



ARTICLE

ANNDRA-IoT: A Deep Learning Approach for Optimal Resource Allocation in Internet of Things Environments

Abdullah M. Alqahtani^{1,*}, Kamran Ahmad Awan², Abdulaziz Almaleh³ and Osama Aletri⁴

¹Department of Electrical and Electronic Engineering, College of Engineering and Computer Science, Jazan University, Jazan, 45142, Saudi Arabia

²Department of Information Technology, The University of Haripur, Haripur, 22620, Pakistan

³College of Computer Science, King Khalid University, Abha, 62529, Saudi Arabia

⁴Department of Computing, College of Engineering and Computing, Umm Al-Qura University, Makkah, 21955, Saudi Arabia

*Corresponding Author: Abdullah M. Alqahtani. Email: amqahtani@jazanu.edu.sa

Received: 25 November 2024; Accepted: 30 January 2025; Published: 03 March 2025

ABSTRACT: Efficient resource management within Internet of Things (IoT) environments remains a pressing challenge due to the increasing number of devices and their diverse functionalities. This study introduces a neural network-based model that uses Long-Short-Term Memory (LSTM) to optimize resource allocation under dynamically changing conditions. Designed to monitor the workload on individual IoT nodes, the model incorporates long-term data dependencies, enabling adaptive resource distribution in real time. The training process utilizes Min-Max normalization and grid search for hyperparameter tuning, ensuring high resource utilization and consistent performance. The simulation results demonstrate the effectiveness of the proposed method, outperforming the state-of-the-art approaches, including Dynamic and Efficient Enhanced Load-Balancing (DEELB), Optimized Scheduling and Collaborative Active Resource-management (OSCAR), Convolutional Neural Network with Monarch Butterfly Optimization (CNN-MBO), and Autonomic Workload Prediction and Resource Allocation for Fog (AWPR-FOG). For example, in scenarios with low system utilization, the model achieved a resource utilization efficiency of 95% while maintaining a latency of just 15 ms, significantly exceeding the performance of comparative methods.

KEYWORDS: Internet of things; resource optimization; deep learning; optimal resource allocation; neural network; efficiency

1 Introduction

The Internet of Things (IoT) has become a significant technological paradigm, redefining the interaction and integration of devices within the digital ecosystem. [1–4]. In IoT systems, the main focus is to achieve seamless integration in several dominions with the help of an elaborate network consisting of connected sensors [5], actuators and intelligent devices [6]. This connected architecture can allow a wide range of applications [7], to be automated, allowing data-driven decision making to be realized; these include addressing various critical quality attributes. However, the increasing scale and complexity of IoT ecosystems implies the need for robust mechanisms that can ensure optimal performance in the presence of challenges related to computational resource constraints [8–10]. The resolution of these challenges is very significant for the widespread adoption of IoT in practical applications in the real world [11–14]. Resource allocation is one of the major challenges to the implementation of IoT systems [15]. Resource allocation needs to be efficiently distributed within the growing device network. Conventional static models used for resource allocation do



not adapt in dynamic IoT environments. This lack leads to a deficient utilization of resources [16], increased latency, and high energy consumption [17,18].

Resource allocation remains one of the critical challenges in the deployment of IoT systems [19]. A challenging issue is how to distribute limited resources effectively among the increasing number of interconnected devices [20]. The traditional static models of allocation, though at the foundation, have some inherent limitations. They lack the flexibility needed to adapt to the ever-changing resource demands of IoT ecosystems [21,22]. These models rely on a predefined resource distribution, assuming certain fixed and static requirements. This rigidity significantly reduces the efficiency of IoT systems, whose resource requirements may change over time [23]. Static allocation methods do not fill the gaps in resources dynamically as they appear; this causes unnecessary delays, increased energy consumption, and a lack of efficiency that undermines the sustainability of IoT environments [24–26]. Overcoming these limitations will definitely help IoT systems perform at their best under dynamic real-world conditions.

Conventional static allocation models, which are designed based on fixed resource requirement assumptions, are not able to cope with the dynamically changing requirements typical of the IoT ecosystem. These models consider rigid distributions of resources, thus not fitting in well for situations where resource needs vary over time. This limited adaptability often introduces a number of inefficiencies, which manifest themselves through increased latency or unnecessary/extra energy consumptions, which impact the operational efficiency and sustainability of IoT infrastructures accordingly [27–29]. These inefficiencies limit Adaptive Neural Network-based Dynamic Resource Allocation (ANNDRRA-IoT) approaches, necessitating long- and short-term memory (LSTM) networks with both long- and short-term memories for a proposal. It sets the performance optimally by using real-time analytical resources, hence dynamically adjusting the resources according to their ever-changing conditions via IoT systems. This kind of adaptability can be achieved very well in diverse IoT ecological surroundings, which was beyond the efficacy of previous ones. The proposed ANNDRRA-IoT has mainly performed the following:

- LSTM-Based Real-Time Dynamic Resource Allocation in the IoT. The model outsmarts traditional static allocation methods by learning to adapt to unique patterns in data flows in the IoT.
- Optimized distribution of resources by predicting and adapting to changes in IoT networks. This would result in reductions in latency and energy consumption.
- The effectiveness in scalability and adaptability to manage resources in the various IoT scenarios. This flexibility expands the applicability of ANNDRRA-IoT in many applications.

The following sections are arranged as follows. [Section 2](#) provides a comprehensive review of the literature. [Section 3](#) presents the proposed ANNDRRA-IoT approach. [Section 4](#)-Experimental Simulation & Results. [Section 5](#) presents a detailed discussion of the results. Finally, [Section 6](#) serves as the conclusion of the article.

2 Related Work

The IoT has become a transformative technology, significantly altering the way computing and communication systems function. It has introduced new paradigms of connectivity and interaction across a wide range of devices and systems. However, efficiently managing resources in IoT environments remains a significant challenge. This section analyzes recent advances and methodologies in the management of IoT resources. [Table 1](#) summarizes these approaches, highlights their strengths, and identifies the gaps that the proposed ANNDRRA-IoT model aims to address.

Table 1: Comparative review of advances and challenges in IoT resource management techniques

Ref.	Advances	Challenges
[3]	Developed a dynamic optimization framework for balancing loads in IoT systems.	Compatibility with diverse IoT environments remains an issue.
[18]	Introduced a refined strategy for task scheduling with preemptive actions in fog layers of IoT.	Struggles in scenarios with highly fluctuating IoT workloads.
[30]	Proposed an adaptive task allocation model using deep learning for cloud environments.	Did not fully explore potential in edge computing scenarios.
[31]	Formulated a predictive resource allocation framework for the industrial IoT sector.	Did not fully delve into unique industrial IoT challenges.
[32]	Reviewed distributed AI techniques for enhancing resource efficiency in IoT.	Absence of detailed case studies and practical applications.
[33]	Suggested a model for efficient energy management in IoT's sensor networks.	Discussion on model's adaptability to different network types was limited.
[34]	Investigated strategies for improving communication efficiency at the IoT edge.	Lack of consideration for integration with existing legacy systems.
[35]	Examined multi-resource allocation in IoT with an emphasis on fairness and efficiency.	Did not evaluate the impact on overall system performance.
[36]	Applied an imperialist competitive algorithm for service deployment in fog computing.	Performance issues under rapidly changing network scenarios.
[37]	Addressed energy-efficient federated learning approaches for edge computing in IoT.	Consideration for broader network architecture was missing.
[38]	Offered a hybrid approach for authentication in heterogeneous IoT settings.	Challenges with scalability and real-time operations in diverse environments were overlooked.
[39]	Explored distributed service placement strategies in fog environments via optimization.	Comprehensive computational overhead analysis was lacking.
[40]	Showcased a strategy for energy-efficient offloading in mobile edge computing for IoT.	Depth of adaptability to changing network conditions was not thoroughly investigated.

2.1 Existing State-of-the-Art Approaches

In [3], the authors proposed a dynamic load balancing mechanism to improve the efficiency and adaptability of the IoT environment. Their research work highlighted how dynamic resource management approaches can provide a solution to fluctuating workloads in IoT environments. Similarly, study [18] proposed an optimized task-scheduling mechanism specifically for fog-assisted IoT environments. It develops this mechanism to introduce preemptive features in scheduling architectures. This concept has been closely aligned with our efforts to improve responsiveness in IoT systems. The study in [30] presented an overview of the secure and adaptive deep learning cloud task scheduling. In their approach, by marrying security with efficiency, there is a developed focus on managing computational tasks efficiently, hence laying down

the principles that could echo the ANNDRA-IoT framework objectives. Meanwhile, in [31], the authors have proposed an autonomic framework for workload prediction and resource allocation in industrial IoT systems. Predictive analysis and autonomic allocation strategies in their work share common goals with ANNDRA-IoT and point out the necessity for intelligent resource management.

The authors in [32] conducted a comprehensive review of resource-efficient distributed AI methods for IoT applications. Their review highlighted the increasing demand for AI-driven solutions, laying the foundations for integrating AI to optimize IoT resource allocation—a central theme in our research. On the other hand, study [33] developed an energy management model for wireless sensor networks in IoT systems, emphasizing resource optimization. Finally, reference [34] studied the AI-driven communication-efficient method at the edge of IoT. The contributions of [34] complement our work in minimizing communications overhead in IoT settings. Furthermore, study [35] proposed a multiresource allocation approach with fairness constraints; the ANNDRA-IoT model developed has considered the fairness of coverage consideration for resource allocation.

In [36], an imperialist competitive algorithm was used to implement IoT services in fog computing, focusing on resource utilization. Similarly, study [37] proposed a federated learning technique to achieve energy efficiency and resource optimization within environmentally sustainable IoT edge intelligence systems. These studies align with the goals of ANNDRA-IoT to maintain energy-efficient operations. Lastly, study [38] presented a hybrid authentication architecture for heterogeneous IoT systems, blending centralized and blockchain-based methods. This approach addresses security and efficiency, critical to the secure functioning of IoT environments. Works such as [39] and [40] further contribute to this discourse by exploring distributed IoT service placement and energy-efficient task offloading strategies, respectively.

2.2 Comparison of LSTM in IoT Resource Management

The management of resources within the IoT has seen significant advancements through the adoption of LSTM neural networks. Table 2 provides a comparison including rule-based systems, classic machine learning techniques, and different neural network designs.

Table 2: Comparison of ANNDRA-IoT with other approaches in IoT resource management

Approach	Performance	Scalability	Ease of Implementation
LSTM-based neural networks	enhanced capturing temporal dependencies, leading to enhanced forecasting accuracy.	Highly scalable with the ability to process large-scale time-series data efficiently.	Requires expertise in neural network configuration and training.
Rule-based systems	Limited by static rules, less effective in dynamic environments.	Scalability is constrained by the complexity of rules and their maintenance.	Relatively easier implementation but difficult to adapt to new environment without manual intervention.
Traditional machine learning algorithms (e.g., SVMs, DTs)	Effective in structured data environments but struggles with high-dimensional and sequential data.	Moderate scalability; performance may degrade with very large datasets.	Implementation complexity varies with the algorithm; significant domain knowledge is required for feature engineering.
Alternative neural network architectures (e.g., CNNs, GRUs)	Varies by architecture; CNNs are effective for spatial data, while GRUs offer advantages in temporal data processing.	Scalability is generally high, similar to LSTMs.	Implementation complexity is dependent on the specific architecture and the problem domain.

LSTM, a type of recurrent neural network, has the distinct ability to identify and retain temporal patterns, making it essential for managing resources in IoT scenarios characterized by time-series data. In [41], experimental evidence demonstrated that LSTMs exhibited superior predictive performance by effectively capturing temporal dynamics. This capability improves resource allocation strategies, thus improving the responsiveness of IoT systems [13,41,42]. In contrast, rule-based systems perform inadequately in dynamic IoT environments due to their static nature. Their inability to adapt to evolving data patterns renders them ineffective compared to LSTM models, which can seamlessly integrate new information and maintain consistent performance as system configurations evolve.

