International Journal of Advancement in Electronics and Computer Engineering (IJAECE)
Volume 14, Issue 1, January. 2025, pp. 119-123 ISSN 2278 -1412

Exploring Hybrid AI Strategies for Intelligent Resource Management in Fog Computing: A Review

Priyanka Singh[1], Nitya Khare[2] , Swati Khanve[3]
[1] Research Scholar, priyankasinghcs2014@gmail.com, CSE SIRTE, Bhopal, India
[2]HOD, CSE SIRTE, Bhopal, India
[3]Assis. Prof., swatikhanve55.sk@gmail.com, CSE SIRTE, Bhopal, India

***Abstract** – Fog computing has emerged as a pivotal paradigm in bridging the gap between cloud and edge computing, enabling efficient resource management and low-latency processing for IoT-driven applications. However, the dynamic nature of fog environments presents significant challenges in resource allocation, workload balancing, and energy efficiency. In response, hybrid Artificial Intelligence (AI)-based frameworks have gained prominence, integrating machine learning, deep learning, and heuristic optimization techniques to enhance decision-making and adaptability in fog computing ecosystems.*

This review provides an in-depth analysis of existing AI-driven strategies for optimized resource management in fog computing. It explores the synergy between rule-based heuristics, reinforcement learning, federated learning, and swarm intelligence in addressing computational and network constraints. Additionally, the paper highlights key research gaps, security considerations, and the potential of emerging AI-driven methodologies, such as neuromorphic computing and explainable AI, in transforming fog infrastructure.

Keywords: Fog Computing, Hybrid AI, Resource Management, Machine Learning, Deep Learning, Heuristic Optimization, Edge Intelligence, Workload Balancing, Energy Efficiency

I. INTRODUCTION

The exponential growth in the use of mobile devices and computers has significantly transformed how individuals and organizations interact with technology. This evolution has led to a surge in data generation, with electronic devices and sensors producing massive volumes of information daily. Managing such enormous datasets presents significant challenges for organizations, necessitating the adoption of advanced data management solutions. The transition to cloud computing has been pivotal in addressing these challenges, offering benefits such as scalability, availability, and cost-effectiveness through pay-as-you-use models. Cloud computing provides a wide array of services, including platforms, software, and infrastructure, making it a cornerstone of modern digital transformation. However, cloud-based systems are not always feasible for handling real-time data processing due to inherent latency and the inefficiencies of transferring large IoT datasets to centralized servers.

Fog computing (FC) has emerged as a promising paradigm. Unlike traditional cloud computing, fog computing brings data storage and processing capabilities closer to the data source, typically at the network edge. This approach minimizes latency, enhances efficiency, and reduces the volume of data transmitted to centralized servers for analysis and storage. In addition to performance improvements, fog computing offers benefits in terms of security and compliance, making it a preferred choice for applications requiring real-time processing and low-latency operations. By leveraging the computing power of devices near end users, fog computing ensures that critical tasks are executed efficiently and promptly.

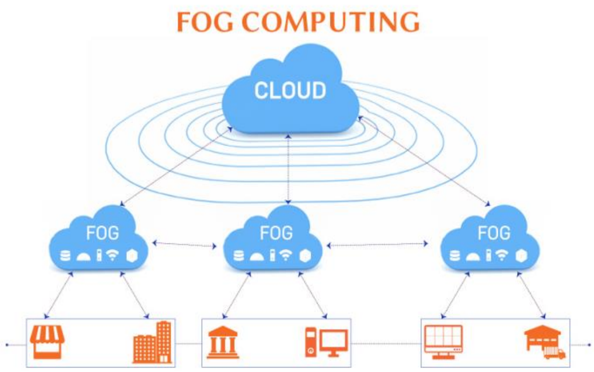

Fig 1.1 Shows Fog computing

Artificial intelligence (AI) and machine learning (ML) play a pivotal role in enhancing the capabilities of fog computing. AI and ML algorithms enable intelligent decision-making by analyzing vast amounts of data generated by IoT devices. These technologies have been instrumental in overcoming networking challenges such as routing, security, traffic engineering, and resource allocation. Furthermore, they enable the development of smart and autonomous environments where systems can self-manage and optimize operations. For IoT applications, ML is indispensable for functional tasks such as monitoring, clustering, and preprocessing data.

