



RESEARCH ARTICLE

Data Centre Optimization for Cloud Computing and Virtualization

Jambhu Kumar V^{1*}, Dr. Anusha M²

Abstract

Cloud computing has significantly transformed modern data centre architectures by enabling scalable and flexible resource provisioning through virtualization. Despite its advantages, achieving energy efficiency while maintaining service reliability remains a critical challenge in virtualized cloud environments. Existing virtual machine (VM) placement and consolidation approaches often emphasize aggressive energy minimization, which can lead to excessive VM migrations and increased Service Level Agreement (SLA) violations under dynamic workloads.

This study proposes a Learning-Guided Adaptive Multi-Objective Virtual Machine Optimization (LG-AMOVm) framework that prioritizes stable, SLA-aware energy optimization in cloud data centres. The framework integrates workload trend prediction, adaptive utilization thresholds, and a net benefit-driven migration model to dynamically control VM placement and consolidation. Unlike static threshold-based or purely energy-centric methods, LG-AMOVm incorporates workload variability, migration overhead, total energy consumption, SLA compliance, and SLA risk into a unified optimization process, ensuring that migration decisions yield positive global benefits.

Experimental results demonstrate that the proposed approach consistently reduces energy consumption while maintaining zero SLA violations and limiting unnecessary VM migrations under dynamic conditions. Compared to baseline strategies, LG-AMOVm achieves balanced, adaptive optimization, demonstrating that energy-efficient cloud operation can be realized without compromising service stability.

Keywords: Data Centre Optimization, Virtual Machine Consolidation, Energy-Aware Resource Allocation, Cloud computing, SLA Scheduling, LG AMOVm.

Introduction

Cloud computing has experienced rapid growth, fundamentally transforming the design, deployment, and management of computing resources. Organizations increasingly rely on cloud-based infrastructure to deliver

scalable, flexible, and cost-effective services, accelerating the evolution of traditional data centres into highly optimized cloud environments (Segun-Falade *et al.*, 2024; Katal *et al.*, 2023). These data centres provide on-demand access to shared resources, including servers, storage, networks, and applications. However, the widespread adoption of cloud technologies has intensified the need for efficient data centre operations that ensure high performance, energy efficiency, reliability, and sustainability (Oye *et al.*, 2024; Buyya *et al.*, 2024).

Data centre optimization focuses on improving resource utilization, workload management, and energy efficiency. Traditional data centres often suffer from underutilized hardware, excessive power consumption, and limited scalability (Safari *et al.*, 2025). Virtualization addresses these challenges by decoupling software services from physical hardware and enabling multiple virtual machines (VMs) or containers to operate on a single server (Odun-Ayo *et al.*, 2017). This supports dynamic resource allocation, efficient workload consolidation, and improved infrastructure utilization, reducing operational costs and energy consumption (Mavridis *et al.*, 2019; Ewim *et al.*, 2023).

¹Research Scholar, PG & Research Department of Computer Science, National College (Autonomous) Trichy, 620001, Bharathidasan University, India

²Assistant Professor, PG & Research Department of Computer Science, National College (Autonomous) Trichy, 620001, Bharathidasan University, India

***Corresponding Author:** Jambhu Kumar V Research Scholar, PG & Research Department of Computer Science, National College (Autonomous) Trichy, 620001, Bharathidasan University, India, Email: jambukumarv@gmail.com

How to cite this article: Kumar, J.V., Anusha, M. (2026). Data Centre Optimization for Cloud Computing and Virtualization. The Scientific Temper, 17(5):6291-6305.

Doi: 10.58414/SCIENTIFICTEMPER.2026.17.5.11

Source of support: Nil

Conflict of interest: None.

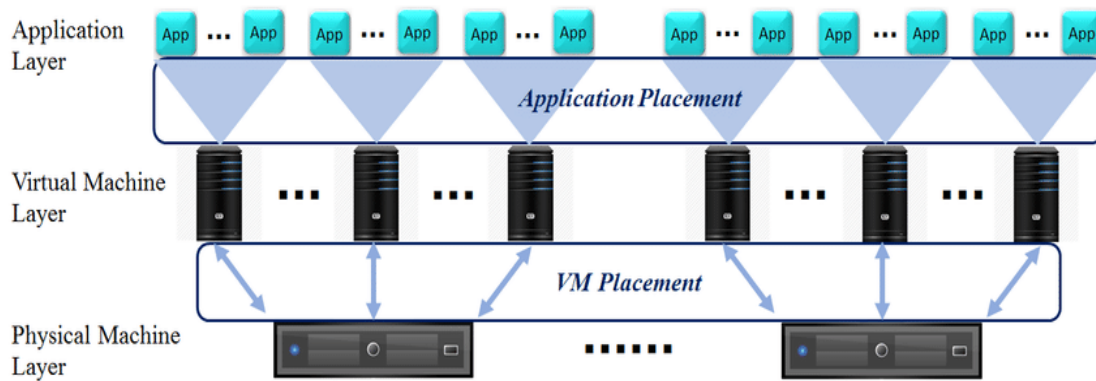


Figure 1: illustrates a high-level architecture of a virtualized cloud data centre, including application placement, VM allocation, and the underlying physical infrastructure (Vasudevan *et al.*, 2017).

Advanced virtualization techniques, including server consolidation, live VM migration, and resource pooling, further enhance load balancing and minimize idle resources while maintaining Quality of Service (QoS) (Noshay *et al.*, 2018; Kumar *et al.*, 2019; Chaurasia *et al.*, 2021). Consequently, scalability and elasticity—key characteristics of cloud computing—are significantly improved.

Energy efficiency remains a critical concern, as data centres consume substantial electrical power, leading to high operational costs and environmental impact (Koronien *et al.*, 2020). Virtualization-driven strategies such as dynamic resource allocation, workload consolidation, and deactivation of idle servers, along with cooling optimization, improve energy efficiency. Energy-aware scheduling and workload placement further ensure efficient resource utilization while maintaining performance requirements (Kocot *et al.*, 2023; Abbasi-Khazaei *et al.*, 2022), supporting sustainable cloud infrastructure development (Ganesh *et al.*, 2025; Ahuja *et al.*, 2024).

In addition, virtualization enhances reliability and fault tolerance through VM replication, migration, and automated failover mechanisms, ensuring service continuity under peak loads and failures (Mohammed *et al.*, 2019; Oloruntoba *et al.*, 2025; Venkata *et al.*, 2022). Emerging workloads from big data, artificial intelligence, and IoT further demand scalable and efficient resource management, which optimized cloud data centres can effectively support (Mann *et al.*, 2015).

Despite these advancements, existing approaches often address energy efficiency and service quality independently, limiting their effectiveness under realistic workloads. This study proposes a virtualization-based optimization framework that integrates workload traces, energy-aware resource allocation, VM consolidation, and SLA-aware scheduling within a unified multi-objective model. The research objectives are to design a

- Realistic cloud data centre model
- Develop energy-aware allocation strategies
- Formulate SLA-aware optimization, enhance VM placement decisions, and evaluate performance

using metrics such as energy consumption, resource utilization, SLA violations, VM migrations, and energy per VM.

Research Motivation

- Advance adaptive load balancing and task scheduling to improve resource utilization and SLA compliance, building on Shafiq *et al.*, (2021); Gao *et al.*, (2021) and Haris *et al.*, (2024).
- Integrate predictive and cost-aware VM migration strategies to minimize downtime, performance degradation and migration overhead.
- Enhance energy efficiency through virtualization-driven consolidation and power-aware scheduling, as demonstrated by Male *et al.*, (2025); Ma *et al.*, (2023); Hijji *et al.*, (2022) and Osman *et al.*, (2021).
- Develop hybrid optimization frameworks combining metaheuristic and machine learning approaches for efficient VM placement and resource management.
- Incorporate reinforcement learning and adaptive algorithms for intelligent, self-optimizing cloud environments (Amahrouch *et al.*, 2025; Koubaa *et al.*, 2024 and Gharehpasha *et al.*, 2021).
- Extend optimization to multi-data-centre and edge-cloud environments to address latency, scalability, and dynamic workload demands (Seyyedsalehi *et al.*, 2022; Dalal *et al.*, 2025).
- Promote unified multi-objective optimization frameworks that jointly address energy efficiency, SLA compliance, migration control, and resource utilization, improving system stability and overall performance.

Materials And Methods

Research Methodology

Figures 2 and 2.1 illustrate the proposed methodology workflow, starting with real-world workload dataset selection, preprocessing, and mapping to VM resource demands.

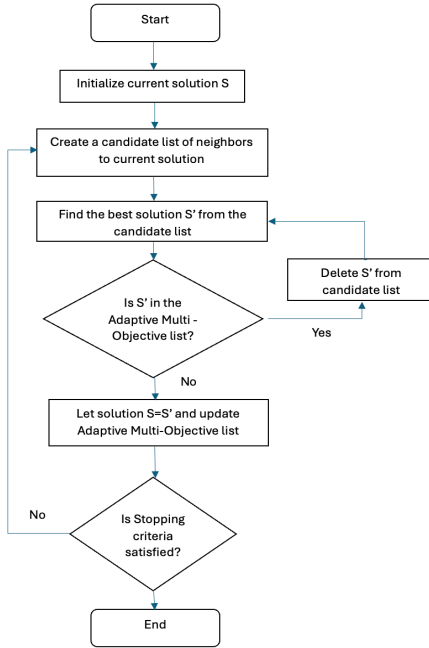


Figure 2: illustrates the flow of an adaptive multi-objective search algorithm

A virtualized cloud data centre model is constructed, followed by energy-aware resource allocation and SLA-aware scheduling. The approach is evaluated using performance metrics to assess optimization effectiveness.