Conventional machine learning techniques, including support vector machines and decision trees, played a fundamental role in the early stages of IoT resource management. However, their reliance on manual feature selection and significant domain expertise limits their applicability in complex, high-dimensional sequential data scenarios [43,44]. LSTMs address these limitations by processing intricate sequential information with remarkable efficiency, which makes them particularly suited to the sophisticated demands of IoT frameworks. Other neural network architectures, such as convolutional neural networks (CNNs) and gated recurring units (GRUs), have their own strengths in the management of spatial and temporal data [45]. Hybrid models that integrate CNNs with LSTM have proven especially effective in tasks such as solar irradiance prediction and photovoltaic power [46]. These combinations balance computational efficiency with predictive accuracy, overcoming the constraints of single-architecture models in IoT applications [47].

3 Proposed ANNDRA-IoT Approach

This ANNDRA-IoT framework implements a type of neural network known as long-short-term memory neural networks to address disorders in resource allocation in heterogeneous IoT environments. This framework has considered IoT devices' heterogeneity by being able to adapt at run-time to different device capabilities and demands. The implementation details of ANNDRA-IoT will be elaborated here, which mentions the key contributions towards increasing system performance with the optimization of resource utilization.

3.1 Overview of Proposed Approach (ANNDRA-IoT)

ANNDRA-IoT has been proposed for resource management in highly heterogeneous IoTs. Introduce reconfigurability for adaptation in versatile and complex scenarios related to IoT ecosystems. By incorporating LSTM neural networks with Artificial Neural Networks (ANNs), ANNDRA-IoT enables dynamic allocation of resources, thus performing self-adaptation even under complex operational conditions. This system uses prediction capabilities as an intelligent one, working in real time to manage network and computational resources through device utilization patterns.

The architecture of ANNDRA-IoT includes LSTM units responsible for processing input data streams from a variety of IoT devices. These units analyze resource usage patterns and generate resource allocation decisions. Fig. 1 illustrates the architecture, showcasing the interaction among the system's components and the data flow within ANNDRA-IoT.

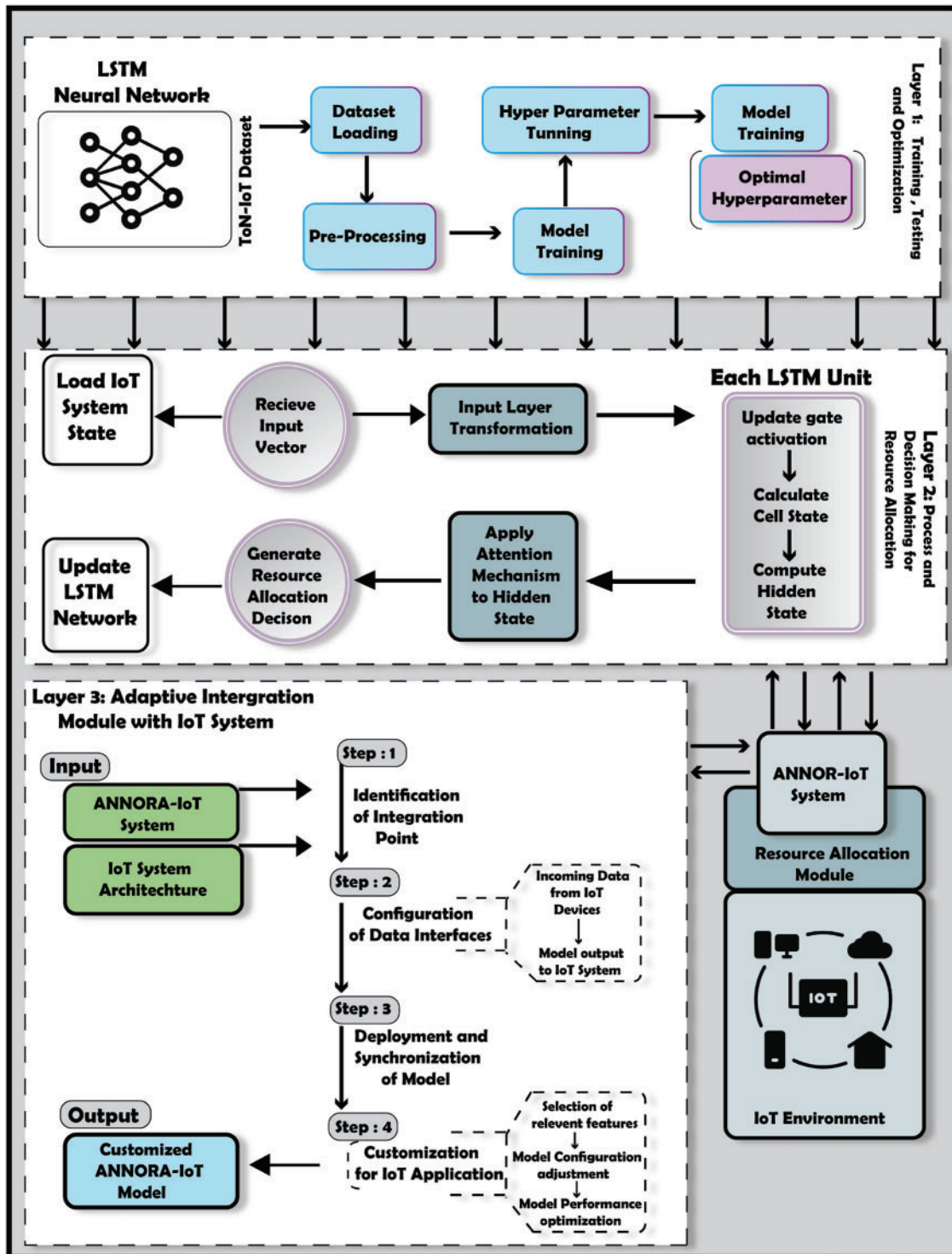


Figure 1: The architecture of the proposed ANNORA-IoT approach

3.2 LSTM-Based Real-Time IoT Resource Allocation

This section introduces a new LSTM neural network that is suitable for real-time IoT resource allocation. The framework is based on stochastic calculus, tensor analysis, variational optimization, differential geometry, and information theory; thus, it can learn sophisticated temporal and multi-dimensional dynamics of an IoT system.

Consider an IoT environment comprising a set of devices $\mathcal{D} = \{d_1, d_2, \dots, d_N\}$, each generating multidimensional data streams over continuous time. Let the state of the entire IoT system at time t be represented by a stochastic process $\mathbf{X}(t) \in \mathbb{R}^n$, where n denotes the dimensionality of the state space of the system. The objective is to determine an optimal resource allocation policy π^* that minimizes a cumulative cost functional over a finite time horizon $[0, T]$:

$$\pi^* = \operatorname{argmin}_{\pi \in \Pi} \mathbf{E} \left[\int_0^T \mathcal{L}(\mathbf{X}(t), \pi(\mathbf{X}(t), t), t) dt + \Phi(\mathbf{X}(T)) \right], \quad (1)$$

where Π is the set of admissible policies, \mathcal{L} is an instantaneous cost function capturing resource utilization and performance metrics, and Φ is a terminal cost function. The evolution of the state of the IoT system $\mathbf{X}(t)$ is governed by a controlled stochastic differential equation (SDE):

$$d\mathbf{X}(t) = \mathbf{f}(\mathbf{X}(t), \pi(\mathbf{X}(t), t), t) dt + \mathbf{G}(\mathbf{X}(t), t) d\mathbf{W}(t), \quad (2)$$

where $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^m \times [0, T] \rightarrow \mathbb{R}^n$ is the drift term, $\mathbf{G} : \mathbb{R}^n \times [0, T] \rightarrow \mathbb{R}^{n \times k}$ is the diffusion term and $\mathbf{W}(t)$ is a Wiener process of dimensions k representing environmental uncertainties. To approximate the optimal policy π^* , we construct an LSTM neural network enhanced with tensor operations and higher-order gates. Let $\mathbf{x}_t \in \mathbb{R}^p$ denote the input vector in time t , representing sensor readings and device statuses. The LSTM cell is redefined using the tensor calculus to handle multidimensional interactions:

$$\mathbf{i}_t = \sigma(\mathcal{W}_i * \mathbf{x}_t + \mathcal{U}_i * \mathbf{h}_{t-1} + \mathcal{V}_i * \mathbf{c}_{t-1} + \mathbf{b}_i), \quad (3)$$

$$\mathbf{f}_t = \sigma(\mathcal{W}_f * \mathbf{x}_t + \mathcal{U}_f * \mathbf{h}_{t-1} + \mathcal{V}_f * \mathbf{c}_{t-1} + \mathbf{b}_f), \quad (4)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tanh(\mathcal{W}_c * \mathbf{x}_t + \mathcal{U}_c * \mathbf{h}_{t-1} + \mathbf{b}_c), \quad (5)$$

$$\mathbf{o}_t = \sigma(\mathcal{W}_o * \mathbf{x}_t + \mathcal{U}_o * \mathbf{h}_{t-1} + \mathcal{V}_o * \mathbf{c}_t + \mathbf{b}_o), \quad (6)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t), \quad (7)$$

where $*$ denotes tensor contraction, \odot is the element-wise Hadamard product, and σ is the sigmoid activation function. Tensors $\mathcal{W}_i, \mathcal{U}_i, \mathcal{V}_i$, etc., are higher-order weight tensors that capture interactions across multiple dimensions. The LSTM network parameters are optimized using variational principles. We define an action functional \mathcal{S} associated with the policy π_θ parameterized by the network weights θ :

$$\mathcal{S}[\theta] = \mathbb{E} \left[\int_0^T \mathcal{L}(\mathbf{X}^\theta(t), \pi_\theta(\mathbf{X}^\theta(t), t), t) dt + \Phi(\mathbf{X}^\theta(T)) \right], \quad (8)$$

where $\mathbf{X}^\theta(t)$ denotes the state trajectory under policy π_θ . Optimization proceeds by computing the functional derivative $\delta \mathcal{S} / \delta \theta$ and updating θ by gradient descent in the function space. Recognizing that the state space of the IoT system may exhibit manifold structures, we introduce differential geometric concepts. Let \mathcal{M} be a smooth manifold representing the state space with a Riemannian metric $g_{\mathbf{X}}$. The geodesic distance between states \mathbf{X}_1 and \mathbf{X}_2 is given by:

$$d_{\mathcal{M}}(\mathbf{X}_1, \mathbf{X}_2) = \inf_{\gamma} \int_0^1 \sqrt{g_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))} dt, \quad (9)$$

where $\gamma : [0, 1] \rightarrow \mathcal{M}$ is a smooth path connecting \mathbf{X}_1 and \mathbf{X}_2 . We redefine the LSTM input transformation to map inputs onto \mathcal{M} :

$$\mathbf{z}_t = \exp_{\mathbf{h}_{t-1}}(\mathcal{W}_z * \mathbf{x}_t + \mathbf{b}_z), \quad (10)$$

where $\exp_{\mathbf{h}_{t-1}}$ is the Riemannian exponential map at \mathbf{h}_{t-1} . To capture non-Gaussian characteristics and higher-order dependencies in the data, we introduce gates based on higher-order statistics:

$$\mathbf{k}_t = \kappa(\mathcal{W}_k * \mathbf{x}_t + \mathcal{U}_k * \mathbf{h}_{t-1} + \mathbf{b}_k), \quad (11)$$

$$\mathbf{s}_t = \zeta(\mathcal{W}_s * \mathbf{x}_t + \mathcal{U}_s * \mathbf{h}_{t-1} + \mathbf{b}_s), \quad (12)$$

