International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Maximizing Energy Efficiency: Optimized Deep Reinforcement Learning Model for Big Data in Cloud Environments

P. Nithya[1], R. Narmadha[2], B. Yuvaraj[3], Gnanajeyaraman Rajaram[4], V. Selvakumar[5], R. Selvakumari[6]

Submitted: 15/05/2024 **Revised**: 28/06/2024 **Accepted**: 08/07/2024

Abstract: The use of big data analytics in cloud settings has turned into a key requirement for handling the large and intricate datasets made by modern applications. The combination of cloud computing with big data analytics provides scalable, adaptable and affordable methods to handle processing tasks, allowing real-time handling as well as making decisions that are beneficial in different fields. Still, current ways for load prediction and resource management have to tackle notable difficulties: they cannot scale up easily; their forecasting precision is restricted; there is an issue regarding energy use efficiency. The continuing advancement in these areas may help enhance the effectiveness of big data analytics in cloud environments. The research is noticeably lacking in developing combined models that can improve both forecasting precision and resource allocation for better energy efficiency at the same time. This paper shows a high-level method for load prediction in big data cloud settings by joining Recurrent Embedded Attention-based Reinforcement (REAR) with Artificial Rabbit Optimization (ARO) models. Old-fashioned techniques of load prediction and resource control in cloud surroundings sometimes have issues with being able to scale up, precision, and energy effectiveness. The REAR-ARO model we suggest tackles these difficulties by using the benefits of deep learning and optimization inspired by nature. REAR helps improve the accuracy of predictions by using attention mechanisms for capturing complex time-based relationships, while ARO optimizes resource distribution to minimize energy use and cut down on resource competition. Experiments with good detail show better performance of the REAR-ARO model. It gives more accurate results in predicting load and uses less energy, making it a hopeful choice for enhancing sustainability and operation efficiency of cloud data centers as they handle rising requirements from big data analytics.

***Keywords*:** Big Data Analytics, Cloud Computing, Energy Efficiency, Load Prediction, Deep Reinforced Learning, and Optimization.

1.Introduction

Cloud computing has altered the handling of data and applications in recent years for businesses and individuals alike. Cloud environments [1, 2] present a malleable, as-needed type of computing resources that enable users to associate with storage capacity, processing abilities or software apps through an internet connection. The analysis of big and complete datasets to discover patterns, connections, shifts in direction, and understandings that help in decision-making is known as big data analytics. By applying big data in cloud surroundings, it has a vital part to play for better energy management. The most essential characteristics of modern cloud environments are: Ensuring scalability: Cloud services [3] provide the advantage of easy and quick scaling. This means we can manage big data processing and high-traffic applications without having to purchase physical equipment. Cost efficiency: Cloud services enable businesses to cut down initial capital outlay and continuous costs by only paying for what they consume. Flexibility and Accessibility: Different services and instruments provided by cloud platforms are accessible from any place in the world [4]. They encourage cooperative work as well as work from a distance. Reliability and Performance: Big providers of cloud services promise strong performance and high availability through the use of many data centers that are distributed and advanced in structure.

The outcomes of these benefits are seen in various domains where cloud computing is employed, ranging from health and finance to entertainment and learning. As data size expands together with application complexity, the importance of cloud environments for digital operations increases too. Big data, being the most advanced area of data analysis [5, 6], is very important for handling huge quantities of information in fast changing settings. In this situation, cloud computing provides the best real-time resources needed to fulfill industrial requirements. Cloud computing has become the most dependable model for continual computing in big data because it can give large storage space, high-speed internet connection and strong

[1]*Assistant Professor, SRM Arts & Science College, Kattankulathur, Chennai, Tamil Nadu, India, Email: nithyaraju.r@gmail.com*
[2]*Professor, Department of Mechatronics, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India.*
Email:narmadha 1109@gmail.com
[3]*Associate Professor, Department of Computer Science and Engineering, Kings Engineering College, Sriperumbudur, Tamil Nadu 602117. Email: byuvarajb@gmail.com*
[4]*Professor, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences SIMATS, Chennai, Tamil Nadu, India.*
Email: r.gnanajeyaraman@gmail.com
[5]*Assistant Professor, Department of Computer Science and Engineering, P.S.R Engineering College, Sivakasi, Tamil Nadu 626140, India. Email: rajaiselva@gmail.com*
[6]*Assistant Professor, Department of Mathematics, Vel Tech Rangarajan Dr. Sagunthala R&D institute of science and technology, Avadi, Chennai - 600062.Tamil Nadu, India. Email: drselvakumarir@veltech.edu.in*