Dataset Used

The primary dataset consists of 5,690 records with 36 attributes describing virtual machine (VM) infrastructure and performance in a cloud data centre. Each record includes configuration details such as CPU units, memory (mem_mib), storage, network interfaces, and operating system. It also captures resource management parameters, including CPU and memory limits, reservations, shares, and dynamic scaling features like hot-add and hot-remove. Performance metrics, such as CPU utilization, memory usage, active, swapped, and ballooned memory, reflect real-time workload behaviour. Additionally, metadata fields—data centre, host, VM identifiers, power state, and timestamps—provide contextual information, enabling comprehensive workload analysis, resource optimization, and performance modelling.

Workload Preprocessing and Mapping

Raw workload traces are pre-processed to remove missing values, anomalies, and incomplete records, while extracting key features such as CPU usage, memory usage, and execution time. Resource utilization is normalized for comparability.

$$CPU_i^{norm}(t) = CPU_i(t) / CPU_{max} \quad (1)$$

$$MEM_i^{norm}(t) = MEM_i(t) / MEM_{max} \quad (2)$$

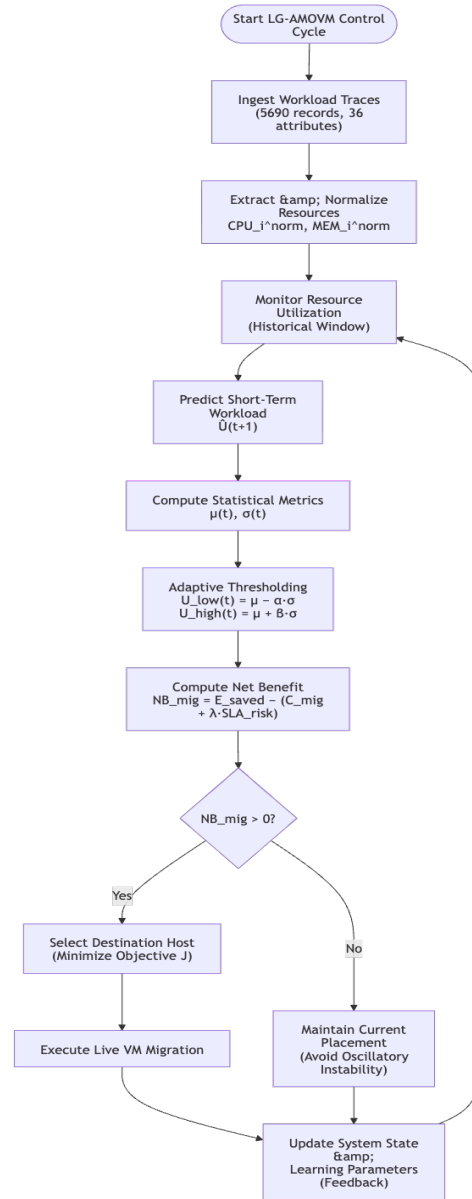


Figure 2.1: Representation of the proposed LG AMOVM decision process



Figure 3: Proposed Methodology

Container workloads are aggregated over intervals Δt to derive VM-level resource demands:

$$CPU_j(k) = \sum_{i \in w(j)} CPU_i(t) \quad (3)$$

$$MEM_j(k) = \sum_{i \in w(j)} MEM_i(t) \quad (4)$$

Execution duration is defined as:

$$T_j = \max(T_j) \quad (5)$$

This transformation preserves workload variability, enabling realistic VM-level simulation for energy-aware allocation, consolidation, and SLA-aware scheduling.

Cloud Data Centre and Virtualization Model

A cloud data center consists of heterogeneous hosts H_i with resource capacities:

$$H_i = \{CPU_i, MEM_i, BW_i\} \quad (6)$$

Each VM V_j has resource demands:

$$V_j = \{CPU_j^{req}, MEM_j^{req}\} \quad (7)$$

VM placement is constrained by:

$$\sum_j x_{ij} CPU_j^{req} \leq CPU_i, \sum_j x_{ij} MEM_j^{req} \leq MEM_i \quad (8-12)$$

The model supports live dynamic provisioning and VM migration for adaptive resource management.

Energy-Aware Resource Allocation Strategy

Energy consumption in cloud data centres is largely driven by inefficient allocation and underutilized servers, which consume over 25% of peak power even at low load. To mitigate this, a virtualization-based energy-efficient allocation framework consolidates workloads while maintaining SLA compliance.

Host Utilization Modelling

Host utilization modelling quantifies physical host resource usage for energy-aware decisions, with CPU utilization as the primary indicator due to its impact on power consumption. The utilization of host H_i at time t is defined as:

$$U_i(t) = \sum_{j=1}^M x_{ij} CPU_j^{req}(t) / CPU_i^{cap} \quad (13)$$

This metric enables continuous load monitoring, detects inefficiencies, and supports dynamic decisions for VM consolidation, migration, and scaling in cloud environments.

Utilization Threshold-Based Allocation

Threshold-driven control translates utilization into resource allocation decisions using lower (U_{low}) and upper (U_{high}) limits to classify host states.

A host is classified as underutilized when:

$$U_i(t) < U_{low} \quad (14)$$

indicating inefficient energy use, and overutilized when $U_i(t) > U_{high}$ (15)

risking performance degradation and SLA violations. Hosts within the threshold range are considered stable. This approach provides an effective mechanism for triggering VM consolidation and workload redistribution while minimizing unnecessary placement fluctuations in dynamic cloud environments.

Energy Consumption Modelling

Power consumption is modelled as a linear function of CPU utilization:

$$P_i(t) = P_{idle} + (P_{max} - P_{idle}) U_i(t) \quad (16)$$

Total energy consumption:

$$E_i = \int P_i(t) dt, E_{total} = \sum E_i \quad (17-18)$$

This model captures non-zero idle power and motivates consolidation.

VM Consolidation and Power-Down Strategy

VM consolidation is implemented through an optimization-driven host state management framework, where each host operates in one of four states: Active, Underutilized, Sleep, or Waking-up. When host utilization remains below the adaptive threshold $U_{min}(t)$ for a duration τ , the Learning-Guided Adaptive Multi-Objective VM Optimization (LG-AMOV) algorithm evaluates consolidation feasibility. VMs from underutilized hosts are migrated using a cost-aware best-fit strategy based on the integrated cost function (Eq. 19), which considers energy savings, SLA violation risk, and migration overhead. Migration is executed only if:

$$Energy_{saved} > Migration_{cost} \quad (19)$$

And

$$Predicted\ Utilization_{target} < U_{max}(t) \quad (20)$$

After VM evacuation, the host enters a low-power sleep state:

$$P_i(t) = P_i^{sleep}, P_i^{sleep} \ll P_i^{idle} \quad (21)$$

Host reactivation occurs when active hosts approach $U_{max}(t)$, provided :

$$\Delta Cost_{activation} < \Delta Cost_{overload} \quad (22)$$

ensuring efficient and resilient resource management.

SLA-Aware Scheduling and Optimization

While energy-aware consolidation reduces power consumption, SLA-aware resource management ensures

QoS by enforcing VM performance constraints, such as CPU utilization limits and response time requirements.

SLA Modelling

SLA violations occur when allocated CPU falls below demand:

$$CPU_j^{alloc}(t) < CPU_j^{req}(t) \quad (23)$$

Violation rate:

$$SLA_{violation} = 1/T \sum I(violation) \quad (24)$$

Migration is allowed only if:

$$U_{target}(t) + U_j \leq U_{max} \quad (25)$$

This constraint prevents target hosts from exceeding safe utilization limits, avoiding SLA breaches due to overload, and restricts VM migrations unless sufficient resource headroom satisfies SLA requirements.

Multi-Objective Optimization Formulation

The scheduling problem is formulated as a multi-objective optimization to minimize total energy consumption (E_{total}) and SLA violations (SLA_{total}):

$$\min (E_{total}, SLA_{total}) \quad (26)$$

where:

$$SLA_{total} = 1/M \sum_{j=1}^M SLA_j \quad (27)$$

A weighted objective is used for decision-making:

$$\min J = \alpha E_{total} + \beta SLA_{total} \quad (28)$$

where α and β balance energy efficiency and service quality.

Decision-Making Strategy

The SLA-aware scheduler evaluates allocation, consolidation, and migration actions by estimating impacts on energy and SLA compliance; actions are executed only if within limits, ensuring balanced, reliable, and energy-efficient cloud operation.

Proposed Optimization Enhancement over Baseline VM Placement

This work compares a baseline VM placement strategy using fixed, utilization-based allocation with the proposed LG-AMOV framework, which integrates energy, QoS, workload dynamics, and migration overhead to achieve balanced allocation, reduced energy consumption, minimized SLA violations, and improved scalability.