Without ML, IoT systems would struggle to meet diverse quality-of-service (QoS) demands, particularly given the resource constraints of IoT devices. Despite its potential, fog computing faces challenges due to the dynamic, complex, and heterogeneous nature of fog nodes. These nodes often have limited memory, low communication bandwidth, and inadequate processing power, making efficient management difficult. However, advancements in computing technologies, such as GPU acceleration, cloud integration, and enhanced hardware capabilities, have significantly bolstered the effectiveness of fog computing. ML frameworks like Weka and Scikit-learn are now being integrated into high-level fog nodes, enabling the execution of complex AI applications.

The synergy between fog computing and ML has paved the way for intelligent fog applications capable of delivering deep analytics and actionable insights. By harnessing ML algorithms, fog nodes can optimize tasks such as clustering, routing, duty-cycle management, and data aggregation. This integration not only enhances system efficiency but also ensures the seamless operation of IoT ecosystems. As research in this field continues to evolve, fog computing combined with ML holds immense potential to revolutionize data processing and decision-making across industries.

II. LITERATURE SURVEY

Fady E. F. Samann et al (2021) analysis revealed that different studies often used varying performance metrics, except for those focusing on security issues, where metrics were more consistent. Despite the progress in integrating ML with FC, simulations of proposed ML models often fall short of addressing the heterogeneous nature of the FC paradigm. This highlights the need for more robust and practical implementations to fully realize the potential of ML in enhancing FC systems. To address these challenges, machine learning (ML) has garnered significant attention as a means to enhance FC's efficiency and resolve its limitations. The integration of ML in FC applications has become increasingly popular, offering solutions for resource management, security, reduced latency, and optimized power consumption. Researchers have also explored intelligent FC to tackle challenges in fields such as Industry 4.0, bioinformatics, blockchain, and vehicular communication systems. Given the critical role ML plays in the FC paradigm, this work focuses on recent research leveraging ML within FC environments. Background information on ML and FC is provided to establish a foundational understanding. The reviewed studies were categorized into three groups based on the primary goals of ML implementation. These studies were comprehensively analyzed and compared, with findings summarized in tabular formats.

Shreshth Tuli et al (2023) research has primarily focused on leveraging AI to enhance existing systems across diverse domains, including resource provisioning, application deployment, task placement, and service management. This review explores the development of data-driven, AI-augmented technologies and their profound influence on computing systems. It examines cutting-edge techniques and provides critical insights into how AI methods are being applied to manage resources in Edge, Fog, and Cloud environments. Additionally, it highlights how AI can transform traditional applications to deliver superior Quality of Service (QoS) through seamless resource integration. The survey also delves into the latest advancements and emerging trends, particularly in optimizing AI models deployed within or tailored for computing systems. It outlines a strategic roadmap for future research, emphasizing key areas such as resource management for QoS enhancement and improving service reliability. Lastly, it introduces innovative concepts and envisions this work as a foundational reference for future studies on AI-driven computing paradigms, offering a forward-looking perspective on the evolution of these transformative technologies.

Chinmoy Bharadwaj et al (2024) Traditional centralized cloud computing is having considerable issues due to the rapid proliferation of mobile internet and Internet of Things applications. These issues include data transfer delays, low efficiency in exploiting available spectrum, and connectivity that is inflexible based on machine type. Driven by the need to overcome these challenges, a new technology is prompting a change in the role of centralized cloud computing to edge devices within networks. Driven by the need to overcome these challenges, a revolutionary technology is prompting a change in the role of centralized cloud computing to edge devices within the networks. Edge computing systems have emerged from a variety of sources, with the goal of reducing latency, improving SE (security and efficiency), and enabling wide connectivity among devices. Given the broad growth and advancement of cloud technologies, the existing cloud computing model is insufficient to address the critical needs for providing services to applications such as smart grid, healthcare systems, and augmented reality. These specifications include low latency, mobility support, and context awareness. Fog and edge computing have developed as technologies that meet these requirements while providing great performance and dependability. Machine learning entails detecting many characteristics such as automated procedures, increased decision support, and diagnostic capabilities within operational settings. This study exposes previous research that focuses on diverse applications of machine learning, identifying potential hazards, and developing solutions for the edge paradigm as a whole. It also underlines the outstanding issues and challenges that may remain in the coming decades.