where κ and ζ are functions that extract kurtosis and skewness, respectively. The cell state update now includes these higher-order terms:

$$\begin{aligned} \mathbf{c}_t = & \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \mathbf{g}_t + \mathbf{k}_t \odot (\mathbf{c}_{t-1} \otimes \mathbf{c}_{t-1}) \\ & + \mathbf{s}_t \odot (\mathbf{c}_{t-1} \otimes \mathbf{c}_{t-1} \otimes \mathbf{c}_{t-1}), \end{aligned} \quad (13)$$

where \otimes denotes the element-wise tensor power. An advanced attention mechanism is incorporated using information-theoretic measures. The attention weights are defined on the basis of the mutual information between hidden states and outputs:

$$\alpha_t = \frac{\exp(I(\mathbf{h}_t; \mathbf{y}_t))}{\sum_{s=1}^T \exp(I(\mathbf{h}_s; \mathbf{y}_t))}, \quad (14)$$

where $I(\mathbf{h}_t; \mathbf{y}_t)$ is the mutual information between \mathbf{h}_t and the output \mathbf{y}_t . This attention mechanism ensures that the most informative hidden states are emphasized in the output computation:

$$\mathbf{y}_t = \phi\left(\sum_{s=1}^T \alpha_s \mathbf{h}_s\right), \quad (15)$$

with ϕ being a nonlinear activation function. The optimal control problem is connected to the Hamilton-Jacobi-Bellman (HJB) equation. The value function $V(\mathbf{X}, t)$ satisfies:

$$\begin{aligned} -\frac{\partial V}{\partial t} = & \min_{\pi} \left\{ \mathcal{L}(\mathbf{X}, \pi, t) + \left\langle \frac{\partial V}{\partial \mathbf{X}}, \mathbf{f}(\mathbf{X}, \pi, t) \right\rangle \right. \\ & \left. + \frac{1}{2} \text{Tr} \left(\mathbf{G} \mathbf{G}^\top \frac{\partial^2 V}{\partial \mathbf{X}^2} \right) \right\}, \end{aligned} \quad (16)$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product, and Tr is the trace operator. The LSTM network aims to approximate the policy π that minimizes the right-hand side of the HJB equation. To train the LSTM network in this complex setting, we employ Stochastic Gradient Hamiltonian Monte Carlo (SGHMC), which combines stochastic gradient descent with Hamiltonian dynamics. The parameter updates are given by:

$$d\theta = -\nabla_{\theta} \mathcal{L}[\theta] dt + \sqrt{2D(\theta)} dB_t, \quad (17)$$

where $D(\theta)$ is a diffusion matrix capturing the curvature of the loss landscape, and B_t is a Brownian motion.

3.3 Training and Optimization Framework for the LSTM Model

This section presents a novel mathematical framework for training and optimizing the LSTM model within the ANNDRA-IoT architecture. Let $\mathcal{D} = \{(\mathbf{x}_t, \mathbf{y}_t)\}_{t=1}^T$ denote the training dataset derived from the preprocessed TON_IoT data, where $\mathbf{x}_t \in \mathbb{R}^p$ represents the input features at time t , and $\mathbf{y}_t \in \mathbb{R}^q$ denotes the corresponding target output for resource allocation decisions. The primary objective is to find the optimal network parameters θ^* that minimize a composite loss function $\mathcal{J}(\theta)$, including the empirical risk and regularization terms:

$$\theta^* = \arg \min_{\theta \in \Theta} \mathcal{J}(\theta) = \arg \min_{\theta \in \Theta} [\mathcal{L}(\theta) + \lambda \mathcal{R}(\theta)], \quad (18)$$

where Θ is the parameter space, $\mathcal{L}(\theta)$ is the empirical loss function defined over the training data, $\mathcal{R}(\theta)$ is a regularization functional, and $\lambda > 0$ is a regularization coefficient controlling the trade-off between data fitting and model complexity. The empirical loss function is defined using a time-averaged functional that accounts for temporal dependencies.

$$\mathcal{L}(\theta) = \frac{1}{T} \sum_{t=1}^T \ell(\mathbf{y}_t, \hat{\mathbf{y}}_t(\theta)) + \frac{\beta}{2} \int_0^T \left\| \frac{d}{dt} \hat{\mathbf{y}}_t(\theta) \right\|^2 dt, \quad (19)$$

where $\ell(\cdot, \cdot)$ is a loss function that measures the discrepancy between the true output \mathbf{y}_t and the predicted output $\hat{\mathbf{y}}_t(\theta)$, $\|\cdot\|$ denotes the Euclidean norm, and $\beta > 0$ is a smoothing parameter that enforces the temporal smoothness through the Sobolev norm of the predictions. The regularization functional $\mathcal{R}(\theta)$ is defined using concepts from reproducing kernel Hilbert spaces (RKHS) to promote smoothness and generalization:

$$\mathcal{R}(\theta) = \|\theta\|_{\mathcal{H}}^2 = \sum_i \|W_i\|_{\mathcal{H}}^2 + \|U_i\|_{\mathcal{H}}^2 + \|b_i\|_{\mathcal{H}}^2, \quad (20)$$

where $\|\cdot\|_{\mathcal{H}}$ denotes the RKHS norm, and W_i, U_i, b_i are the weight matrices and bias vectors associated with the LSTM gates. This regularization encourages the parameters to lie in a smooth function space, enhancing the model's ability to generalize to unseen data. To solve the optimization problem, we employ stochastic gradient descent (SGD) enhanced with adaptive moment estimation and preconditioning matrices derived from second-order information:

$$\theta_{k+1} = \theta_k - \eta_k \mathbf{P}_k^{-1} \nabla_{\theta} \mathcal{J}(\theta_k), \quad (21)$$

where η_k is the learning rate in iteration k , $\nabla_{\theta} \mathcal{J}(\theta_k)$ is the loss function gradient, and \mathbf{P}_k is a preconditioning matrix calculated as:

$$\mathbf{P}_k = \varepsilon \mathbf{I} + (\hat{\mathbf{v}}_k)^{1/2}, \quad (22)$$

with ε being a small constant to ensure numerical stability, \mathbf{I} the identity matrix, and $\hat{\mathbf{v}}_k$ the exponentially weighted moving average of the squared gradients:

$$\hat{\mathbf{v}}_k = \gamma \hat{\mathbf{v}}_{k-1} + (1 - \gamma) (\nabla_{\theta} \mathcal{J}(\theta_k) \odot \nabla_{\theta} \mathcal{J}(\theta_k)), \quad (23)$$

where $\gamma \in [0, 1)$ is the decay rate and \odot denotes element-wise multiplication. This makes the learning rates of each parameter adaptive, which gives a better convergence behavior. Yet another variant of this

further incorporates information geometry in order to use the natural gradient that takes into account the Riemannian structure of the parameter space:

$$\theta_{k+1} = \theta_k - \eta_k \mathbf{F}^{-1}(\theta_k) \nabla_{\theta} \mathcal{J}(\theta_k), \quad (24)$$

where $\mathbf{F}(\theta_k)$ is the Fisher information matrix in iteration k , defined as:

$$\mathbf{F}(\theta) = \mathbb{E}_{\mathbf{x}_t} [\nabla_{\theta} \log p_{\theta}(\mathbf{y}_t | \mathbf{x}_t) \nabla_{\theta} \log p_{\theta}(\mathbf{y}_t | \mathbf{x}_t)^{\top}], \quad (25)$$

with $p_{\theta}(\mathbf{y}_t | \mathbf{x}_t)$ being the conditional probability of the output given the input under the model parameterized by θ . The natural gradient descent adapts to the curvature of the parameter space, potentially leading to faster convergence. To capture the uncertainty of the model and improve generalization, we adopt a Bayesian learning framework using variational inference. We define a variational posterior $q(\theta)$ over the model parameters and minimize the variational free energy $\mathcal{F}(q)$:

$$\begin{aligned} \mathcal{F}(q) &= \int q(\theta) \log \frac{q(\theta)}{p(\mathcal{D}, \theta)} d\theta \\ &= \text{KL}(q(\theta) \| p(\theta)) - \mathbb{E}_{q(\theta)} [\log p(\mathcal{D} | \theta)], \end{aligned} \quad (26)$$

where $p(\theta)$ is the prior distribution over parameters, $p(\mathcal{D}, \theta)$ is the joint probability of data and parameters, and KL denotes the Kullback-Leibler divergence. The optimization seeks a variational distribution $q(\theta)$ that approximates the true posterior, balancing the fit and complexity of the model. This study also introduces entropy regularization to encourage exploration during training.

$$\mathcal{J}_{\text{entropy}}(\theta) = \mathcal{J}(\theta) - \mu H[q(\theta)], \quad (27)$$

where $H[q(\theta)]$ is the differential entropy of the variational posterior:

$$H[q(\theta)] = - \int q(\theta) \log q(\theta) d\theta, \quad (28)$$

and $\mu > 0$ is a coefficient that controls the influence of regularization of the entropy. To analyze the convergence properties of the optimization algorithm, the proposed approach utilizes convex analysis and establishes conditions under which the loss function $\mathcal{J}(\theta)$ is convex or satisfies the inequality:

$$\frac{1}{2} \|\nabla_{\theta} \mathcal{J}(\theta)\|^2 \geq \mu_{\text{PL}} (\mathcal{J}(\theta) - \mathcal{J}(\theta^*)), \quad (29)$$

where $\mu_{\text{PL}} > 0$ is the PL constant. Under this condition, gradient descent methods exhibit linear convergence rates. In addition to standard L2 regularization, we incorporate Total Variation (TV) regularization to penalize rapid changes in the parameter space:

$$\mathcal{R}_{\text{TV}}(\theta) = \int \|\nabla \theta(s)\| ds, \quad (30)$$

where $\theta(s)$ denotes the parameters as a function over a continuous index s , and $\nabla \theta(s)$ is the gradient with respect to s . This approach further integrates adaptive learning rates based on the spectral properties of the Hessian matrix $\mathbf{H}(\theta)$ of the loss function:

$$\eta_k = \frac{2}{\lambda_{\max}(\mathbf{H}(\theta_k)) + \lambda_{\min}(\mathbf{H}(\theta_k))}, \quad (31)$$

where λ_{\max} and λ_{\min} are the largest and smallest eigenvalues of $\mathbf{H}(\theta_k)$, respectively. This choice of learning rate ensures stable convergence by taking into account the curvature of the loss landscape. The advanced training algorithm integrates the concepts mentioned above and is outlined as Algorithm 1.

Algorithm 1: Advanced training algorithm for the LSTM model

Result: Optimized LSTM Model for ANNDRA-IoT

1 **Initialization:** Initialize parameters θ_0 and variational posterior $q_0(\theta)$;

2 **for** $k = 0$ **to** K **do**

3 Sample mini-batch \mathcal{B}_k from training data;

4 Compute stochastic gradients $\nabla_{\theta} \mathcal{J}_{\mathcal{B}_k}(\theta_k)$;

5 Update preconditioning matrix \mathbf{P}_k ;

6 Update parameters using adaptive learning rate:

$$\theta_{k+1} = \theta_k - \eta_k \mathbf{P}_k^{-1} \nabla_{\theta} \mathcal{J}_{\mathcal{B}_k}(\theta_k)$$

 Update variational posterior $q_{k+1}(\theta)$ via Bayesian update rules;

7 Adjust learning rate η_{k+1} based on spectral analysis;

8 **end**

We incorporate constraints into the optimization problem to ensure feasibility with respect to resource limitations:

$$\text{Find } \theta^* \text{ such that } \theta^* = \arg \min_{\theta \in \Theta} \mathcal{J}(\theta), \quad \text{s.t. } \mathcal{C}(\theta) \leq C_{\max}, \quad (32)$$

where $\mathcal{C}(\theta)$ is a cost function representing computational or energy resources consumed by the model, and C_{\max} is the maximum allowable cost. This constrained optimization can be addressed using Lagrangian multipliers or penalty methods. To ensure the global convergence of the training algorithm, we employ Lyapunov stability analysis. We define a Lyapunov function $V(\theta_k) = \mathcal{J}(\theta_k) - \mathcal{J}(\theta^*)$ and show that:

$$V(\theta_{k+1}) - V(\theta_k) \leq -\alpha V(\theta_k), \quad (33)$$

for some $\alpha > 0$, implying exponential convergence to the global minimum.