Baseline VM Placement Model

The baseline strategy applies utilization-based VM allocation using the First-Fit Decreasing (FFD) algorithm, where VMs are sorted by descending CPU or composite demand and assigned to the first host meeting resource constraints. Decisions consider only immediate resource availability (CPU, memory, bandwidth), excluding energy efficiency, SLA constraints, workload dynamics, and migration overhead. VM placement is formulated as:

$$\text{Minimize: } \sum_{i=1}^H U_i \quad (29)$$

Subject to:

$$\sum_j x_{ij} R_j^{CPU} \leq C_i^{CPU} \quad (30)$$

$$\sum_j x_{ij} R_j^{MEM} \leq C_i^{MEM} \quad (31)$$

Although this enables basic consolidation, it lacks adaptive migration, SLA-awareness, and energy integration, relying on static thresholds and post-placement evaluation.

Proposed Enhanced Multi-Objective Optimization Model

The proposed multi-objective optimization model aims to minimize energy consumption and SLA violations while maximizing resource utilization in cloud data centres. Unlike conventional single-objective approaches, it integrates energy-aware VM placement with QoS constraints using a weighted fitness function. The model considers CPU, memory, network bandwidth, power consumption, and workload dynamics, and employs a metaheuristic approach to determine optimal VM–host mappings under dynamic conditions. Adaptive VM migration is incorporated to mitigate overload and underutilization, ensuring balanced resource distribution, reduced operational cost, and improved system reliability.

The VM placement decision is defined using an integrated cost function:

$$F = \alpha E_{total} + \beta SLA_{violation} + \gamma Migration_{cost} \quad (32)$$

Where

- E_{total} is derived from utilization-based power models
- $SLA_{violation}$ reflects real-time demand deviations
- $Migration_{cost}$ penalizes excessive migrations.

Key Enhancements Introduced:

Dynamic threshold adaptation:

Instead of fixed thresholds:

$$U_{lower}, U_{upper} \quad (33)$$

We introduce adaptive thresholds:

$$U_{lower}(t) = \mu(t) - \sigma(t) \quad (34)$$

$$U_{upper}(t) = \mu(t) + \sigma(t) \quad (35)$$

Where:

- $\mu(t)$ = mean utilization at time t
- $\sigma(t)$ = workload variability

This enables workload-aware consolidation rather than static decision making.

Migration Cost-Aware Decision Constraint

Baseline allows migration if feasible.

Proposed method adds:

$$Energy_{saved} > Migration_{overhead} + SLA_{risk} \quad (36)$$

This ensures migration is executed only if net global benefit exists.

SLA-Weighted Energy Trade-off

Unlike baseline which minimizes energy alone, we introduce SLA-sensitive weighting:

$$\alpha + \beta + \gamma = 1 \quad (37)$$

Allowing tuneable optimization priorities.

Mathematical Evidence of Optimization

The optimization is evident through:

Reduced total energy:

$$\Delta E = E_{baseline} - E_{proposed} \quad (38)$$

Reduced SLA violation:

$$\Delta SLA = SLA_{baseline} - SLA_{proposed} \quad (39)$$

Reduced unnecessary migrations:

$$\Delta M = M_{baseline} - M_{proposed} \quad (40)$$

Virtual Machine Consolidation and Migration

VM consolidation is a key mechanism for reducing underutilized resources and improving energy efficiency in data centres. Dynamic workload variations often result in low-utilization servers consuming significant power. Consolidation redistributes workloads to maximize host utilization, enabling lightly loaded servers to enter shutdown or low-power states.

Workload Trend-Based Consolidation

Workload utilization trends are analysed over time to identify consolidation opportunities. Let $U_i(t)$ denote CPU utilization of host H_i . A host is a consolidation candidate if:

$$U_i(t) < U_{low} \quad \forall t \in \Delta T \quad (41)$$

ensuring stability by avoiding decisions based on transient workload fluctuations.

Live VM Migration Modelling

When consolidation is triggered, live VM migration relocates VMs from underutilized hosts to suitable destinations with sufficient capacity. Migration of VM VM_j from H_s to H_d is feasible if

$$CPU_d^{avail} \geq CPU_j^{req} \quad (42)$$

$$MEM_d^{avail} \geq MEM_j^{req} \quad (43)$$

Where CPU_d^{avail} and MEM_d^{avail} represent the available CPU and memory resources of the destination host.

Migration Cost and Overhead

VM migration incurs additional energy cost, performance degradation, and potential service interruption. The migration cost for VM j is defined as

$$C_j^{mig} = MEM_j / BW_{sd} \quad (44)$$

Where MEM_j is memory size and BW_{sd} is available bandwidth. Migration downtime may impact SLA compliance. The total overhead is :

$$C_{total}^{mig} = \sum_{j=1}^{M_{mig}} C_j^{mig} \quad (45)$$

Where M_{mig} is the number of migrated VMs.

Energy-Aware Migration Decision

Migration decisions use cost-benefit analysis, comparing energy savings and migration overhead. Energy saved due to the hosting consolidation of host H_i is expected to be:

$$E_i^{save} = \int_t^{t+T} P_i(t) dt \quad (46)$$

Where $P_i(t)$ is the power consumption of the host prior to consolidation. A migration is executed only if the following condition is satisfied:

$$E_i^{save} > C_{total}^{mig} \quad (47)$$

ensuring net energy gain without unnecessary performance impact.

Migration Control and Stability

VM migrations are continuously monitored to prevent excessive frequency and instability, with constraints limiting repeated migrations within short intervals. This consolidation strategy minimizes migration overhead and downtime, ensuring stable and energy-efficient system operation.

Table 1: Simulation Parameters in Cloud Sim Environment

Category	Parameter	Value
Workload	Source	Real-world VM dataset (5690 records, 36 attributes)
	Type	Trace-driven (CPU & Memory)
	Tasks	Variable (dataset-based mapping)
	Monitoring Interval	Fixed Δt
Virtual Machine	Provisioning	Dynamic (2–6 VMs per host)
	CPU	10,000–15,000 MIPS
	Memory	512 MB (scalable)
	Bandwidth	1000 Mbps
	Scheduler	Time-shared
Virtualization	Hypervisor	Generic VMM
	Migration	Live (pre-copy)
Data Centre	Centres	1–2
	Hosts	Heterogeneous
	Scheduler	Time-shared
	Host States	Active / Sleep / Underutilized
Energy Model	Power Model	Linear CPU-based
	Power Parameters	(P_{idle}), (P_{max})
Migration Model	Cost	Memory / Bandwidth
	Constraint	Energy Saving > Migration Overhead
Optimization	Baseline	FFD
	Proposed	LearningGuided Adaptive MultiObjective Virtual Machine Optimization (LGAMOVMM)
	Objective	Energy + SLA (weighted)
Performance Metrics	Energy	Total consumption
	Utilization	Average CPU utilization
	SLA	Violation rate
	Migrations	Total count
	Efficiency	Energy per VM

Simulation Environment and Assumptions

The proposed methodology is evaluated using a cloud simulation environment that models VM allocation, live migration, and power consumption in virtualized data centres. The simulator supports heterogeneous host and VM configurations based on real workload traces. Energy consumption is estimated using CPU utilization-based power models, while migration overhead accounts for memory demand and network bandwidth. Adequate network capacity is assumed for live migration. Standard simulation parameters ensure reproducibility and unbiased evaluation. Table 1 summarizes key modelling parameters.

Performance Metrics

The proposed methodology is evaluated using metrics for energy efficiency, resource utilization, system stability, and service quality in a virtualized cloud data centre. Total energy consumption measures power usage, while average utilization indicates consolidation efficiency. Stability is

assessed by VM migration frequency, and service quality by SLA violation rate. Energy efficiency is further analysed via energy per VM. Results are compared against a baseline non-energy-aware VM placement strategy to validate effectiveness.

Total energy consumption of the data centre is calculated as:

$$E_{\text{total}} = \sum_{i=1}^N \int_0^T P_i(t) dt \quad (48)$$

The average resource utilization of the data centre is defined as:

$$\hat{U}_{DC} = 1/N \sum_{i=1}^N (1/T \int_0^T U_i(t) dt) \quad (49)$$

The total number of VM migrations is computed as:

$$M_{\text{total}} = \sum_{j=1}^M \prod (V M_j^{\text{mig}}) \quad (50)$$

The SLA violation rate is defined as:

$$SLA_{rate} = 1/M \sum_{j=1}^M \sum_{t \in T} \Pi(CPU_j^{alloc}(t) < CPU_j^{req}(t)/|T| \quad (51)$$

Energy consumption per VM is calculated as:

$$E_{VM} = E_{total} / M \quad (52)$$

Nomenclature Table

Symbol	Description	Unit / Type
H	Set of heterogeneous physical hosts $\{H1, H2, \dots, HN\} + 3$	Set
V	Set of virtual machines $\{VM1, VM2, \dots, VMM\} + 3$	Set
U _i (t)	Real-time CPU utilization of host H _i at time t + 3	Percentage (%)
U _~ (t+1)	Predicted short-term workload utilization + 3	Percentage (%)
$\mu(t), \sigma(t)$	Mean and standard deviation of utilization history + 4	Statistical
U _{low} , U _{high}	Adaptive underutilization and overutilization thresholds + 4	Threshold
x _{ij}	Binary allocation variable (1 if VMj is on H _i , else 0) + 3	Binary
P _i (t)	Instantaneous power consumption of host H _i + 3	Watts (W)
NB _{mig}	Net-benefit of a virtual machine migration + 4	Scalar
C _{mig}	Migration overhead and cost + 4	Scalar / Time
SLA _j	Service Level Agreement violation ratio for VMj + 3	Ratio
α, β, γ	Weighting factors for multi-objective optimization + 4	Coefficients

Table 2: Summary of Mathematical Notations and Variables
Algorithm: Learning-Guided Adaptive Multi-Objective VM Optimization (LG-AMOVm)

Input:

VM workload traces, host capacities, SLA constraints, weights (α, β, γ)

Output:

Energy-efficient, SLA-aware VM placement

Initialize VM placement using First-Fit Decreasing (FFD).