Amira Bourechak et al (2022) paper seeks to explore the intersection of AI and edge computing across various application domains, leveraging existing research and uncovering new perspectives. The convergence of edge computing and AI plays a crucial role in improving user experiences in emergency scenarios, such as in the Internet of Vehicles, where delays or inaccuracies can lead to accidents and significant harm. These factors are often used to evaluate the success of edge-based applications. Provide a comprehensive analysis of the current state of AI in edge-based applications, focusing on eight key application areas: smart agriculture, smart environment, smart grid, smart healthcare, smart industry, smart education, smart transportation, and security and privacy. We then present a qualitative comparison that highlights the core objectives of this convergence, the roles of AI at the network edge, and the key enabling technologies for edge analytics. Furthermore, we discuss the open challenges, future research directions, and perspectives in this field. Finally, the paper concludes by summarizing the key findings and insights.

Ali Shakarami et al (2022) paper explores open issues and future challenges in resource provisioning that have been insufficiently addressed. These include optimizing resource performance, managing resource location constraints, handling uncertainties, ensuring resource elasticity, and addressing resource migration. By highlighting these challenges, the survey aims to provide a foundation for future research and innovation in resource provisioning for fog and edge computing environments. Despite the importance of resource provisioning in these computing environments, there has yet to be a systematic, comprehensive, and detailed survey conducted on the topic. This paper seeks to bridge that gap by providing a review of resource provisioning approaches within fog and edge computing. It introduces a standard classification to examine existing methods on this critical topic while identifying open issues that remain to be addressed. The classification organizes resource provisioning approaches into five primary categories: framework-based methods, heuristic and meta-heuristic approaches, model-based techniques, machine learning-driven solutions, and game-theoretic mechanisms. These approaches are then analyzed and compared based on essential features, such as performance metrics, case studies, employed methodologies, and evaluation tools. This comparative analysis offers valuable insights into the advantages and limitations of each method.

Hongquan Gui et al (2022) proposed a system significantly reduces the volume of transferred data. Using the data-based model, the volume is reduced by 11/16, while the designed system further reduces the transferred thermal data volume to 1/10. By incorporating a precision threshold, the system enhances predictive accuracy by 8.31% compared to a system without this threshold. With the implementation of the MEFCS, the accuracy level of tooth profile deviation improved from ISO level 5 to ISO level 3. Additionally, the total execution times for different architectures were measured, showing that the mist-cloud structure took 206 seconds, the mist-edge-cloud structure 200 seconds, the mist-fog-cloud structure 186 seconds, and the mist-edge-fog-cloud structure 167 seconds. A finite element model was developed to validate the effectiveness of the bidirectional long short-term memory (Bi-LSTM) network in this context. To enhance optimization, a cosine and sine gray wolf optimization (CSGWO) algorithm was designed to optimize the batch size. Building on this, the CSGWO-Bi-LSTM network error model was proposed. The predictive accuracy achieved with this model showed progressive improvements across various methods: 90.80% for the multiple linear regression model, 94.57% for the recurrent neural network, 95.77% for the LSTM network, 96.79% for the Bi-LSTM network, 97.51% for the CSGWO1-Bi-LSTM network, 98.45% for the CSGWO2-Bi-LSTM network, and 98.92% for the CSGWO3-Bi-LSTM network.

Omid Bushehrian et al (2025) paper proposes a novel two-phase method for the adaptive creation and deployment of IoT machine learning tasks within a heterogeneous multi-layer fog computing architecture. In the first phase, a Deep Reinforcement Learning (DRL) approach is used to determine the optimal number of tasks and their respective sensor coverage. In the second phase, these tasks are deployed across the fog computing layers using a greedy deployment strategy. The task creation and deployment process is formulated as a three-objective optimization problem, aiming to minimize deployment latency, reduce deployment costs, and lower the evaluation loss of the machine learning job when trained in a federated manner across edge, fog, and cloud nodes. To solve this problem efficiently and adaptively, the study employs a Deep Deterministic Policy Gradient (DDPG) algorithm. The experimental results, based on the deployment of multiple IoT machine learning jobs with varying profiles on a heterogeneous fog computing testbed, demonstrate the efficacy of the proposed two-phase DRL-based method. The approach outperformed traditional Edge-IoT and Cloud-IoT baseline methods, achieving up to a 32% improvement in the total deployment score. These findings underscore the potential of the proposed method in optimizing the creation and deployment of IoT machine learning tasks in complex computing environments.