3.4 Advanced Framework for Adaptation and Scalability of the LSTM Model in IoT Systems

This framework incorporated decentralized optimization, variational inference, and nonlinear dynamical systems to address the challenges posed by the heterogeneous and dynamic nature of IoT environments.

3.4.1 Manifold Embedding for Adaptive Representations

Let \mathcal{X} denote the high-dimensional input space of IoT data streams. We assume that the data lie on a lower-dimensional manifold $\mathcal{M} \subset \mathbb{R}^n$ embedded in \mathcal{X} . To capture the intrinsic geometry of the data, we map inputs to a manifold-adapted feature space using a diffeomorphic mapping $\phi: \mathcal{M} \rightarrow \mathbb{R}^d$. The LSTM cell is redefined to operate on manifold-valued inputs:

$$\mathbf{i}_t = \sigma(\mathbf{W}_i \phi(\mathbf{x}_t) + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i), \quad (34)$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_f \phi(\mathbf{x}_t) + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f), \quad (35)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tanh(\mathbf{W}_c \phi(\mathbf{x}_t) + \mathbf{U}_c \mathbf{h}_{t-1} + \mathbf{b}_c), \quad (36)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o \phi(\mathbf{x}_t) + \mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o), \quad (37)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t). \quad (38)$$

By operating on manifold embeddings, the model adapts to the underlying data structure, improving generalization. To ensure scalability, we model the LSTM parameters as elements of a Hilbert space \mathcal{H} endowed with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. The optimization of the network parameters $\theta \in \mathcal{H}$ is formulated as minimizing a functional $\mathcal{J} : \mathcal{H} \rightarrow \mathbb{R}$:

$$\theta^* = \arg \min_{\theta \in \mathcal{H}} \{ \mathcal{L}(\theta) + \lambda \|\theta\|_{\mathcal{H}}^2 \}, \quad (39)$$

where $\mathcal{L}(\theta)$ is the empirical loss and $\lambda > 0$ is a regularization parameter. The use of functional spaces allows for scalability through dimensionality reduction and efficient parameterization. In distributed IoT systems, data is generated across multiple devices. We employ decentralized optimization to train the LSTM model collaboratively. Let there be N devices, each with local data \mathcal{D}_i . The global objective is to minimize:

$$\theta^* = \arg \min_{\theta \in \mathcal{H}} \left\{ \frac{1}{N} \sum_{i=1}^N \mathcal{L}_i(\theta) + \lambda \|\theta\|_{\mathcal{H}}^2 \right\}. \quad (40)$$

We use a consensus-based algorithm where each device updates its local parameters θ_i and communicates with neighboring devices to reach agreement:

$$\theta_i^{(k+1)} = \theta_i^{(k)} - \eta \left(\nabla \mathcal{L}_i(\theta_i^{(k)}) + \sum_{j \in \mathcal{N}_i} (\theta_i^{(k)} - \theta_j^{(k)}) \right), \quad (41)$$

where \mathcal{N}_i is the set of neighboring devices of device i , and η is the learning rate. To adapt to non-stationary data distributions, we model the LSTM parameters probabilistically using variational inference. We define a posterior distribution $q(\theta)$ over the parameters and optimize the evidence lower bound (ELBO):

$$\mathcal{L}_{\text{ELBO}} = \mathbb{E}_{q(\theta)} [\log p(\mathcal{D}|\theta)] - \text{KL}(q(\theta) \| p(\theta)), \quad (42)$$

where $p(\theta)$ is the prior distribution, and $\text{KL}(\cdot)$ denotes the Kullback-Leibler divergence. This approach allows the model to adaptively update its beliefs about the parameters in response to new data. We analyze the stability of the LSTM dynamics by modeling the hidden states as a nonlinear discrete-time dynamical system:

$$\mathbf{h}_{t+1} = \mathbf{F}(\mathbf{h}_t, \mathbf{x}_t; \theta). \quad (43)$$

Using Lyapunov's direct method, we establish conditions for the asymptotic stability of the system by finding a Lyapunov function $V : \mathbb{R}^n \rightarrow \mathbb{R}_+$ such that:

$$V(\mathbf{h}_{t+1}) - V(\mathbf{h}_t) \leq -W(\mathbf{h}_t), \quad (44)$$

where $W(\mathbf{h}_t)$ is a positive definite function. We define a scalability metric S using mutual information to quantify the model's capacity to handle increasing data complexity:

$$S = \frac{I(\mathbf{Y}; \mathbf{H})}{C}, \quad (45)$$

where $I(\mathbf{Y}; \mathbf{H})$ is the mutual information between the output \mathbf{Y} and the hidden states \mathbf{H} , and C is the computational cost. A higher S indicates better scalability.

3.5 Integration with IoT Systems

The development of the integrative ANNDRA-IoT model in existing and future IoT systems is the last significant step towards a deployed model with full capabilities, as shown in Algorithm 2.

Algorithm 2: Integration and customization process for ANNDRA-IoT in IoT systems

Result: Integrated and Customized ANNDRA-IoT Model for IoT Systems

1 **Input:** ANNDRA-IoT Model, IoT System Architecture;

2 **Output:** Integrated and Customized ANNDRA-IoT Model;

3 **Step 1: Identify Integration Points;**

4 **for each** $node\ n\ in\ N_{IoT}$ **do**

5 **if** $Impact(n) > \theta$ **then**

6 Add n to I_p ;

7 **end**

8 **end**

9 **Step 2: Configure Data Interfaces;**

10 Set up D_{intf_in} for incoming data from IoT devices;

11 Set up D_{intf_out} for model's output to IoT system;

12 **Step 3: Deploy and Synchronize Model;**

13 Deploy ANNDRA-IoT model at I_p ;

14 Synchronize t_{model} with t_{IoT} using S_{sync} ;

15 **Step 4: Customize for IoT Applications;**

16 **for each** $IoT\ application\ App$ **do**

17 Select relevant features F_{app} based on $Relevance(f, App)$;

18 Adjust model configuration with $C_{dynamic}$;

19 Optimize model's performance for App using O_{app} ;

20 **end**

21 **Return** Integrated and Customized ANNDRA-IoT Model;

3.5.1 Process of Integrating ANNDRA-IoT in IoT Environments

The integration of ANNDRA-IoT within diverse IoT environments is a multifaceted process, involving several key steps to ensure seamless functionality and compatibility. This process is not just a mere deployment of the model but a harmonious fusion with the existing IoT ecosystem. The first step involves identifying specific points within the IoT infrastructure where the ANNDRA-IoT model can be most effective. This involves analyzing the architecture of the IoT network and pinpointing nodes where resource allocation decisions have the most significant impact. Let N_{IoT} represent the set of all nodes in the IoT network, the integration points I_p can be defined as:

$$I_p = \{n \in N_{IoT} | Impact(n) > \theta\}, \quad (46)$$

where $Impact(n)$ quantifies the influence of node n on network performance and θ is a predefined threshold. Configuring the data interface is crucial for the flow of information between the ANNDRA-IoT and the IoT system. This involves setting up data pipelines that feed real-time IoT data into the model and retrieve the model's allocation decisions. The data interface D_{intf} is represented as:

$$\begin{aligned} D_{intf_in}(x_{IoT}) &= x_{IoT} \rightarrow \text{ANNDRA-IoT}, \\ D_{intf_out}(y_{ANNORA}) &= y_{ANNORA} \rightarrow \text{IoT System}, \end{aligned} \quad (47)$$

where x_{IoT} represents the input data from the IoT devices and y_{ANNORA} is the output of the ANNDRA-IoT model. Deploying the ANNDRA-IoT model within identified integration points involves not just the installation of the model, but also its synchronization with the IoT system's operational tempo. This synchronization ensures that the model's resource allocation decisions are timely and contextually relevant. The synchronization function S_{sync} can be represented as:

$$S_{sync}(t_{model}, t_{IoT}) = \text{Align}(t_{model}, t_{IoT}), \quad (48)$$

where t_{model} and t_{IoT} are the time cycles of the ANNDRA-IoT model and the IoT system, respectively, and Align ensure their alignment.

3.5.2 Customization for Various IoT Applications

Customization is one of the building blocks to integrate ANNDRA-IoT into different IoT applications. Every IoT application has its unique features and requirements, and hence a tailored approach to its integration and realization. The customization process starts with selecting the relevant features for every tailored IoT application. Let F_{total} represent the total set of features available in the ANNDRA-IoT model. The subset of characteristics F_{app} for a specific application is selected based on its relevance and impact, formulated as:

$$F_{app} = \{f \in F_{total} | \text{Relevance}(f, \text{App}) > \tau\}, \quad (49)$$

where $\text{Relevance}(f, \text{App})$ quantifies the importance of the feature f for the application and τ is a threshold that determines the selection. The ANNDRA-IoT configuration model dynamically adjusts to the complexity and nature of the applied IoT. It includes the number of layers, units in the LSTM, and learning parameters. The dynamic configuration function for the application $C_{dynamic}$ is expressed as follows:

$$C_{dynamic}(N_{lstm}, N_{layers}, \alpha, \text{App}) = \text{Adjust}(N_{lstm}, N_{layers}, \alpha, \text{App}), \quad (50)$$

where N_{lstm} , N_{layers} , and α are LSTM units, the number of layers, and the learning rate, respectively, and Adjust the adjustment. The performance model has yet to be adjusted to the application for optimal performance loss through fine-tuning optimization. The optimization function O_{app} is defined as:

$$O_{app}(L_{loss}, M_{metrics}, \text{App}) = \text{Optimize}(L_{loss}, M_{metrics}, \text{App}), \quad (51)$$

where L_{loss} is the loss function, $M_{metrics}$ is a set of performance metrics, and Optimize optimizes the parameters for IoT applications.

4 Experimental Simulation and Results

This section elaborates the evaluation of the ANNDRA-IoT model through the experimental simulations. The experimental simulations were conducted with the use of simulation of falling edge computing technology within the IoTSim-Edge simulator. To ensure the robustness of the ANNDRA-IoT system, objective and subjective performance evaluations were performed. Objective evaluations include quantitative metrics such as resource utilization efficiency, response time, system performance, load balance effectiveness, and latency. Subjective evaluations emphasize practical implications, demonstrating the system's effectiveness in addressing real-world challenges like dynamic resource allocation and scalability in IoT environments. Extensive simulations have been conducted against several cutting-edge approaches to

demonstrate the superiority of ANNDRA-IoT in optimizing resource allocation. Specifically, we compared ANNDRA-IoT with the following models: DEELB [3], OSCAR [18], CNN-MBO [30] and AWPR-FOG [31].

4.1 Simulation Setup

IoTSim-Edge simulator emulates the complex IoT ecosystem within the simulation environment. The basis of our training and testing stages comprised data from the TON_IoT Dataset, obtained from UNSW Research. The realization also utilizes IoTSim-Edge's API for integrating the model within the simulated IoT environment.