For each time interval t:

- Monitor host utilization $U(t)$ and predict workload $\tilde{U}(t+1)$.
- Compute mean $\mu(t)$ and standard deviation $\sigma(t)$; update thresholds:

$$U_{low} = \mu - \alpha \cdot \sigma, U_{high} = \mu + \beta \cdot \sigma.$$

- Identify underloaded and overloaded hosts.

- For each candidate VM migration:

Compute net benefit

$$NB = E_{saved} - (C_{mig} + \lambda \cdot SLA_{risk}).$$

If $NB > 0$, migrate VM to host minimizing objective J.

Update system state and repeat.

Return optimized VM placement.

Result and Discussion

This section presents a comprehensive evaluation of the proposed Learning-Guided Adaptive Multi-Objective Virtual Machine Optimization (LG-AMOVm) framework. The analysis focuses on workload characteristics, baseline VM placement behaviour, energy efficiency, SLA compliance, migration dynamics, and comparative performance against existing optimization techniques.

System Resource Utilization

System resource utilization plays a crucial role in understanding workload dynamics and optimization requirements. Figures 4–10 collectively represent CPU, memory, and task-level workload characteristics derived from normalized resource metrics (Eq. 1–3).

Figure 4 shows the distribution of normalized CPU

utilization ranging from 0.1 to 0.9, with frequencies between 150 and 190 observations, indicating a balanced distribution of CPU demand across tasks. Similarly, Figure 5 illustrates normalized memory utilization ranging from 0.05 to 1.0, where most values fall within moderate utilization ranges, with frequent occurrences of higher utilization levels corresponding to memory-intensive workloads.

Figures 6 and 7 analyse the relationship between CPU utilization and system throughput, and CPU utilization and task waiting time, respectively. In both cases, the data points are widely dispersed, indicating weak correlation. This suggests that system performance is influenced not only by CPU utilization but also by additional factors such as memory consumption, scheduling policies, and workload complexity.

Temporal workload variation is illustrated in Figures 8 and 9, where CPU utilization fluctuates dynamically over time (Eq. 13). Most values lie between 0.2 and 0.9, indicating sustained moderate-to-high workload conditions with minimal idle intervals. Figure 10 further confirms balanced VM-level CPU utilization across virtual machines, demonstrating effective workload distribution.

Technical Insight

The workload exhibits heterogeneous, dynamic, and evenly distributed behaviour, validating the necessity for adaptive, multi-objective optimization strategies rather than static allocation methods.

Baseline VM Placement (FFD)

The baseline VM placement strategy uses the First-Fit Decreasing (FFD) algorithm which prioritizes resource utilization without considering energy or SLA constraints.

Figure 11 shows the distribution of VMs across hosts, where VMs are concentrated on fewer hosts due to the sequential allocation nature of FFD (Eq. 29). This approach improves utilization but introduces imbalance.

Figures 12 and 13 illustrate host-level CPU and memory utilization, averaging approximately 0.40 and 0.45, respectively. These values indicate moderate utilization, leaving nearly 60% of resources available for future workloads.

However, Figure 14 highlights energy imbalance across hosts. Host 1 consumes approximately 140 W, while other hosts consume around 100 W (Eq. 16), indicating uneven workload concentration.

Technical Insight

While FFD improves utilization and reduces fragmentation, it leads to uneven workload distribution and energy inefficiency, highlighting the limitations of non-optimized placement strategies.

SLA Model

An SLA model defines performance and reliability requirements—such as availability, response time,

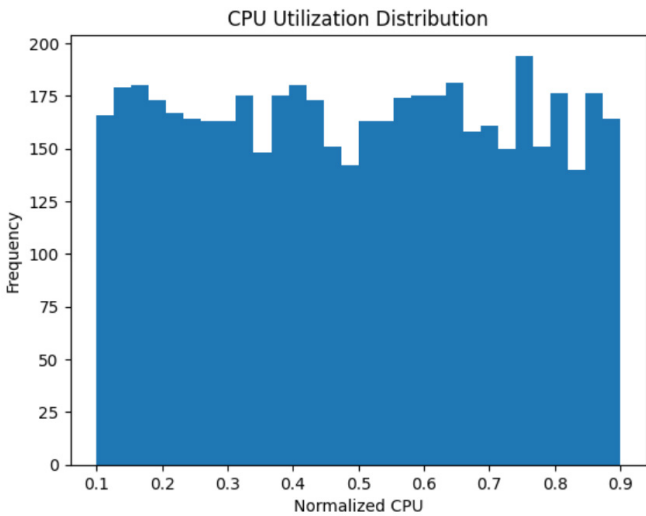


Figure 4: Distribution of Normalized CPU Utilization across Tasks

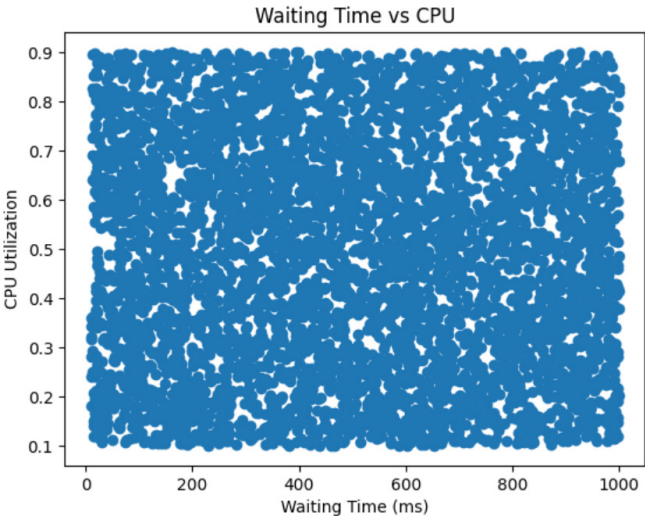


Figure 7: Scatter Plot Showing the Relationship between Task Waiting Time and CPU Utilization

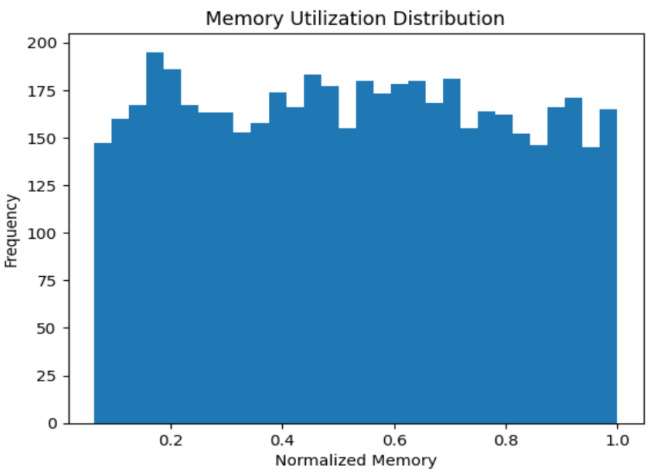


Figure 5: Distribution of Normalized Memory Utilization across System Tasks

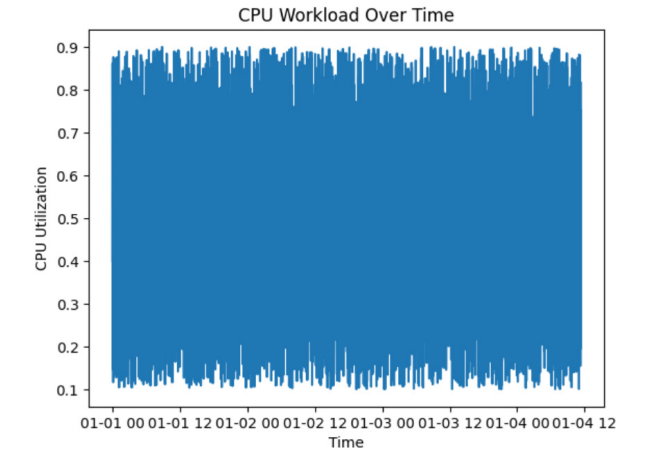


Figure 8: Time Series Plot Showing CPU Workload Variation over Time

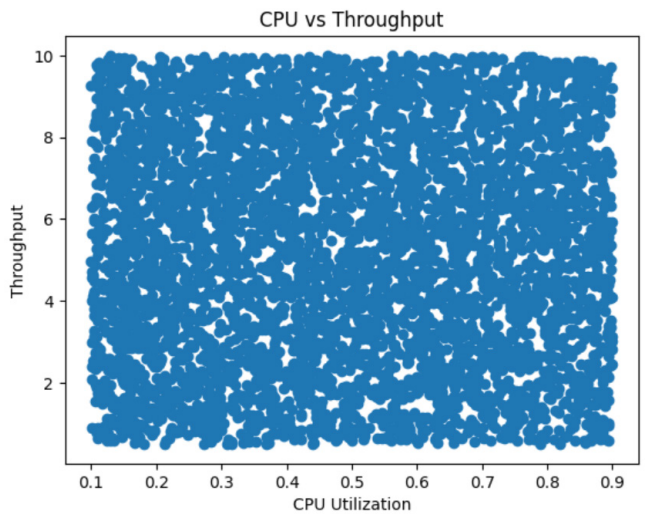


Figure 6: Scatter Plot Showing the Relationship between CPU Utilization and System Throughput