Gonçalo Carvalho et al (2020) study conducts a comprehensive literature review on computation offloading in EC systems, comparing approaches that incorporate AI techniques with those that do not. It focuses on analyzing various AI techniques, particularly ML-based methods, that have demonstrated significant promise in overcoming the limitations of traditional approaches for coordinating computation offloading. ML algorithms are categorized into distinct classes to facilitate a detailed analysis. The study places special emphasis on the use of AI in the context of Vehicular Edge Computing Networks, a technology that has gained significant relevance in recent years. This technology enables numerous solutions for computation and data offloading, highlighting the practical applications of AI in EC systems.The study explores the key advantages and limitations of computation offloading, with and without the application of AI techniques. By offering insights into these aspects, the work provides a deeper understanding of the role of AI in enhancing the effectiveness of computation offloading in EC environments.

Belen Bermejo et al (2023) article seeks to explore whether the integration of artificial intelligence can effectively enhance the sustainability of cloud, fog, edge, and IoT ecosystems. To achieve this, a systematic literature review is conducted to examine the relationship between AI techniques and sustainability in these ecosystems. The paper also presents a classification framework for the analyzed studies, categorizing them based on various aspects of these ecosystems, their environmental impact, and the role of AI in improving sustainability. By providing a comprehensive analysis, the study aims to contribute valuable insights into the potential of AI to mitigate the environmental challenges associated with these rapidly evolving technologies. The rapid growth in the use of services across cloud, fog, edge, and IoT ecosystems has become increasingly significant in recent years. While these ecosystems provide numerous technological advancements, they also pose challenges to environmental sustainability due to their substantial energy consumption, which leads to increased CO2 emissions. Furthermore, the COVID-19 pandemic has significantly accelerated the adoption of these ecosystems, amplifying their environmental impact. To address these challenges, it is essential to implement policies and techniques aimed at enhancing the sustainability of these ecosystems. Among the promising solutions are approaches that leverage artificial intelligence (AI). However, the application of AI-driven methods also raises concerns, as their processing requirements can lead to additional resource consumption.

III. PROBLEM STATEMENT

In the modern digital landscape, the exponential growth of data generated by IoT devices, sensors, and connected systems has posed significant challenges to traditional cloud computing architectures. These centralized systems struggle to meet the demands of low latency, high bandwidth, and real-time processing required by applications such as autonomous vehicles, healthcare monitoring, smart cities, and industrial automation. Additionally, cloud-based solutions often suffer from issues such as network congestion, high energy consumption, and increased operational costs, which hinder their ability to deliver reliable and efficient services.

Edge and Fog Computing have emerged as transformative paradigms that bring computational resources closer to data sources, reducing latency and improving system responsiveness. However, these decentralized architectures introduce complexities in task offloading, resource allocation, and workload balancing. Deciding when, where, and how to offload computational tasks while ensuring optimal performance, energy efficiency, and reliability remains a non-trivial challenge.

Artificial Intelligence (AI) and Machine Learning (ML) offer promising solutions to these challenges by enabling intelligent decision-making and predictive capabilities. AI and ML can optimize resource utilization, enhance task scheduling, and enable adaptive systems that learn and evolve over time. Despite these advantages, integrating AI/ML with Edge/Fog Computing presents its own set of challenges, including limited computational power of edge devices, scalability issues, and the need for robust models that can operate under dynamic and resource-constrained environments.

This research aims to explore the integration of AI and ML into Edge/Fog Computing systems, addressing critical challenges such as task offloading, resource optimization, and real-time analytics. By leveraging advanced AI/ML techniques, this study seeks to develop innovative solutions that enhance the performance, scalability, and adaptability of Edge/Fog Computing systems, ultimately unlocking their full potential across various domains and applications..

IV. CONCLUSION

Game-Changer for Data Management: The integration of edge and fog computing with Artificial Intelligence (AI) and Machine Learning (ML) is revolutionizing how data is managed in an increasingly interconnected world. As the volume of data generated by IoT devices, sensors, and smart systems continues to grow, traditional centralized cloud models face challenges in efficiently processing and storing this data. Edge and fog computing address these challenges by bringing computation closer to data sources, significantly reducing the need for constant data transfer to centralized servers. When coupled with AI and ML, these paradigms transform raw data into actionable insights, enabling smarter and more efficient systems. This integration allows for enhanced data handling capabilities, improved resource allocation, and the ability to process data in near real-time, making it a true game-changer in modern data management practices.