4.2 Performance Evaluation Metrics

This section discusses the key performance metrics to be used to assess the efficacy of the ANNDRA-IoT model in optimal resource allocation within IoT environments. This indicates that 100 training epochs of the ANNDRA-IoT model accuracy metric are tracked. In a comparative analysis, the internal accuracy of the ANNDRA-IoT model over time was benchmarked compared to four important models in this field, namely DEELB, AWPR-FOG, and OSCAR. At epoch 100, the ANNDRA-IoT model outperformed the competing models with a training accuracy of 0.976 and a validation accuracy of 0.968 as illustrated in Table 3. In contrast, the DEELB model reported a training accuracy of 0.925 and a validation accuracy of 0.917, while the OSCAR model achieved a training accuracy of 0.902 and a validation accuracy of 0.895. The CNN-MBO model presented a training accuracy of 0.940 and a validation accuracy of 0.932, and the AWPR-FOG model showed a training accuracy of 0.950 and a validation accuracy of 0.942.

Table 3: Comparative analysis of accuracy and precision

Model	Accuracy		Precision	
	Training	Validation	Training	Validation
ANNDRA-IoT	0.976	0.968	0.97	0.957
DEELB	0.925	0.917	0.89	0.88
OSCAR	0.902	0.895	0.85	0.83
CNN-MBO	0.940	0.932	0.91	0.90
AWPR-FOG	0.950	0.942	0.93	0.92

The precision and recall of the ANNDRA-IoT model were observed for the 100 training epochs. The precision and recall of ANNDRA-IoT were compared with other algorithms. It can be seen in Table 4 that upon completion of the training, the ANNDRA-IoT model exhibited a training precision value of 0.97 and a recall value of 0.96. The precision and recall of the ANNDRA-IoT model in the classification states were recorded as the model was trained over 100 training epochs to assess its classification efficacy. At the end of the training, the ANNDRA-IoT model achieved training and validation precision of 0.97 and 0.96, and recall of 0.96 and 0.96, surpassing the corresponding metrics of the algorithms compared as represented in Table 4. The DEELB model was trained to achieve a precision of 0.89 and 0.88 and a recall of 0.87 and 0.86, for training and validation, respectively. OSCAR was trained and validated to reach a precision of 0.85 and 0.84 and a recall of 0.83 and 0.82, respectively. The CNN-MBO approach achieved a training precision of 0.91 and a validation precision of 0.90, a training recall of 0.90 and validation recall of 0.89. Finally, the AWPR-FOG model needed to achieve training precision of 0.93 and recall of 0.92, with corresponding validation measurements of 0.92 and 0.91.

Table 4: Comparative analysis of recall and F1 score

Model	Recall		F1 score	
	Training	Validation	Training	Validation
ANNDRA-IoT	0.96	0.954	0.965	0.955
DEELB	0.87	0.86	0.88	0.87
OSCAR	0.84	0.82	0.84	0.83
CNN-MBO	0.90	0.89	0.91	0.90
AWPR-FOG	0.92	0.91	0.93	0.92

For benchmarking the performance of the ANNDRA-IoT model, the F1 score's performance was compared to that of DEELB, OSCAR, CNN-MBO, and AWPR-FOG, where this comparison was done against the F1 score at the last epoch, which is 100. Herein, the ANNDRA-IoT model has given the highest performance with 0.965 of the training F1 score and 0.955 of the validation F1 score.

In this comparison, the DEELB model reached 0.88 for a training F1 score and 0.87 for a validation F1 score according to [3]. Besides, OSCAR, which was proposed in [18], reported 0.84 for a training F1 score and 0.83 for a validation F1 score, while the CNN-MBO approach of [30] had 0.91 for a training F1 score and 0.90 for a validation F1 score. The last model, AWPR-FOG, proposed in [31] reached 0.93 F1 score during training, and its F1 score when validated was 0.92.

4.3 Resource Allocation Optimization Metrics

In a comprehensive way of assessing the effectiveness of ANNDRA-IoT as a framework to optimize resource allocation in IoT environments, a set of custom metrics is used, including efficiency, response time, and overall system performance.

4.3.1 Resource Utilization Efficiency

Efficiency in resource utilization is one of the important metrics concerned with IoT systems; hence, a significant amount of performance analysis needs to be taken into consideration regarding variable load conditions. It was estimated that, when the load was low, the ANNDRA-IoT model gained an efficiency rate of 94%, which definitely outperformed the compared models: DEELB with 88%, OSCAR 85%, CNN-MBO 90%, and AWPR-FOG 92%. Thus, these numeric results depict that this model has great capability for the optimization of resources in under utilization conditions. Shown here in Fig. 2 are the performance results in different scenarios.

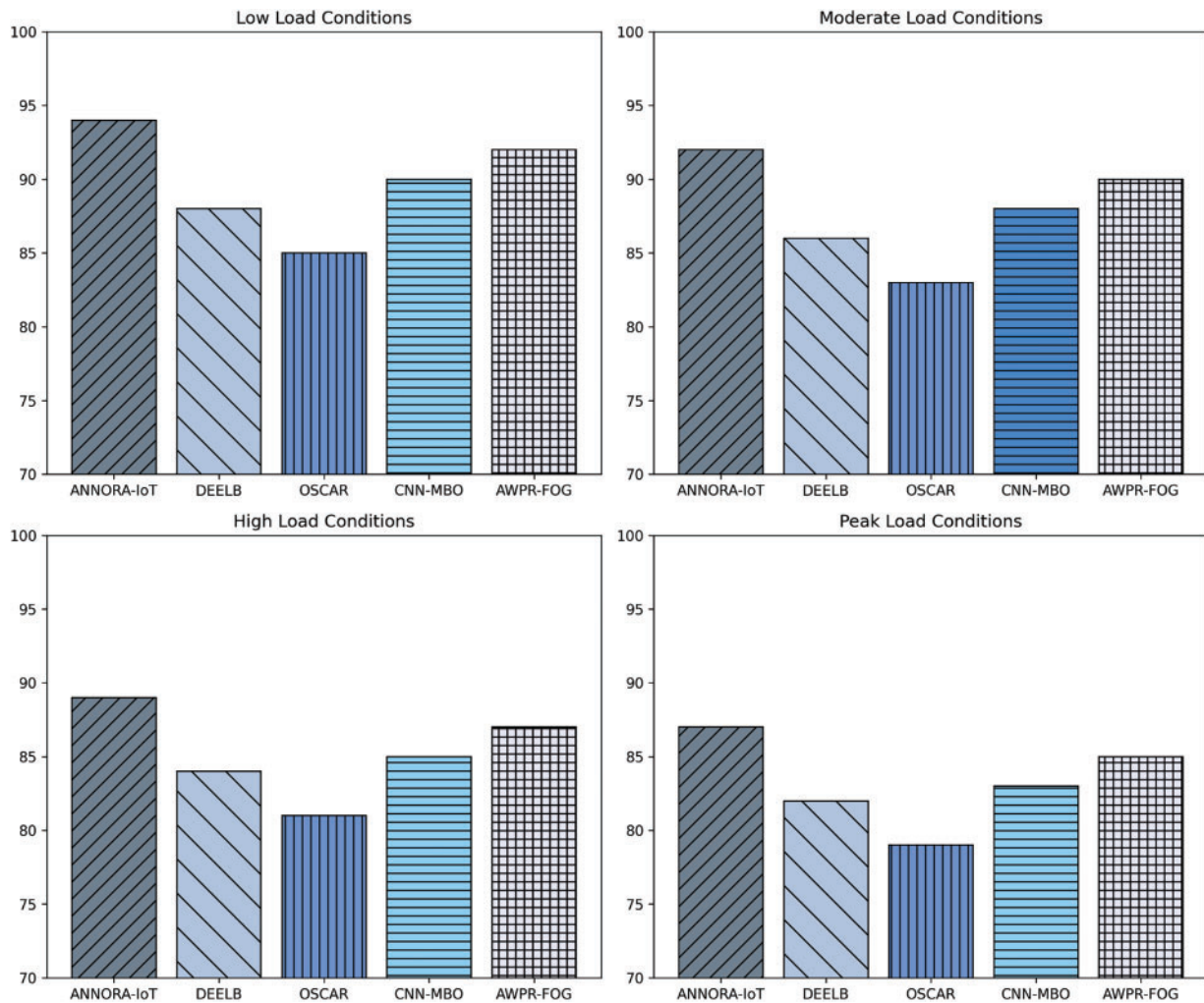


Figure 2: Load-varying efficiency comparison of IoT resource management models

In the case of the application of moderate load, ANNORA-IoT maintained an efficiency of 92%, while DEELB achieved an efficiency of 86%, OSCAR similarly reached 83%, CNN-MBO recorded 88%, and AWPR-FOG recorded 90%. Finally, for increasing load conditions, the efficiency of the ANNORA-IoT model was at 89%, well above that of DEELB at 84%, OSCAR at 81%, CNN-MBO at 85%, and AWPR-FOG at 87%, reflecting the capability of the model to manage resources in times of high load with much more efficiency. At peak load conditions, the ANNORA-IoT model achieved an efficiency of 87%, exceeding the performance of DEELB (82%), OSCAR (79%), CNN-MBO (83%) and AWPR-FOG (85%), which demonstrated superior performance in resource management under maximum demand.

4.3.2 Resource Allocation Response Time

Resource allocation responsiveness time is one of the useful performance indices of a model under varied operational conditions in IoT systems for a useful decision. In idle scenarios, characterized by minimal resource demands, the ANNORA-IoT model demonstrated a rapid response time of 20 ms (see Fig. 3), significantly faster than DEELB (35 ms), OSCAR (40 ms), CNN-MBO (30 ms) and AWPR-FOG (28 ms).

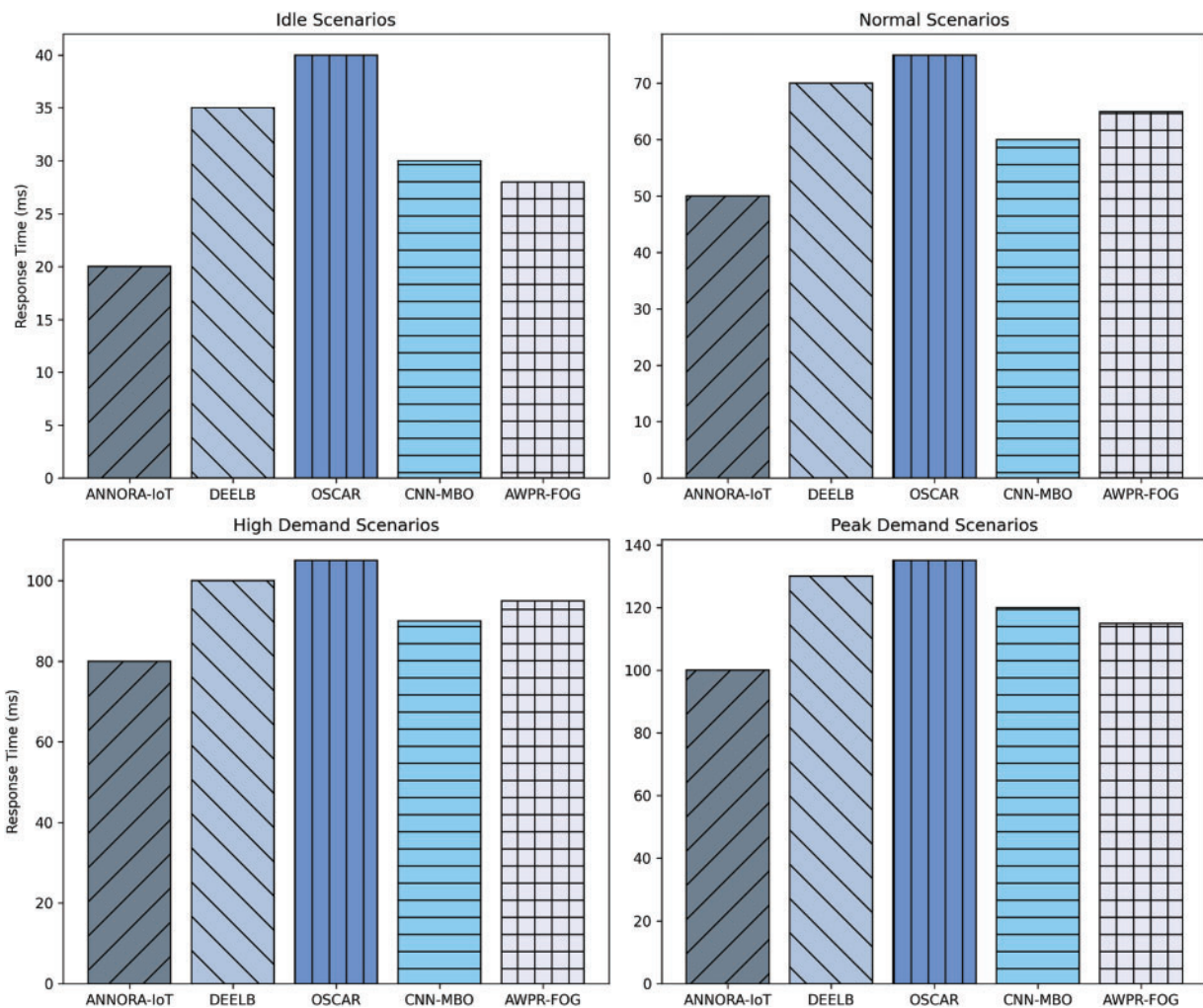


Figure 3: Comparative analysis of resource allocation response times under varying operational scenarios