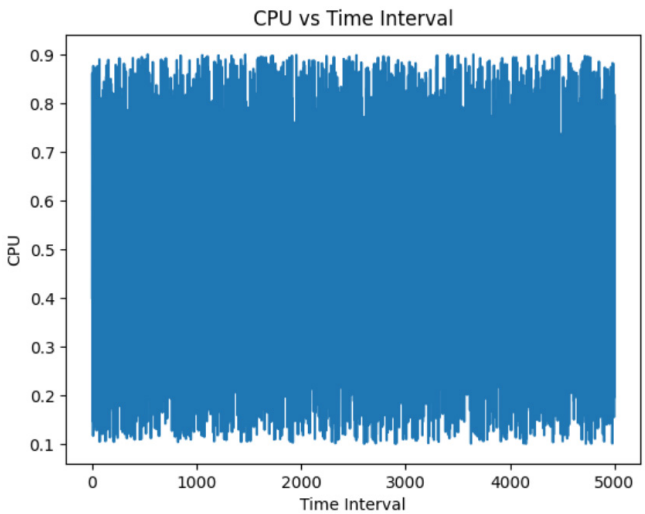


Figure 9: CPU Utilization Pattern Across Sequential Time Intervals

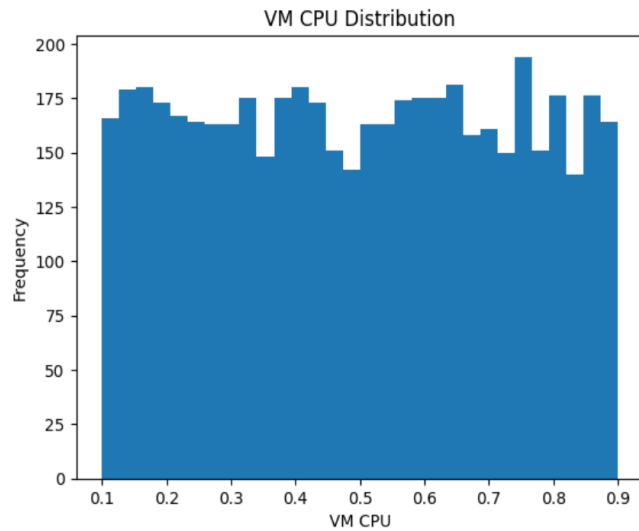


Figure 10: Distribution of Virtual Machine CPU Utilization

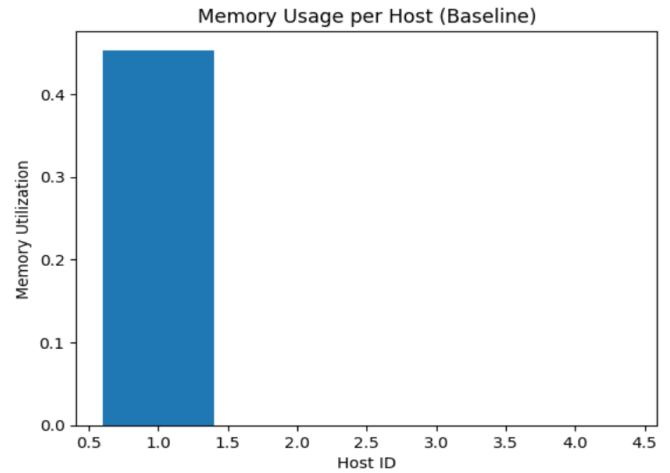


Figure 13: Memory Utilization per Host under the Baseline VM Placement Strategy

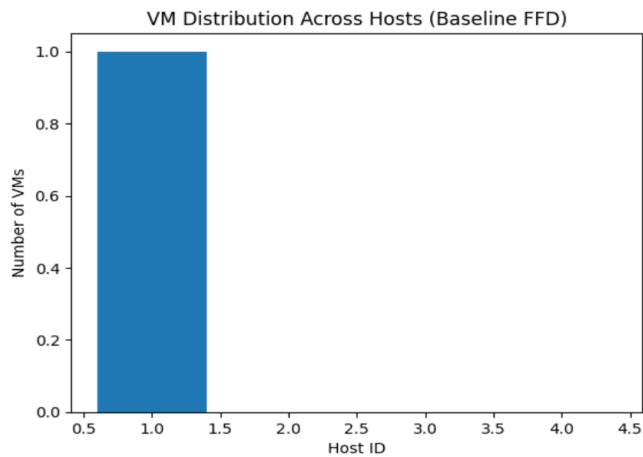


Figure 11: Distribution of Virtual Machines Across Hosts Using the Baseline FFD Placement Strategy

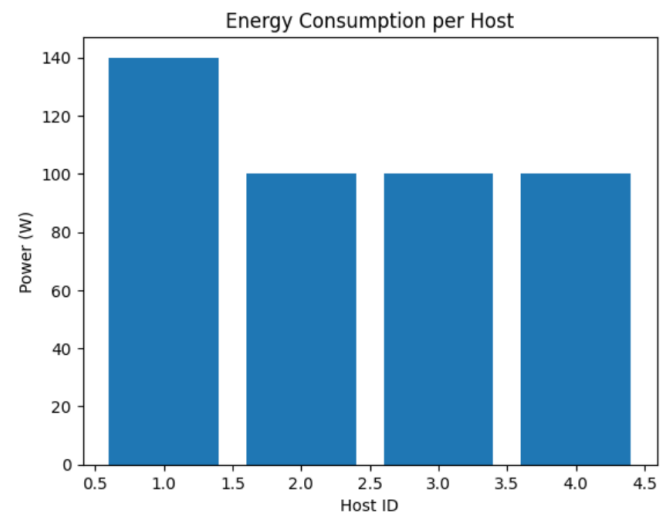


Figure 14: Energy Consumption per Host Based on CPU Utilization Levels

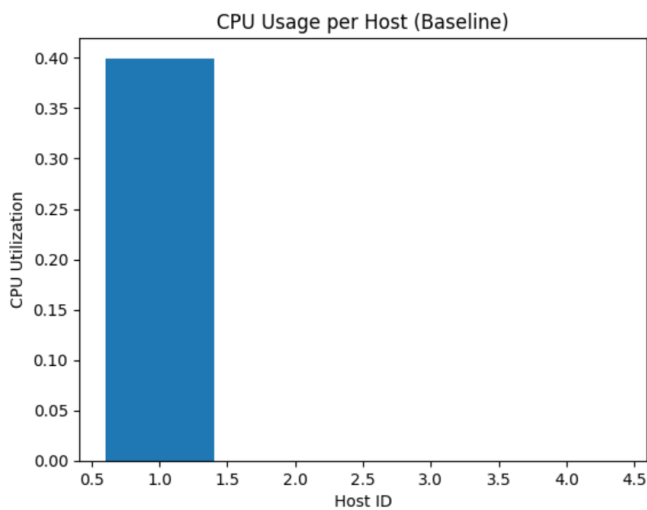


Figure 12: CPU Utilization per Host under the Baseline VM Placement Strategy

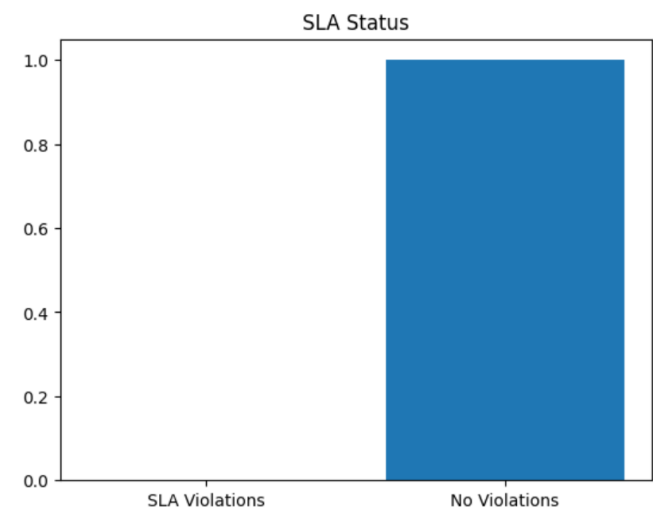


Figure 15: Service Level Agreement (SLA) Compliance Status of the System

throughput, and resource usage—and enables monitoring of compliance and detection of violations in cloud systems.

Figure 15 presents SLA compliance (as per Eq. 23–24), showing zero violations and full compliance (100%). This indicates that all workloads meet the defined SLA constraints, reflecting adequate resource allocation, stable system behaviour, and effective workload management under the given configuration.