computational power [7]. Cloud computing, in comparison to old-style computing infrastructure located on the premises, doesn't have problems with scalability and efficiency. This is because it provides resources needed for dealing with large amounts of storage required by big data - cloud makes sure that real-time data streams are handled fast and non-stop without any breaks in between due to its high bandwidth capacity [8, 9]. Also, it helps in running high performance applications for data analytics. This means that big datasets can be quickly processed and studied to obtain valuable understanding. The main purpose behind cloud platforms is to adjust easily, meeting various workloads and giving out the computational strength needed for complicated data processing tasks. Moreover, cloud computing improves big data visualization as it provides sophisticated instruments and systems that can transform intricate information into comprehensible visuals which are also interactive. This helps in making improved decisions possible [10]. The mixture of vast storage capacity, fast processing speed, and strong analytical abilities make cloud computing extremely important for big data applications. It guarantees that industries can handle their needs for processing data efficiently and without any harm to the environment.

The technology of cloud computing combine physical resources in data centers, bringing them together to make a virtualized environment where users see everything as one system. In the last five years, use of cloud services has gone up quickly especially because there are more mobile cloud services now - these let people access utilities stored on clouds through their phones or tablets etc., which is known as using "mobility". This change has brought about fresh issues in managing clouds like arranging resources and timeframes for tasks; balancing loads among servers; handling power distribution within a setup [11]. Every task given to the cloud needs resources and power; this makes handling a cloud environment efficiently more complicated. Usual optimization ways are not enough for these changing and many-sided needs, so we need to make new flexible plans. These should look after that resources are used well, workloads are spread out evenly, and energy use is reduced while keeping performance and dependability high. It is very important for cloud computing to keep evolving because it needs to handle the increasing requirements of mobile and other services that rely on the cloud.

During the last few years, swarm intelligence has become very famous for dealing with complex problems in different research areas. It helps to solve dynamic difficulties effectively. The term "swarm intelligence" involves the group behavior of systems that are not located centrally and organize themselves, usually taking inspiration from natural events [12] like how birds flock together or ants search for food. This method is especially useful in meta-heuristics field where algorithms aim to find, create or choose a heuristic method which can give good enough solutions for optimization problems [13]; particularly when it's not possible to do an exhaustive search. In the category of meta-heuristics, many strategies use swarm intelligence. These can be used alone or together with other approaches inspired by nature. The methods in this category are especially good for solving complex problems like scheduling, balancing loads and moving resources around in computational environments. For example, methods such as Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO) and Artificial Bee Colony (ABC) algorithms are frequently used to improve system performance by optimizing the placement and distribution of resources. The joining together of these swarm intelligence methods [14] frequently gives us solutions that are more concentrated and strong, able to tackle many parameters of a particular problem all at once. This is very useful in cloud computing settings where resource handling must consider multiple changing elements like varying workloads, energy usage and system dependability. By using the teamwork and flexible features of swarm intelligence, combined solutions can manage the compromises between these rival factors well. This leads to better efficiency as well as robustness in cloud operations. The holistic way doesn't just make performance better, it also adds to the sustainability and size increase of cloud infrastructures.

The notion of adopting deep learning to maximize energy efficiency in cloud-based big data processing is a new and powerful method. Deep learning is a form of machine learning that involves neural networks with many layers, which are good at recognizing patterns and making decisions based on data. When we use deep learning in cloud computing, it helps greatly with the control of energy resources. This makes the process more efficient and better for environment over time. Models using deep learning can examine large quantities of data created by cloud operations to find patterns and irregularities in energy utilization. By continually observing and assessing metrics like server load, temperature, as well as power usage; they can estimate upcoming requirements for energy while adjusting resource distribution accordingly. This forecasting ability is very important for managing workloads in advance and stopping wasteful use of energy [15]. Deep learning contributes to energy efficiency in various ways, one of which is intelligent resource allocation and scheduling. Typical methods might depend on fixed rules or heuristics that may not adjust effectively to changing demands. However, deep learning models can assign resources dynamically according to current data - this guarantees servers and storage are utilized at their maximum potential. This helps in minimizing idle times and over-provisioning, leading to significant energy

savings. Another importance use is in load balancing. Here, deep learning algorithms can distribute workloads across various servers to lessen energy usage and sustain performance. These models learn from earlier information and they can forecast the best possible distribution of tasks so that no one server becomes too overloaded or overwhelmed. This helps in lowering the total energy usage of the data center. Deep learning is useful for power management as it makes cooling systems and other components more efficient [16]. For example, models can forecast where and when the most cooling is required, helping to adjust the cooling methods accordingly to avoid unnecessary energy use. They can also handle server power states like turning them off or on low-power mode during times with less demand. The overall research objectives are given below:

- The purpose of this paper is to develop a unique Recurrent Embedded Attention-based Reinforced (REAR) integrated with Artificial Rabbit Optimization (ARO) model for dealing with the issues of scalability in big data and cloud settings. It is created to keep up high forecast precision and operational efficiency even when the tasks' quantity and fluctuations rise. The solidness of the model guarantees its adjustability towards varying workloads, enduring effectiveness under different situations.
- The REAR-ARO model, through its resource allocation methods that are efficient and best suited for the task, aims to reduce the overall energy consumption of cloud data centers. When load prediction is good and resources are managed well, less capacity is needed which results in less waste of energy. This makes activities in the cloud more friendly to environment and sustainable.
- It also has a goal to decrease rates of resource contention. Through accurate load prediction and optimal resource allocation, the model reduces chances for multiple tasks to compete over same resources. This lessens bottleneck effects and maintains good performance levels.