Although ANNDRA-IoT had a higher response time during high-demand situations, where resource allocation becomes more difficult, of 80 ms vs. DEELB, OSCAR, CNN-MBO, and AWPR-FOG, who had 100, 105, 90, and 95 ms. Finally, in the peak demand condition, which represents the testing point at both the resource and system limit, ANNDRA-IoT records a response time of 100 ms, overcoming DEELB (130 ms), OSCAR (135 ms), CNN-MBO (120 ms), and AWPR-FOG (115 ms).

4.3.3 System Throughput

This study computed the overall work handled by the IoT system under the operational scenarios considered, which inherently represent quite challenging and distinctly different conditions. However, the results are presented in Fig. 4. In the idle scenario, where the system load was minimal, the throughput in the ANNDRA-IoT model was 1000 tasks per hour, much better than that of DEELB (850 tasks/h), OSCAR (800 tasks/h), CNN-MBO (900 tasks/h) and AWPR-FOG (920 tasks/h). In regular demands of operations, a moderate use scenario was observed through the throughput of the ANNDRA-IoT model at 2000 tasks per hour compared to DEELB (1700 tasks/h), OSCAR (1650 tasks/h), CNN-MBO (1800 tasks/h) and AWPR-FOG (1850 tasks/h).

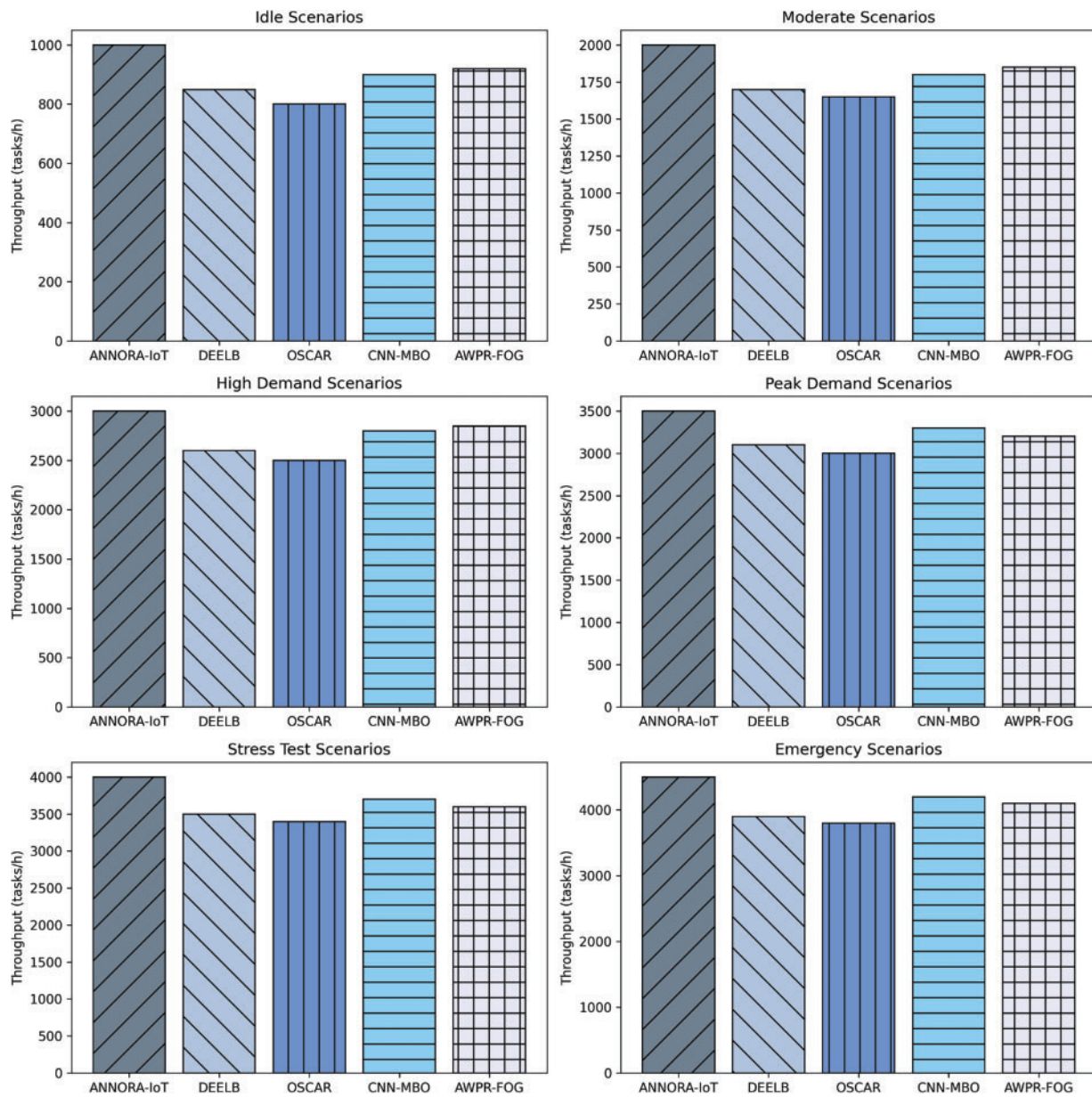


Figure 4: Comparative system throughput analysis across various operational scenarios

The use of the system is very high, in such a way that the demands for this system increase. ANNORA-IoT can handle a throughput of 3000 tasks per hour, which means that it is faster than DEELB (2600 tasks/h), OSCAR (2500 tasks/h), CNN-MBO (2800 tasks/h) and AWPR-FOG (2850 tasks/h). Under peak demand scenarios, i.e., the limits of the system being tested, the throughput arrived at for the ANNORA-IoT model was 3500 tasks per hour against DEELB (3100 tasks/h), OSCAR (3000 tasks/h), CNN-MBO (3300 tasks/h), and AWPR-FOG (3200 tasks/h). In such conditions, the ANNORA-IoT model registered a throughput of 4000 tasks per hour in a stress test scenario, leaving DEELB at 3500 tasks per hour, OSCAR at 3400 tasks per hour, and CNN-MBO at 3700 tasks per hour, while AWPR-FOG remained at 3600 tasks per hour. Finally, in the worst-case situation of an emergency scenario, with demands being both urgent and unexpected and

that these demands reach a high volume, the ANNDRA-IoT throughput model in this work is 4500 tasks per hour, while that of DEELB is 3900 tasks/h, OSCAR is 3800 tasks/h, CNN-MBO is 4200 tasks/h, and AW.

4.3.4 Load Balancing Effectiveness

This study evaluated the performance of the ANNDRA-IoT model under four distinct operational scenarios: low utilization, average utilization, high utilization, and peak utilization, while the result of these is presented in Fig. 5. The first low utilization scenario is implemented with minimal system load, the ANNDRA-IoT model achieved a load balancing efficiency of 95%, indicating its superior capability to evenly distribute work. This was significantly higher than DEELB (87%), OSCAR (85%), CNN-MBO (90%), and AWPR-FOG (88%), demonstrating ANNDRA-IoT's advanced load management under underutilized conditions. In addition, in the average utilization scenario that represents regular operational demands, the ANNDRA-IoT model maintained a load balancing efficiency of 92%, outperforming DEELB (84%), OSCAR (80%), CNN-MBO (86%) and AWPR-FOG (85%).

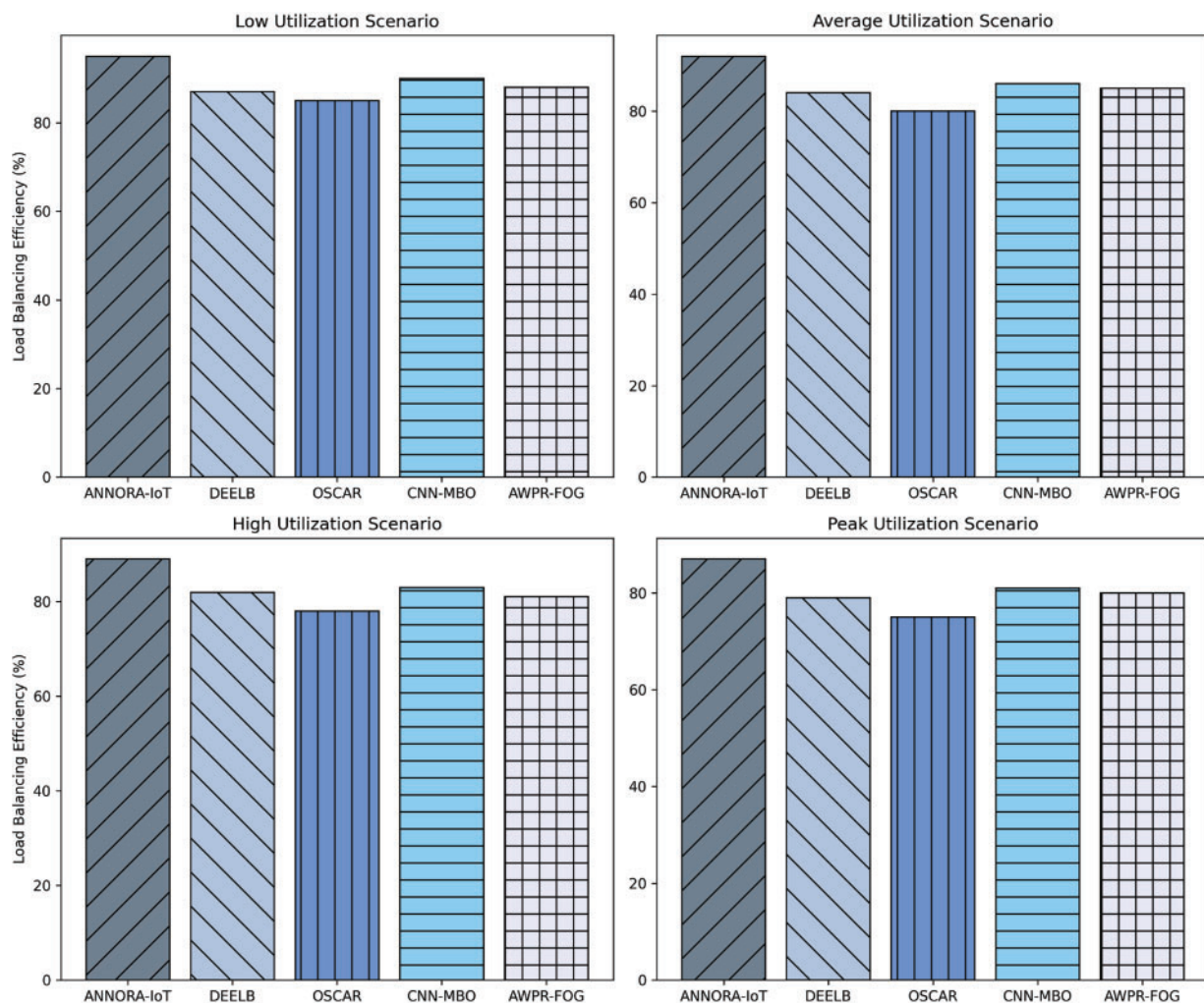


Figure 5: Comparative load balancing effectiveness across various utilization scenarios

In a high utilization scenario implemented under increased demands, where effective load balancing becomes more challenging, ANNDRA-IoT showcased an efficiency of 89%. In comparison, DEELB recorded 82%, OSCAR 78%, CNN-MBO 83%, and AWPR-FOG 81%. Upon further evaluation of performance, another scenario is implemented named the Peak Utilization Scenario. During peak utilization tests, the ANNDRA-IoT model demonstrated its exceptional load balancing capability with an efficiency of 87%, surpassing DEELB (79%), OSCAR (75%), CNN-MBO (81%) and AWPR-FOG (80%).