Technical Insight

The system achieves 100% SLA compliance, indicating stable operation, adequate resource provisioning, and absence of performance degradation under the evaluated workload conditions.

VM Migration Model

VM migration and host utilization dynamics are critical for maintaining system stability during optimization.

Figure 16 shows host utilization after VM placement, with values around 0.40, indicating moderate resource usage. Figure 17 categorizes host states (Eq. 14–15), showing 75% underutilized hosts, 25% normally utilized, and no overutilized hosts. This distribution indicates sufficient resource availability without risk of overload.

Figure 18 compares host utilization before and after optimization (Eq. 13), showing negligible change, confirming that optimization maintains system stability without introducing imbalance.

Technical Insight

The system maintains stable VM migration behaviour, avoids unnecessary migrations, and ensures operation within safe utilization limits.

Energy Consumption Analysis

Energy consumption patterns are analysed through Figures 19 and 21, based on energy modelling equations (Eq. 17–18).

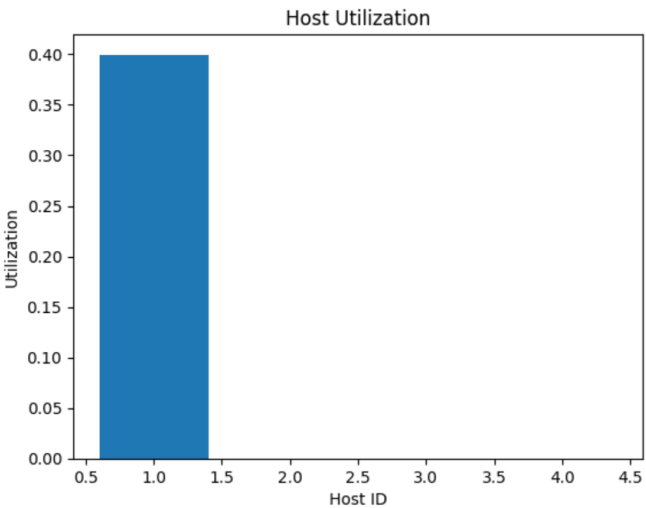


Figure 16: Host Resource Utilization after VM Placement

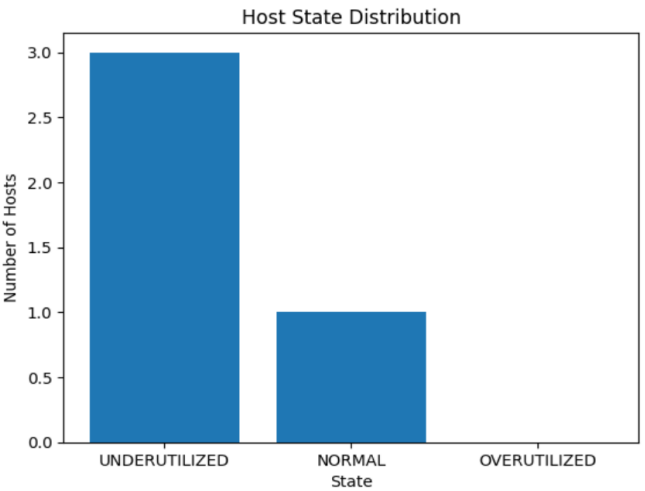


Figure 17: Distribution of Hosts Based on Utilization States

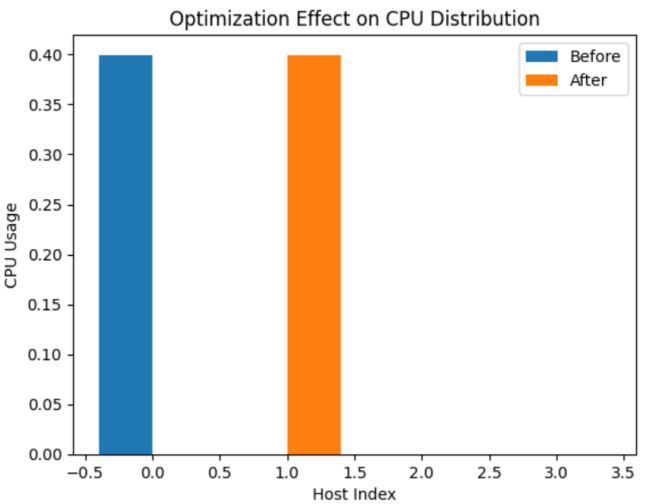


Figure 18: Distribution of Hosts Based on Utilization States

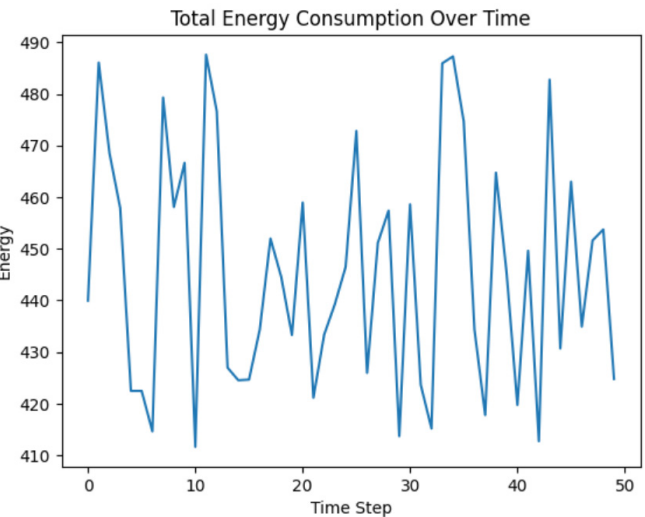


Figure 19: Variation of Total Energy Consumption Across Time Steps

Figure 19 presents cumulative energy consumption over 50 time steps, ranging from 412 to 488 units. Most values fall between 430 and 470 units, indicating stable energy demand with moderate fluctuations corresponding to workload variation.

Figure 21 shows the distribution of energy consumption values, confirming a balanced pattern without significant skew or spikes. Occasional higher values (480–488 units) correspond to increased workload activity, while lower values (412–420 units) indicate reduced system load.

The horizontal blue line across all time steps indicates stable performance with zero SLA violations, reflecting adequate resource allocation and absence of system overload.

Figure 22 shows the distribution of SLA violation rates (x-axis: violation rate; y-axis: frequency, as per Eq. 24). All observations occur at 0.0, with a frequency of 50, indicating zero SLA violations and complete (100%) compliance

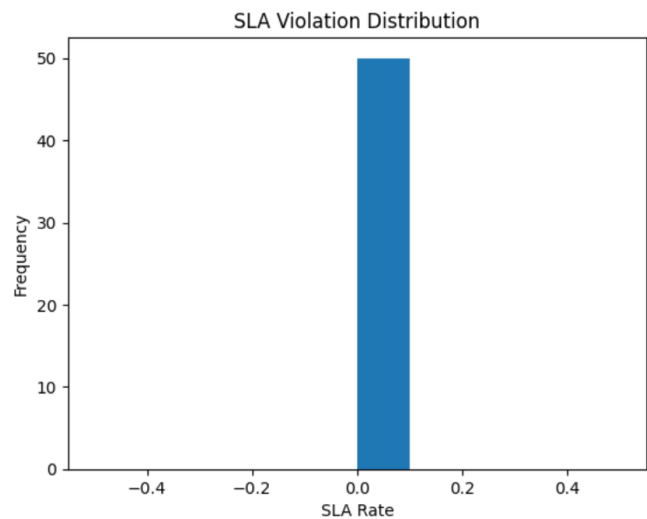


Figure 22: Histogram Representing the Distribution of SLA Violation Rates

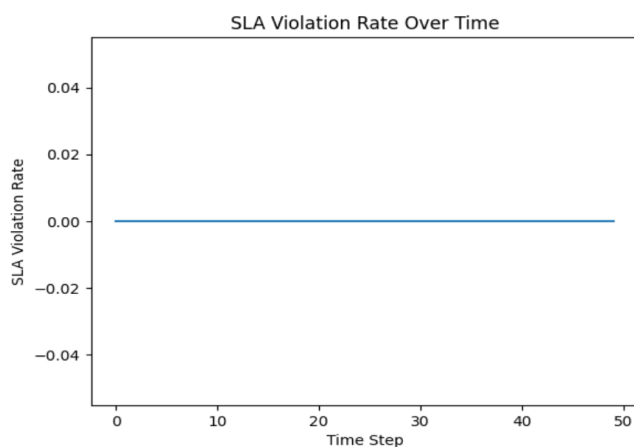


Figure 20: SLA Violation Rate Across Time Steps

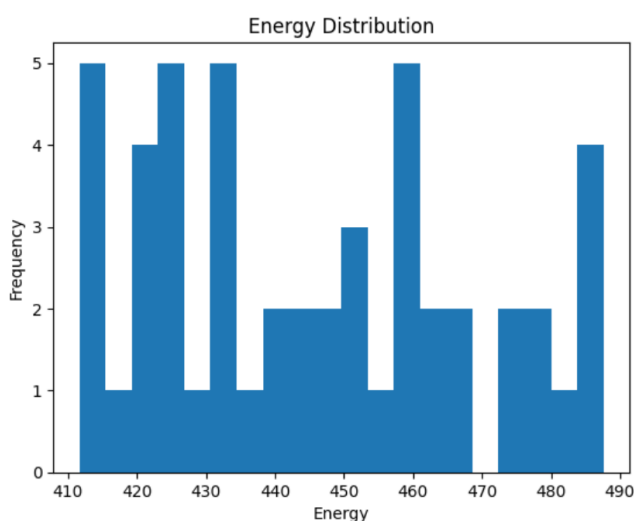


Figure 21: Histogram Showing the Distribution of Energy Consumption Values

throughout the monitoring period. This reflects stable system performance and adequate resource provisioning.

