01768-8.
[15] D. Baburao, T. Pavankumar, and C. S. R. Prabhu, “Load balancing in the fog nodes using particle swarm optimization-based enhanced dynamic resource allocation method,” Applied Nanoscience (Switzerland), vol. 13, no. 2. 2023. doi: 10.1007/s13204-021-01970-w.
[16] A. R. Hameed, S. ul Islam, I. Ahmad, and K. Munir, “Energy- and performance-aware load-balancing in vehicular fog computing,” Sustainable Computing: Informatics and Systems, vol. 30, 2021, doi: 10.1016/j.suscom.2020.100454.
[17] M. Ebrahim and A. Hafid, “Resilience and load balancing in Fog networks: A Multi-Criteria Decision Analysis approach,” Microprocess Microsyst, vol. 101, 2023, doi: 10.1016/j.micpro.2023.104893.
[18] L.Sun, M.Liu,J. Guo, X.Yu,andS.Wang,‘‘Deep reinforcement learning empowered resource allocation in vehicular fog computing,’’ IEEE Trans. Veh. Technol., vol. 73, no. 5, pp. 7066–7076, May 2024.
[19] D. H. Abdulazeez and S. K. Askar, ‘‘Offloading mechanisms based on reinforcement learning and deep learning algorithms in the fog computing environment,’’ IEEE Access, vol. 11, pp. 12555–12586, 2023.
[20] D.R.Patil, ‘‘Dynamic resource allocation and memory management using machinelearning for cloud environments,’’ Int. J. Adv. Trends Comput. Sci. Eng., vol. 9, no. 4, pp. 5921–5927,. Aug. 2020
[21] A. Lakhan, M. A. Mohammed, O. I. Obaid, C. Chakraborty, K. H. Abdulkareem, and S. Kadry, ‘‘Efficient deep-reinforcement learning aware resource allocation in SDN-enabled fog paradigm,’’ Automated Softw. Eng., vol. 29, no. 1, pp. 1–25, May 2022.
[22] G. Shruthi, M. R. Mundada, B. J. Sowmya, and S. Supreeth, ‘‘Mayfly Taylor optimisation-based scheduling algorithm with deep reinforcement learning for dynamic scheduling in fog-cloud computing,’’ Appl. Comput. Intell. Soft Comput., vol. 2022, pp. 1–17, Aug. 2022.
[23] M. Kumar, A. Kishor, J. K. Samariya, and A. Y. Zomaya, ‘‘An autonomic workload prediction and resource allocation framework for fog-enabled industrial IoT,’’ IEEE Internet Things J., vol. 10, no. 11, pp. 9513–9522, Jun. 2023.
[24] S. Mishra, M. N. Sahoo, S. Bakshi, and J. J. P. C. Rodrigues, ‘‘Dynamic resource allocation in fog-cloud hybrid systems using multicriteria AHP techniques,’’ IEEE Internet Things J., vol. 7, no. 9, pp. 8993–9000, Sep. 2020.
[25] T. Tsokov and H. Kostadinov, ‘‘Dynamic network-aware container allocation in cloud/fog computing with mobile nodes,’’ Internet Things, vol. 26, Jul. 2024, Art. no. 101211.
[26] F. M. Talaat, S. H. Ali, A. I. Saleh, and H. A. Ali, “Effective Load Balancing Strategy (ELBS) for Real-Time Fog Computing Environment Using Fuzzy and Probabilistic Neural Networks,” Journal of Network and Systems Management, vol. 27, no. 4, 2019, doi: 10.1007/s10922-019-09490-3.
[27] I. Z. Yakubu and M. Murali, ‘‘An efficient meta-heuristic resource allocation with load balancing in IoT-Fog-cloud computing environment,’’ J. Ambient

Intell. Humanized Comput., vol. 14, no. 3, pp. 2981–2992, Mar. 2023.