4.3.5 Latency Analysis

The latency performance of the ANNDRA-IoT model was compared to DEELB, OSCAR, CNN-MBO, and AWPR-FOG in four distinct scenarios. A baseline scenario was established without the ANNDRA-IoT model to highlight its impact on system latency. The results for each scenario are presented below:

- **Idle Environment:** In scenarios with minimal resource demand, ANNDRA-IoT demonstrated a latency of **10 ms**, outperforming DEELB (18 ms), OSCAR (22 ms), CNN-MBO (15 ms), and AWPR-FOG (12 ms). Without ANNDRA-IoT, the system latency increased to **25 ms**.
- **Normal Operations:** During typical operations, ANNDRA-IoT maintained a latency of **15 ms**, significantly lower than DEELB (25 ms), OSCAR (28 ms), CNN-MBO (22 ms), and AWPR-FOG (20 ms). In systems without ANNDRA-IoT, the latency rose to **30 ms**.
- **High Demand:** Under conditions of high demand, ANNDRA-IoT achieved a latency of **25 ms**, compared to DEELB (35 ms), OSCAR (40 ms), CNN-MBO (30 ms), and AWPR-FOG (28 ms). Systems without ANNDRA-IoT recorded a latency of **45 ms**.
- **Peak Demand:** During peak demand scenarios, ANNDRA-IoT maintained a latency of **40 ms**, outperforming DEELB (50 ms), OSCAR (55 ms), CNN-MBO (45 ms), and AWPR-FOG (42 ms). Systems without ANNDRA-IoT exhibited a latency of **60 ms**.

Fig. 6 depicts the outcome in these four scenarios that demonstrates the adaptability and efficiency of the ANNDRA-IoT model in diverse operational environments. These results validate the ANNDRA-IoT model as a superior solution for the allocation of real-time IoT resources.

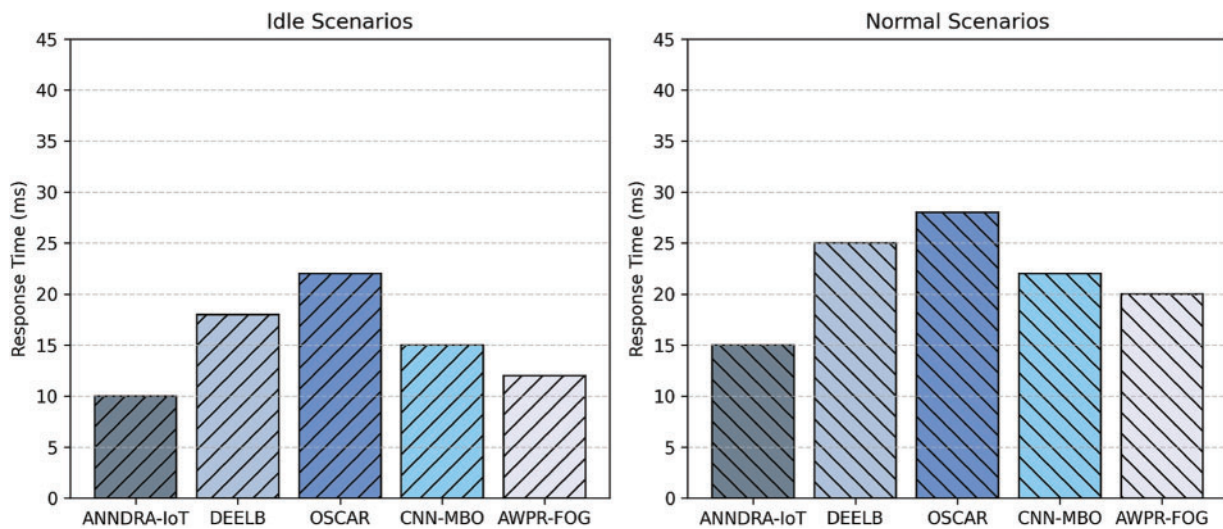


Figure 6: (Continued)

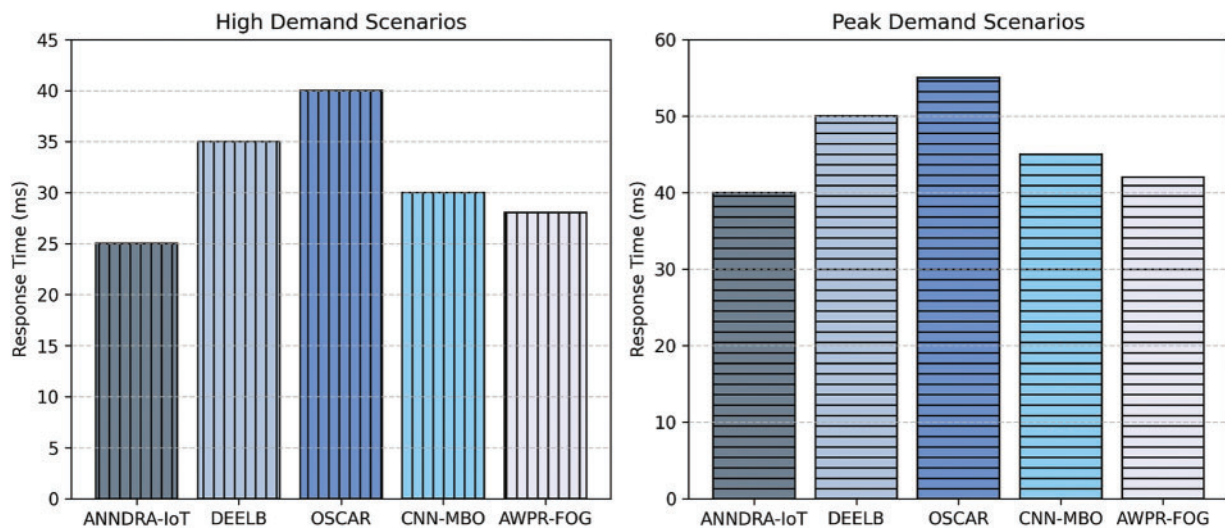


Figure 6: Comparative load balancing effectiveness across various utilization scenarios

5 Discussion

The proposed ANNDRA-IoT model is a novel architecture based on adaptive neural networks that addresses the complexities and dynamic nature of IoT environments. The adaptive approach employed by ANNDRA-IoT facilitates optimal resource distribution, ensuring efficient utilization while minimizing latency and improving overall system throughput. The key to this approach is its ability to intelligently interpret and respond to different IoT scenarios, thus boosting the performance and reliability of IoT systems. The implementation of LSTM-based neural networks for IoT resource management faces several key challenges. The large-scale data collection and processing carried out by these models poses a potential threat to data privacy, necessitating the development of privacy-preserving mechanisms.

Our simulation results demonstrate that ANNDRA-IoT outperforms state-of-the-art solutions such as DEELB, OSCAR, CNN-MBO, and AWPR-FOG. For example, ANNDRA-IoT achieves 95% efficiency in highly loaded scenarios, a significant improvement over DEELB's 87% and OSCAR's 85%. In terms of latency reductions, the proposed model reduces latency to 15 ms, as opposed to 22 ms in CNN-MBO and 20 ms in AWPR-FOG. These results demonstrate not only the effectiveness of ANNDRA-IoT in diverse operational conditions but also its potential to revolutionize IoT resource management.

The possible future expansion would be the incorporation of state-of-the-art machine learning algorithms for predictive analytics, allowing the model to predict future resource requirements based on trends in historical data. In addition, research for the incorporation of some edge computing paradigms may enhance the processing power of the model to capture, trace, and retrieve the nearest possible data from its source, reducing latency further with an exceptionally superfast decision-making processes' speed.

6 Conclusion

This paper introduces the ANNDRA-IoT model that provides a significant improvement in resource allocation in environmental settings of the IoT. The ANNDRA-IoT model effectively combats the complex and dynamic natures associated with IoT systems through its innovative neural network-based architecture. The implementation of the model has led to many achievements in improving system performance. ANNDRA-IoT lowered latency to 15 ms in simulation scenarios as compared to CNN-MBO's 22 ms and AWPR-FOG's 20 ms. The efficiency in resource utilization under low utilization scenarios hit peaks at 95%

with DEELB only at 87% and OSCAR only at 85%. Moreover, the throughput of the system with ANNDRA-IoT has been measured in high-demand scenarios that were 3000 tasks per hour, better than the DEELB and OSCAR results of 2600 tasks/h and 2500 tasks/h, respectively. The efficiency in load balancing was also superior in high utilization scenarios and registered an efficiency 89% compared to DEELB 82% and OSCAR 78%. The primary limitations of this study include the computational complexity of implementing LSTM networks on resource-constrained devices. Future extension of the proposed approach can include more diversity in the data set to increase efficiency in resource allocation.

Acknowledgement: None.

Funding Statement: The authors gratefully acknowledge the funding of the Deanship of Graduate Studies and Scientific Research, Jazan University, Saudi Arabia, through Project Number: ISP-2024.