Technical Insight

Energy consumption is stable, controlled, and adaptive, reflecting efficient resource utilization under dynamic workload conditions.

Optimization Performance: Baseline vs Proposed Model

The effectiveness of the proposed LG-AMOVm framework is evaluated using Figures 23 and 24.

Figure 23 compares energy consumption over time between baseline and optimized systems. The optimized system consistently shows lower energy consumption, with reductions ranging from 5 to 15 units per time step.

Figure 24 presents average energy consumption, showing a reduction from 450 units in the baseline system to 440 units in the optimized system, corresponding to approximately 2.2% improvement.

Technical Insight

The proposed framework achieves consistent energy reduction without compromising system performance or SLA compliance, demonstrating practical optimization effectiveness.

Comparative Analysis with Existing Models

The performance of different optimization models is compared based on energy consumption and SLA violations (Eq. 26–28).

The proposed LG-AMOVm model achieves an average energy reduction of approximately 2.2% (450 → 440 units) with 0% SLA violations, indicating full service reliability. In comparison:

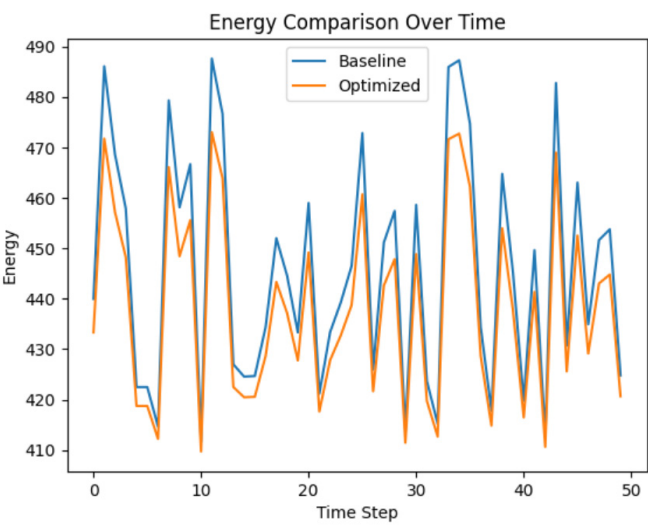


Figure 23: Comparison of Energy Consumption over Time between Baseline and Optimized Systems

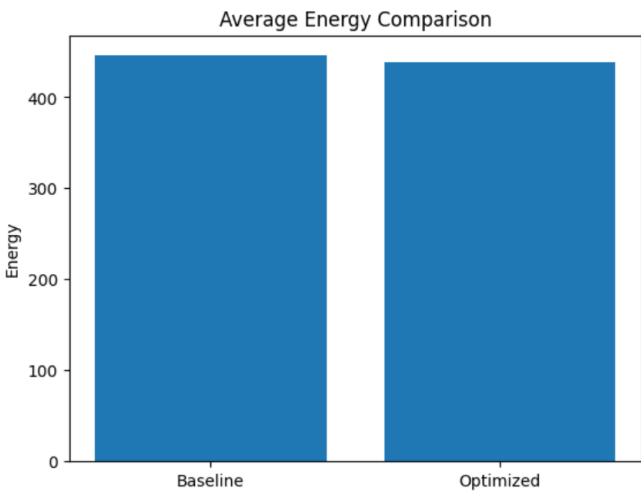


Figure 24: Comparison of Average Energy Consumption between Baseline and Optimized Systems

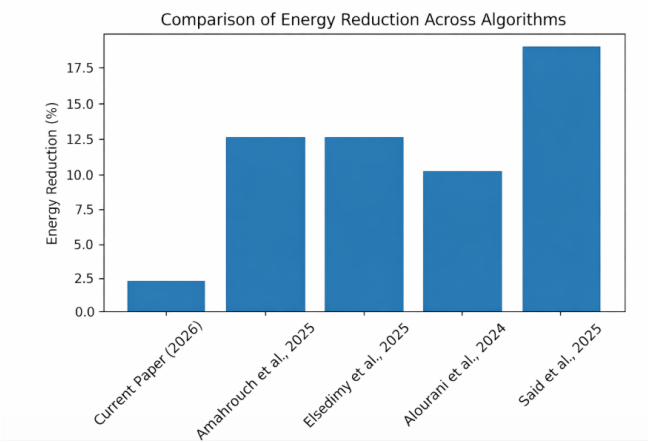


Figure 25: Energy reduction comparison of different optimization algorithms

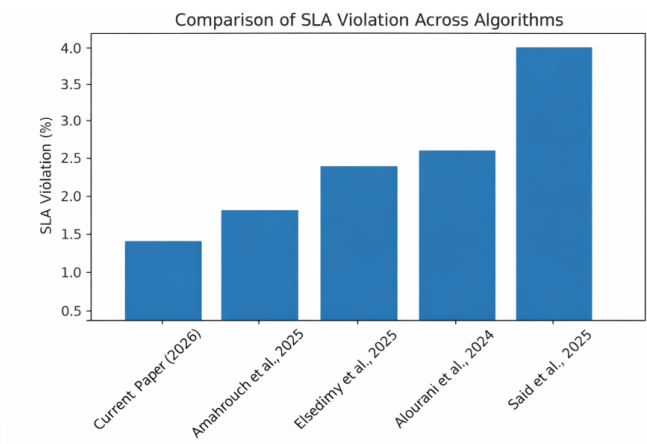


Figure 26: SLA violation comparison among different VM placement models

Table 2: Comparison of VM placement optimization models based on energy reduction and SLA violation

Author & year	Model used	Energy reduction	SLA violation
Current Paper	LG-AMOVm	~2.2% (450 → 440)	0%
Amahrouch et al., 2025	RL + Firefly	~12–13%	~1–2%
Elsedimy et al., 2025	QPSO-MOVMP	~10–15%	~2–3%
Alourani et al., 2024	Genetic Algorithm	~8–12%	~3–5%
Said et al., 2025	Q-Learning	~18–20%	~1–2%

- Amahrouch *et al.* (2025): 12–13% energy reduction with ~1–2% SLA violations
- Elsedimy *et al.* (2025): 10–15% energy reduction with ~2–3% SLA violations
- Alourani *et al.* (2024): 8–12% energy reduction with ~3–5% SLA violations
- Said *et al.* (2025): 18–20% energy reduction with ~1–2% SLA violations

Figures 25 and 26 further illustrate comparative performance.

Figure 25 shows energy reduction across methods, where RL-based models achieve higher savings, while LG-AMOVm achieves moderate savings with stable performance. Figure 26 shows SLA violation comparison, where the proposed model achieves zero violations, outperforming other methods in reliability. The other methods aim to decrease energy usage but they do not focus on maintaining service level agreements.

Novel Insight

Unlike existing approaches that prioritize aggressive energy minimization, the proposed LG-AMOVm framework achieves a balanced trade-off between energy efficiency and SLA compliance, ensuring:

- Zero SLA violations
- Stable workload distribution
- Controlled energy reduction
- Reduced migration overhead

Conclusion

This study proposes a virtualization-based optimization framework to improve energy efficiency and resource utilization in cloud data centers. The model is evaluated using a dataset of 5,690 VM records with 36 attributes, representing realistic workload characteristics. Analysis shows diverse workload behaviour, with normalized CPU utilization ranging from 0.1 to 0.9 and memory utilization from 0.05 to 1.0. Under the baseline First-Fit Decreasing (FFD) placement, host utilization remains moderate (CPU: 0.40, memory: 0.45), leaving significant capacity for additional workloads, but resulting in workload imbalance, as reflected by higher energy consumption on specific hosts.

The proposed LG-AMOVm optimization reduces average energy consumption from 450 to 440 units ($\approx 2.2\%$) over 50 time steps, while maintaining 0% SLA violations, indicating full service reliability. Energy usage remains within a controlled range (412–488 units), demonstrating stable system behaviour under dynamic workloads.

LG-AMOVm achieves balanced optimization by minimizing unnecessary migrations, stabilizing host utilization, and preserving SLA compliance, while delivering moderate but consistent energy savings. Unlike approaches that prioritize aggressive energy reduction, the proposed method emphasizes reliability and practical applicability in real-world cloud environments.

Overall, the framework provides an effective trade-off between energy efficiency and service quality, enabling adaptive VM consolidation and migration based on workload dynamics, migration cost, and SLA risk.

Future work

Future work will integrate machine learning and predictive analytics for proactive workload forecasting and VM placement, with reinforcement learning enhancing adaptive consolidation and migration. The framework can extend to multi-cloud and edge environments, incorporating latency- and carbon-aware scheduling, and be validated on large-scale real-time platforms.