[28] P. Singh et al., "A Fog-Cluster Based Load-Balancing Technique," Sustainability (Switzerland), vol. 14, no. 13, 2022, doi: 10.3390/su14137961.

[29] F. Banaie, M. H. Yaghmaee, S. A. Hosseini, and F. Tashtarian, "Load-Balancing Algorithm for Multiple Gateways in Fog-Based Internet of Things," IEEE Internet Things J, vol. 7, no. 8, 2020, doi: 10.1109/JIOT.2020.2982305.

[30] S. S. Karthik and A. Kavithamani, "Fog computing-based deep learning model for optimization of microgrid connected WSN with load balancing," Wireless Networks, vol. 27, no. 4, 2021, doi: 10.1007/s11276-021-02613-2.

[31] S. P. Singh, A. Sharma, and R. Kumar, "Design and exploration of load balancers for fog computing using fuzzy logic," Simul Model Pract Theory, vol. 101, 2020, doi: 10.1016/j.simpat.2019.102017.

[32] R. Beraldi, C. Canali, R. Lancellotti, and G. P. Mattia, "Randomized Load Balancing under Loosely Correlated State Information in Fog Computing," in MSWiM 2020 - Proceedings of the 23rd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, 2020. doi: 10.1145/3416010.3423244.

[33] K. Cui, W. Sun, B. Lin, and W. Sun, "Load balancing mechanisms of unmanned surface vehicle cluster based on marine vehicular fog computing," in Proceedings - 2020 16th International Conference on Mobility, Sensing and Networking, MSN 2020, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 797–802. doi: 10.1109/MSN50589.2020.00136.

[34] Q. Fan and N. Ansari, "Towards Workload Balancing in Fog Computing Empowered IoT," IEEE Trans Netw Sci Eng, vol. 7, no. 1, 2020, doi: 10.1109/TNSE.2018.2852762.

[35] F. Alqahtani, M. Amoon, and A. A. Nasr, "Reliable scheduling and load balancing for requests in cloud-fog computing," Peer Peer Netw Appl, vol. 14, no. 4, 2021, doi: 10.1007/s12083-021-01125-2.

[36] B. Alamri, M. A. Hossain, and M. S. Jamil Asghar, "Electric power network interconnection: A review on current status, future prospects and research direction," Electronics (Switzerland), vol. 10, no. 17, pp. 1–29, 2021, doi: 10.3390/electronics10172179.

[37] A. Lakhan, M. A. Mohammed, O. I. Obaid, C. Chakraborty, K. H. Abdulkareem, and S. Kadry, ''Efficient deep-reinforcement learning aware resource allocation in SDN-enabled fog paradigm,'' Automated Softw. Eng., vol. 29, no. 1, pp. 1–25, May 2022

[38] F. M. Talaat, S. H. Ali, A. I. Saleh, and H. A. Ali, "Effective Load Balancing Strategy (ELBS) for Real-Time Fog Computing Environment Using Fuzzy and Probabilistic Neural Networks," Journal of Network and Systems Management, vol. 27, no. 4, 2019, doi: 10.1007/s10922-019-09490-3.

[39] J. Yan, J. Wu, Y. Wu, L. Chen, and S. Liu, "Task Offloading Algorithms for Novel Load Balancing in Homogeneous Fog Network," in Proceedings of the 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2021, 2021. doi: 10.1109/CSCWD49262.2021.9437748.

[40] L.Sun, M.Liu,J. Guo, X.Yu,andS.Wang,''Deep reinforcement learning empowered resource allocation in vehicular fog computing,'' IEEE Trans. Veh. Technol., vol. 73, no. 5, pp. 7066–7076, May 2024.

[41] D. H. Abdulazeez and S. K. Askar, ''Offloading mechanisms based on reinforcement learning and deep learning algorithms in the fog computing environment,'' IEEE Access, vol. 11, pp. 12555–12586, 2023.