Author Contributions: The authors confirm contributions to the paper as follows: study conception and design: Abdullah M. Alqahtani, Kamran Ahmad Awan; data collection: Kamran Ahmad Awan, Abdulaziz Almaleh; analysis and interpretation of results: Abdullah M. Alqahtani, Kamran Ahmad Awan, Osama Aletri; draft manuscript preparation: Kamran Ahmad Awan, Osama Aletri. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data used in this study is derived from the publicly available TON_IoT Dataset, which can be accessed at [The TON_IoT Datasets](#) (accessed on 29 January 2025). The code utilized for this research is currently part of ongoing work and will be made available upon request to maintain the integrity and progress of the research.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Zaidi T, Usman M, Aftab MU, Aljuaid H, Ghadi YY. Fabrication of flexible role-based access control based on blockchain for Internet of Things use cases. *IEEE Access*. 2023;11:106315–33. doi:10.1109/ACCESS.2023.3318487.
2. Ahmed SF, Alam MSB, Afrin S, Rafa SJ, Taher SB, Kabir M, et al. Towards a secure 5G-enabled Internet of Things: a survey on requirements, privacy, security, challenges, and opportunities. *IEEE Access*. 2024;12(8):13125–45. doi:10.1109/ACCESS.2024.3352508.
3. Shuaib M, Bhatia S, Alam S, Masih RK, Alqahtani N, Basheer S, et al. An optimized, dynamic, and efficient load-balancing framework for resource management in the Internet of Things (IoT) environment. *Electronics*. 2023;12(5):1104. doi:10.3390/electronics12051104.
4. Zhang Y, He D, Vijayakumar P, Luo M, Huang X. SAPFS: an efficient symmetric-key authentication key agreement scheme with perfect forward secrecy for industrial internet of things. *IEEE Internet Things J*. 2023;10(11):9716–26. doi:10.1109/JIOT.2023.3234178.
5. Mowla MN, Mowla N, Shah AS, Rabie K, Shongwe T. Internet of Things and wireless sensor networks for smart agriculture applications: a survey. *IEEE Access*. 2023;11(2):145813–52. doi:10.1109/ACCESS.2023.3346299.
6. Bakar KA, Zuhra FT, Isyaku B, Sulaiman SB. A review on the immediate advancement of the Internet of Things in wireless telecommunications. *IEEE Access*. 2023;11(70):21020–48. doi:10.1109/ACCESS.2023.3250466.
7. Araújo SO, Peres RS, Filipe L, Manta-Costa A, Lidon F, Ramalho JC, et al. Intelligent data-driven decision support for agricultural systems-ID3SAS. *IEEE Access*. 2023;11:115798–815. doi:10.1109/ACCESS.2023.3324813.
8. Zhang F, Han G, Li A, Lin C, Liu L, Zhang Y, et al. QoS-driven distributed cooperative data offloading and heterogeneous resource scheduling for IIoT. *IEEE Internet Things Mag*. 2023;6(3):118–24. doi:10.1109/IOTM.001.2200264.
9. Singh M, Sahoo KS, Gandomi AH. An intelligent IoT-based data analytics for freshwater recirculating aquaculture system. *IEEE Internet Things J*. 2023;11(3):4206–17. doi:10.1109/JIOT.2023.3298844.

10. Singh SK, Kumar M, Tanwar S, Park JH. GRU-based digital twin framework for data allocation and storage in IoT-enabled smart home networks. *Future Gener Comput Syst.* 2024;153(1):391–402. doi:10.1016/j.future.2023.12.009.
11. Xu A, Hu Z, Zhang X, Xiao H, Zheng H, Chen B, et al. QDRL: queue-aware online DRL for computation offloading in industrial Internet of Things. *IEEE Internet Things J.* 2023;11(5):7772–86. doi:10.1109/JIOT.2023.3316139.
12. Jeribi F, Amin R, Alhameed M, Tahir A. An efficient trust management technique using ID3 algorithm with blockchain in smart buildings IoT. *IEEE Access.* 2022;11:8136–49. doi:10.1109/ACCESS.2022.3230944.
13. Zhang F, Han G, Liu L, Zhang Y, Peng Y, Li C. Cooperative partial task offloading and resource allocation for IIoT based on decentralized multi-agent deep reinforcement learning. *IEEE Internet Things J.* 2023;11(3):5526–44. doi:10.1109/JIOT.2023.3306803.
14. Khan D, Alshahrani A, Almjally A, Al Mudawi N, Algarni A, Al Nowaiser K, et al. Advanced IoT-based human activity recognition and localization using deep polynomial neural network. *IEEE Access.* 2024;12:94337–53. doi:10.1109/ACCESS.2024.3420752.
15. Minhaj SU, Mahmood A, Abedin SF, Hassan SA, Bhatti MT, Ali SH, et al. Intelligent resource allocation in LoRaWAN using machine learning techniques. *IEEE Access.* 2023;11(4):10092–106. doi:10.1109/ACCESS.2023.3240308.
16. George A, Ravindran A, Mendieta M, Tabkhi H. Mez: an adaptive messaging system for latency-sensitive multi-camera machine vision at the IoT edge. *IEEE Access.* 2021;9:21457–73. doi:10.1109/ACCESS.2021.3055775.
17. Zou S, Wu J, Yu H, Wang W, Huang L, Ni W, et al. Efficiency-optimized 6G: a virtual network resource orchestration strategy by enhanced particle swarm optimization. *Dig Commun Netw.* 2024;10(5):1221–33. doi:10.1016/j.dcan.2023.06.008.
18. Wadhwa H, Aron R. Optimized task scheduling and preemption for distributed resource management in fog-assisted IoT environment. *J Supercomput.* 2023;79(2):2212–50. doi:10.1007/s11227-022-04747-2.
19. Peng M, Zhang K. Recent advances in fog radio access networks: performance analysis and radio resource allocation. *IEEE Access.* 2016;4:5003–9. doi:10.1109/ACCESS.2016.2603996.
20. Park H, Kim N, Lee GH, Choi JK. MultiCNN-FilterLSTM: resource-efficient sensor-based human activity recognition in IoT applications. *Future Gener Comput Syst.* 2023;139(15):196–209. doi:10.1016/j.future.2022.09.024.
21. Walia GK, Kumar M, Gill SS. AI-empowered fog/edge resource management for IoT applications: a comprehensive review, research challenges and future perspectives. *IEEE Commun Surv Tutor.* 2023;26(1):619–69. doi:10.1109/COMST.2023.3338015.
22. Xu M. A novel machine learning-based framework for channel bandwidth allocation and optimization in distributed computing environments. *EURASIP J Wirel Commun Netw.* 2023;2023(1):97. doi:10.1186/s13638-023-02310-y.
23. Ke H, Wang J, Wang H, Ge Y. Joint optimization of data offloading and resource allocation with renewable energy aware for IoT devices: a deep reinforcement learning approach. *IEEE Access.* 2019;7:179349–63. doi:10.1109/ACCESS.2019.2959348.
24. Singh J, Singh P, Hedabou M, Kumar N. An efficient machine learning-based resource allocation scheme for SDN-enabled fog computing environment. *IEEE Trans Veh Technol.* 2023;72(6):8004–17. doi:10.1109/TVT.2023.3242585.
25. Otoshi T, Murata M, Shimonishi H, Shimokawa T. Hierarchical Bayesian attractor model for dynamic task allocation in edge-cloud computing. In: *2023 International Conference on Computing, Networking and Communications (ICNC)*; 2023; Honolulu, HI, USA. p. 12–8.
26. Liu L, Yuan X, Chen D, Zhang N, Sun H, Taherkordi A. Multi-user dynamic computation offloading and resource allocation in 5G MEC heterogeneous networks with static and dynamic subchannels. *IEEE Trans Veh Technol.* 2023;72(11):14924–38. doi:10.1109/TVT.2023.3285069.
27. Fan W, Liu X, Yuan H, Li N, Liu Y. Time-slotted task offloading and resource allocation for cloud-edge-end cooperative computing networks. *IEEE Trans Mob Comput.* 2024;23(8):8225–41. doi:10.1109/TMC.2024.3349551.
28. Liu Y, Wang Y, Shen Y. Delay Doppler division multiple access resource allocation in aircraft network. *IEEE Internet Things J.* 2024;11(17):27881–93. doi:10.1109/JIOT.2024.3352030.

29. Ortiz F, Skatchkovsky N, Lagunas E, Martins WA, Eappen G, Daoud S, et al. Energy-efficient on-board radio resource management for satellite communications via neuromorphic computing. *IEEE Trans Mach Learn Commun Netw.* 2024;2:169–89. doi:10.1109/TMLCN.2024.3352569.
30. Badri S, Alghazzawi DM, Hasan SH, Alfayez F, Hasan SH, Rahman M, et al. An efficient and secure model using adaptive optimal deep learning for task scheduling in cloud computing. *Electronics.* 2023;12(6):1441. doi:10.3390/electronics12061441.
31. Kumar M, Kishor A, Samariya JK, Zomaya AY. An autonomic workload prediction and resource allocation framework for fog-enabled industrial IoT. *IEEE Internet Things J.* 2023;10(11):9513–22. doi:10.1109/JIOT.2023.3235107.
32. Baccour E, Mhaisen N, Abdellatif AA, Erbad A, Mohamed A, Hamdi M, et al. Pervasive AI for IoT applications: a survey on resource-efficient distributed artificial intelligence. *IEEE Commun Surv Tutor.* 2022;24(4):2366–418. doi:10.1109/COMST.2022.3200740.
33. Kuthadi VM, Selvaraj R, Baskar S, Shakeel PM, Ranjan A. Optimized energy management model on data distributing framework of wireless sensor network in IoT system. *Wirel Pers Commun.* 2022;127(2):1377–403. doi:10.1007/s11277-021-08583-0.
34. Mwase C, Jin Y, Westerlund T, Tenhunen H, Zou Z. Communication-efficient distributed AI strategies for the IoT edge. *Future Gener Comput Syst.* 2022;131(3):292–308. doi:10.1016/j.future.2022.01.013.
35. Chen S, Yang C, Huang W, Liang W, Ke N, Souri A, et al. Fairness constraint efficiency optimization for multiresource allocation in a cluster system serving Internet of Things. *Int J Commun Syst.* 2023;36(3):e5395. doi:10.1002/dac.5395.
36. Zare M, Sola YE, Hasanpour H. Imperialist competitive based approach for efficient deployment of IoT services in fog computing. *Cluster Comput.* 2023;27:1–14.
37. Salh A, Ngah R, Audah L, Kim KS, Abdullah Q, Al-Moliki YM, et al. Energy-efficient federated learning with resource allocation for green IoT edge intelligence in B5G. *IEEE Access.* 2023;11:16353–67. doi:10.1109/ACCESS.2023.3244099.
38. Khashan OA, Khafajah NM. Efficient hybrid centralized and blockchain-based authentication architecture for heterogeneous IoT systems. *J King Saud Univ Comput Inf Sci.* 2023;35(2):726–39. doi:10.1016/j.jksuci.2023.01.011.
39. Huangpeng Q, Yahya RO. Distributed IoT services placement in fog environment using optimization-based evolutionary approaches. *Expert Syst Appl.* 2024;237(1):121501. doi:10.1016/j.eswa.2023.121501.
40. Diwaker C, Sharma A. ETOSP: energy-efficient task offloading strategy based on partial offloading in mobile edge computing framework for efficient resource management. *J Auton Intell.* 2024;7(3).
41. Lara-Benítez P, Carranza-García M, Riquelme JC. An experimental review on deep learning architectures for time series forecasting. *Int J Neural Syst.* 2021;31(3):2130001. doi:10.1142/S0129065721300011.
42. Liu Y, Liu A, Xia Y, Hu B, Liu J, Wu Q, et al. A blockchain-based cross-domain authentication management system for IoT devices. *IEEE Trans Netw Sci Eng.* 2023;11(1):115–27. doi:10.1109/TNSE.2023.3292624.
43. Muazu T, Yingchi M, Muhammad AU, Ibrahim M, Samuel O, Tiwari P. IoMT: a medical resource management system using edge-empowered blockchain federated learning. *IEEE Trans Netw Serv Manag.* 2023;21(1):517–34. doi:10.1109/TNSM.2023.3308331.
44. Syu JH, Lin JCW, Srivastava G, Yu K. A comprehensive survey on artificial intelligence empowered edge computing on consumer electronics. *IEEE Trans Consum Electron.* 2023;69(4):1023–34. doi:10.1109/TCE.2023.3318150.
45. Nabi ST, Islam MR, Alam MGR, Hassan MM, AlQahtani SA, Aloji G, et al. Deep learning-based fusion model for multivariate LTE traffic forecasting and optimized radio parameter estimation. *IEEE Access.* 2023;11(5):14533–49. doi:10.1109/ACCESS.2023.3242861.
46. Rajagukguk RA, Ramadhan RAA, Lee HJ. A review on deep learning models for forecasting time series data of solar irradiance and photovoltaic power. *Energies.* 2020;13(24):6623. doi:10.3390/en13246623.
47. Ali NF, Atef M. An efficient hybrid LSTM-ANN joint classification-regression model for PPG-based blood pressure monitoring. *Biomed Signal Process Control.* 2023;84(2):104782. doi:10.1016/j.bspc.2023.104782.