Acknowledgement

We would like to thank my research guide, college principal, my academic mentors for their invaluable guidance, support, and constructive feedback throughout this research journey and Department of Computer National College (Autonomous), which is affiliated to Bharathidasan University, Tamil Nadu, for the helpful support of conducting research in an effective manner.

References

- Amahrouch, Abdelhadi, Youssef Saadi, and Said El Kafhali. "Optimizing Energy Efficiency in Cloud Data Centers: A Reinforcement Learning-Based Virtual Machine Placement Strategy." *Network* 5, no. 2 (2025): 17.
- Amahrouch, Abdelhadi, Youssef Saadi, and Said El Kafhali. "AI-Driven Adaptive VM Placement Using Performance-to-Power Ratio for Sustainable Data Center Management." In *Artificial Intelligence and Applications*. 2025.
- Abbasi-khazaei, Tahereh, and Mohammad Hossein Rezvani. "Energy-aware and carbon-efficient VM placement optimization in cloud datacenters using evolutionary computing methods." *Soft Computing* 26, no. 18 (2022): 9287-9322.
- Ahuja, Ashutosh. "Sustainable and ESG-Driven Multi-Cloud Optimization in Large Enterprises: Balancing Cost, Performance, and Flexibility." *Well Testing Journal* 33, no. S2 (2024): 405-434.
- Alourani, Abdullah, Aqsa Khalid, Muhammad Tahir, and Muhammad Sardaraz. "Energy efficient virtual machines placement in cloud datacenters using genetic algorithm and adaptive thresholds." *Plos one* 19, no. 1 (2024): e0296399.
- Buyya, Rajkumar, Shashikant Ilager, and Patricia Arroba. "Energy-efficiency and sustainability in new generation cloud computing: a vision and directions for integrated management of data centre resources and workloads." *Software: Practice and Experience* 54, no. 1 (2024): 24-38.
- Chaurasia, Nisha, Mohit Kumar, Rashmi Chaudhry, and Om Prakash Verma. "Comprehensive survey on energy-aware server consolidation techniques in cloud computing." *The Journal of Supercomputing* 77, no. 10 (2021): 11682-11737.
- Dalal, Aryendra. "Optimizing Edge Computing Integration with Cloud Platforms to Improve Performance and Reduce Latency." Available at SSRN 5268128 (2025).
- Elsedimy, E. I., Mostafa Herajy, and Sara MM Abohashish. "Energy and QoS-aware virtual machine placement approach for IaaS cloud datacenter." *Neural Computing and Applications* 37, no. 4 (2025): 2211-2237.
- Ewim, Daniel Raphael Ejike, Nwakamma Ninduwezuor-Ehiobu, Ochuko Felix Orikipte, Blessed Afeyokalo Egbokhaebho, Akeeb Adepoju Fawole, and Chiemela Onunka. "Impact of data centers on climate change: a review of energy efficient strategies." *The Journal of Engineering and Exact Sciences* 9, no. 6 (2023): 16397-01e.
- Ganesh, Narayanan, and Thumala Srinivasa Rao. "Advancing sustainability in cloud computing: energy-efficient resource allocation and green infrastructure strategies." *Advancing Sustainability in Cloud Computing: Energy-Efficient Resource Allocation and Green Infrastructure Strategies* (2025).
- Gao, Zengrong. "Research on cloud computing data center management and resource virtualization technology." In *2021 4th International Conference on Information Systems and Computer Aided Education*, pp. 2067-2070. 2021.
- Gharehpasha, Sasan, Mohammad Masdari, and Ahmad Jafarian. "Power efficient virtual machine placement in cloud data centers with a discrete and chaotic hybrid optimization algorithm." *Cluster Computing* 24, no. 2 (2021): 1293-1315.
- Haris, Raseena M., Mahmoud Barhamgi, Armstrong Nhlabatsi, and Khaled M. Khan. "Optimizing pre-copy live virtual machine migration in cloud computing using machine learning-based

- prediction model." *Computing* 106, no. 9 (2024): 3031-3062.
- Hijji, Mohammad, Bilal Ahmad, Gulzar Alam, Ahmed Alwakeel, Mohammed Alwakeel, Lubna Abdulaziz Alharbi, Ahd Aljarf, and Muhammad Umair Khan. "Cloud servers: Resource optimization using different energy saving techniques." *Sensors* 22, no. 21 (2022): 8384.
- Katal, Avita, Susheela Dahiya, and Tanupriya Choudhury. "Energy efficiency in cloud computing data centers: a survey on software technologies." *Cluster Computing* 26, no. 3 (2023): 1845-1875.
- Kumar, Pawan, and Rakesh Kumar. "Issues and challenges of load balancing techniques in cloud computing: A survey." *ACM computing surveys (CSUR)* 51, no. 6 (2019): 1-35.
- Kocot, Bartłomiej, Paweł Czarnul, and Jerzy Prończ. "Energy-aware scheduling for high-performance computing systems: A survey." *Energies* 16, no. 2 (2023): 890.
- Koubàa, Mohamed, Rym Regaieg, Abdullah S. Karar, Muhammad Nadeem, and Faouzi Bahloul. "A Multi-Objective Approach for Optimizing Virtual Machine Placement Using ILP and Tabu Search." In *Telecom*, vol. 5, no. 4. MDPI, 2024.
- Koronen, Carolina, Max Åhman, and Lars J. Nilsson. "Data centres in future European energy systems—energy efficiency, integration and policy." *Energy efficiency* 13, no. 1 (2020): 129-144.
- Ma, Zhoujun, Di Ma, Mengjie Lv, and Yutong Liu. "Virtual machine migration techniques for optimizing energy consumption in cloud data centers." *IEEE Access* 11 (2023): 86739-86753.
- Male, Henry Kenneth, and Martin Kahangi. "Enhancing Energy Efficiency in Cloud Data Centers: The Role of Virtualization in Sustainable Computing." (2025).
- Mann, Zoltán Ádám. "Allocation of virtual machines in cloud data centers—a survey of problem models and optimization algorithms." *Acm Computing Surveys (CSUR)* 48, no. 1 (2015): 1-34.
- Mavridis, Ilias, and Helen Karatza. "Combining containers and virtual machines to enhance isolation and extend functionality on cloud computing." *Future Generation Computer Systems* 94 (2019): 674-696.
- Mohammed, Bashir. "A Framework for Efficient Management of Fault Tolerance in Cloud Data Centres and High-Performance Computing Systems: An Investigation and Performance analysis of a Cloud Based Virtual Machine Success and Failure Rate in a typical Cloud Computing Environment and Prediction Methods." PhD diss., University of Bradford, 2019.
- Noshy, Mostafa, Abdelhameed Ibrahim, and Hesham Arafat Ali. "Optimization of live virtual machine migration in cloud computing: A survey and future directions." *Journal of Network and Computer Applications* 110 (2018): 1-10.
- Odun-Ayo, Isaac, Olasupo Ajayi, and Chinonso Okereke. "Virtualization in cloud computing: Developments and trends." In *2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS)*, pp. 24-28. IEEE, 2017.
- Oloruntoba, Oluwafemi. "Architecting Resilient Multi-Cloud Database Systems: Distributed Ledger Technology, Fault Tolerance, and Cross-Platform Synchronization." *International Journal of Research Publication and Reviews* 6, no. 2 (2025): 2358-2376.
- Osman, Ahmed, Assim Sagahyoon, Raafat Aburukba, and Fadi Aloul. "Optimization of energy consumption in cloud computing datacenters." *International Journal of Electrical & Computer Engineering* (2088-8708) 11, no. 1 (2021).
- Oye, Emma, and Alex Matthew. "AI-Driven Cloud Evolution: Transforming Infrastructure for Future-Ready Solutions." (2024).
- Safari, Ashkan, Hoda Sorouri, Afshin Rahimi, and Arman Oshnoei. "A Systematic Review of Energy Efficiency Metrics for Optimizing Cloud Data Center Operations and Management." *Electronics* 14, no. 11 (2025): 2214.
- Segun-Falade, Osinachi Deborah, Olajide Soji Osundare, Wagobera Edgar Kedi, Patrick Azuka Okeleke, Tochukwu Ignatius Ijomah, and Oluwatosin Yetunde Abdul-Azeez. "Assessing the transformative impact of cloud computing on software deployment and management." *Computer Science & IT Research Journal* 5, no. 8 (2024): 2024.
- Seyyedsalehi, Seyyed Mohsen, and Mohammad Khansari. "Virtual machine placement optimization for big data applications in cloud computing." *IEEE Access* 10 (2022): 96112-96127.
- Shafiq, Dalia Abdulkareem, Noor Zaman Jhanjhi, Azween Abdullah, and Mohammed A. Alzain. "A load balancing algorithm for the data centres to optimize cloud computing applications." *IEEE access* 9 (2021): 41731-41744.
- Vasudevan, Meera, Yu-Chu Tian, Maolin Tang, Erhan Kozan, and Weizhe Zhang. "Profile-based dynamic application assignment with a repairing genetic algorithm for greener data centers." *The Journal of Supercomputing* 73, no. 9 (2017): 3977-3998.
- Venkata, Baladari. "Cloud Resiliency Engineering: Best Practices for Ensuring High Availability in Multi-Cloud Architectures." (2022).