[42] D.R.Patil, ''Dynamic resource allocation and memory management using machinelearning for cloud environments,'' Int. J. Adv. Trends Comput. Sci. Eng., vol. 9, no. 4, pp. 5921–5927,. Aug. 2020

[43] A. Lakhan, M. A. Mohammed, O. I. Obaid, C. Chakraborty, K. H. Abdulkareem, and S. Kadry, ''Efficient deep-reinforcement learning aware resource allocation in SDN-enabled fog paradigm,'' Automated Softw. Eng., vol. 29, no. 1, pp. 1–25, May 2022.

[44] G. Shruthi, M. R. Mundada, B. J. Sowmya, and S. Supreeth, ''Mayfly Taylor optimisation-based scheduling algorithm with deep reinforcement learning for dynamic scheduling in fog-cloud computing,'' Appl. Comput. Intell. Soft Comput., vol. 2022, pp. 1–17, Aug. 2022.

[45] M. Kumar, A. Kishor, J. K. Samariya, and A. Y. Zomaya, ''An autonomic workload prediction and resource allocation framework for fog-enabled industrial IoT,'' IEEE Internet Things J., vol. 10, no. 11, pp. 9513–9522, Jun. 2023.

[46] S. Mishra, M. N. Sahoo, S. Bakshi, and J. J. P. C. Rodrigues, ''Dynamic resource allocation in fog-cloud hybrid systems using multicriteria AHP techniques,'' IEEE Internet Things J., vol. 7, no. 9, pp. 8993–9000, Sep. 2020

[47] T. Tsokov and H. Kostadinov, ''Dynamic network-aware container allocation in cloud/fog computing with mobile nodes,'' Internet Things, vol. 26, Jul. 2024, Art. no. 101211.

[48] M. Peng, Z. Zhao, Y. Sun, M. Peng, Z. Zhao, and Y. Sun, ''Dynamic resource allocation in fog radio access networks,'' in Fog Radio Access Networks (F-RAN), 2020, pp. 105–131.

[49] I. Z. Yakubu and M. Murali, ''An efficient meta-heuristic resource allocation with load balancing in IoT-Fog-cloud computing environment,'' J. Ambient Intell. Humanized Comput., vol. 14, no. 3, pp. 2981–2992, Mar. 2023.

[50] H.H.Esmatand B.Lorenzo, ''Deep reinforcement learning based dynamic edge/fog network slicing,'' in Proc. GLOBECOM IEEE Global Commun. Conf., Dec. 2020, pp. 1–6.

[51] S. D. A. Javaheri, R. Ghaemi, and H. M. Naeen, ''An autonomous architecture based on reinforcement deep neural network for resource allocation in cloud computing,'' Computing, vol. 106, no. 2, pp. 371–403, Feb. 2024.

[52] D. Baburao, T. Pavankumar, and C. S. R. Prabhu, ''Load balancing in the fog nodes using particle swarm optimization-based enhanced dynamic resource allocation method,'' Appl. Nanosci., vol. 13, no. 2, pp. 1045–1054, Feb. 2023.

[53] H. Liang, X. Zhang, X. Hong, Z. Zhang, M. Li, G. Hu, and F. Hou, ''Reinforcement learning enabled dynamic resource allocation in the Internet of Vehicles,'' IEEE Trans. Ind. Informat., vol. 17, no. 7, pp. 4957–4967, Jul. 2021.

[54] S. Bhandari, H. Kim, N. Ranjan, H. P. Zhao, and P. Khan, ''Optimal cache resource allocation based on deep neural networks for fog radio access networks,'' J. Internet Technol, vol. 21, pp. 967–975, 2020.

[55] S. Khan, I. A. Shah, M. F. Nadeem, S. Jan, T. Whangbo, and S. Ahmad, ''Optimal resource allocation and task scheduling in fog computing for Internet of Medical Things applications,'' Tech. Rep., 2023.