Real-Time Data Processing: One of the most significant advantages of combining edge and fog computing with AI and ML is the ability to process data in real time. Traditional cloud-based systems often suffer from latency issues due to the time it takes for data to travel to and from centralized servers. Edge and fog computing eliminate this bottleneck by processing data closer to its source, ensuring faster response times. AI and ML further enhance this capability by enabling predictive analytics and real-time decision-making. For instance, in autonomous vehicles, real-time data processing is critical for safety and efficiency, as vehicles need to analyze road conditions, traffic patterns, and potential hazards instantly. Similarly, in healthcare, real-time monitoring of patient vitals can lead to quicker diagnoses and timely interventions, ultimately saving lives.

REFERENCES

[1] Fady E. F. Samann , Adnan Mohsin Abdulazeez Duhok Polytechnic University, Duhok, Iraq, "Fog Computing Based on Machine Learning: A Review" Vol. 15, No. 12, 2021, IJIM, DOI: https://doi.org/10.3991/ijim.v15i12.21313

[2] Shreshth Tuli, Fatemeh Mirhakimi, Samodha Pallewatta, Syed Zawad, Giuliano Casale, Bahman Javadi, Feng Yan, Rajkumar Buyya, Nicholas R. Jennings, "AI augmented Edge and Fog computing: Trends and challenges", Volume 216, July 2023, 103648, Journal of Network and Computer Applications, DOI: https://doi.org/10.1016/j.jnca.2023.103648

[3] Chinmoy Bharadwaj; Atowar Ul Islam; Sippee Bharadwaj; Priyanka Sarma, "A Review and Threat Analysis of Machine Learning in Relation to Fog and Edge Computing", Electronic ISBN:979-8-3503-7994-5, 2024, IEEE, DOI: https://doi.org/10.1109/ICESC60852.2024.10689816

[4] Amira Bourechak, Ouarda Zedadra, Mohamed Nadjib Kouahla, Antonio Guerrieri , Hamid Seridi and Giancarlo Fortino, "At the Confluence of Artificial Intelligence and Edge Computing in IoT-Based Applications: A Review and New Perspectives", Volume 23 Issue 3, 2022, MDPI, DOI: https://www.mdpi.com/1424-8220/23/3/1639#

[5]Ali Shakarami, Hamid Shakarami, Mostafa Ghobaei, Arani, Elaheh Nikougoftar, Mohammad Faraji-Mehmandar, "Resource provisioning in edge/fog computing: A Comprehensive and Systematic Review", Volume 122, January 2022, Journal of Systems Architecture, DOI: https://doi.org/10.1016/j.sysarc.2021.102362

[6] Hongquan Gui, Jialan Liu, Chi Ma, Mengyuan Li, Shilong Wang, "New mist-edge-fog-cloud system architecture for thermal error prediction and control enabled by deep-learning", Volume 109, March 2022, Engineering Applications of Artificial Intelligence, DOI: https://doi.org/10.1016/j.engappai.2021.104626

[7] Omid Bushehrian & Amir Moazeni, "Deep reinforcement learning-based optimal deployment of IoT machine learning jobs in fog computing architecture", Volume 107, article number 15, 2025, Springer

[8] Gonçalo Carvalho, Bruno Cabral, Vasco Pereira, Jorge Bernardino, "Computation offloading in Edge Computing environments using Artificial Intelligence techniques", Volume 95, October 2020, Engineering Applications of Artificial Intelligence, DOI: https://doi.org/10.1016/j.engappai.2020.103840

[9] Belen Bermejo, Carlos Juiz, "Improving cloud/edge sustainability through artificial intelligence: A systematic review", Volume 176, June 2023, Journal of Parallel and Distributed Computing, DOI: https://doi.org/10.1016/j.jpdc.2023.02.006

[10] Swarnava Dey; Arijit Mukherjee, "Implementing Deep Learning and Inferencing on Fog and Edge Computing Systems", Electronic ISBN:978-1-5386-3227-7, 2018, IEEE, DOI: https://doi.org/10.1109/PERCOMW.2018.8480168