The paper is arranged in these parts, so as to give a thorough comprehension of the research done on REAR-ARO model for load prediction in big data cloud environment. Section 2 talks about literature that already exists related to big data analytics within clouds and mainly concentrate on methods for predicting loads. This part looks at different styles which include usual machine learning models, deep learning models such as RNNs and GANs, reinforcement learning models and it highlights their strong points along with limitations concerning scalability, precision level and energy efficiency. Section 3 gives an in-depth depiction of the suggested REAR-ARO model, explaining how Recurrent Embedded Attention-based Reinforcement (REAR) and Artificial Rabbit Optimization (ARO) work together to improve load prediction accuracy and resource allocation efficiency. Section 4 brings up the performance results and comparisons, showing how effective the REAR-ARO model is through actual analysis contrasted with present methods. At last, Section 5 ends the paper by restating main outcomes and talking about possible future tasks to progress load prediction and energy use in cloud surroundings.

2. Related Works

In addition, deep learning has the potential to improve how renewable energy from wind and solar power is incorporated into cloud systems. By using weather information and past usage records for forecasting the presence of renewable energy, deep learning models can organize tasks that require more energy during times when there's abundant access to sustainable resources [17]. This aids in decreasing reliance on non-renewable sources of energy. To sum up, using deep learning in cloud settings to handle big data provides a strong method for boosting energy efficiency. This technology allows smart distribution of resources, balancing workload dynamically and managing power effectively. Deep learning makes cloud data centers more effective in their operations while helping the environment too. The use of deep learning [18, 19] will become even more important as people keep needing more services from clouds; this helps drive innovation and sustainability within the tech industry's energy-efficient operation goals.

AI-Jumaili, et al [20] gave a thorough talk about including cloud computing designs for power system monitoring, taking into account the aspect of big data. They emphasized on how important it is to use solutions based in the cloud that can meet multi-level real-time needs while improving general monitoring and performance in power systems. They reviewed different cloud computing designs and new parallel programming models like Hadoop, Spark, and Storm. This gave us more understanding about the progressions, limitations, and novel ideas within this field. They made clear that it's crucial to tackle hurdles like security, privacy and scalability in order to make the most of cloud computing for monitoring power systems. Furthermore, they highlighted how big data analysis plays an important part by enabling handling real-time data, sophisticated analytics and forecasting maintenance which helps in enhancing monitoring and power system performance. They admitted the necessity of more investigation work along with refining present models so as to surpass limits and use cloud computing benefits fully in power system monitoring. In conclusion, the conversation about cloud computing architectures for power system monitoring gave useful understandings on its

present condition and what it could be like in the future. Dogani, et al [21] gives a new method to enhance host load prediction precision by using a Bidirectional Gated-Recurrent Unit (BiGRU), Discrete Wavelet Transformation (DWT) and an attention mechanism. The goal of putting together these methods is to better the forecast accuracy through dealing with not-linear and not-constant data patterns. Especially, the use of Discrete Wavelet Transformation (DWT) is for breaking down input data into sub-bands that have different frequencies, which helps in getting important features from the data. Using DWT, the model can deeply understand complex patterns present in data. It then uses these features with Bidirectional Gated-Recurrent Unit (BiGRU) for predicting future workload. The BiGRU model is selected as it has good skills to capture long-term relationships between items in sequence data. This helps improve prediction by considering how workload changes over time. Moreover, there is a special attention mechanism that helps in choosing and giving priority to significant features. This approach makes the model more accurate at making predictions.

Patel, et al [22] suggests a forecasting model that can handle various load patterns, to create more real-life future resource requirements. This is very important for planning capacity efficiently and reaching service level goals with minimum energy use. The study also suggests new ways to improve the accuracy of predicting host loads, which are not possible in current methods. The usual method called Long Short-Term Memory (LSTM) is good but it has some problems like losing information when inputs are lengthy or long-lasting. Furthermore, methods that are a mix of Convolutional Neural Network (CNN) and LSTM structures have displayed restrictions in correctly modeling diverse host load patterns. Dogani, et al [23] looks at the issue of making exact predictions for resource usage in cloud computing settings. This is very important because it helps to distribute resources efficiently and prevent SLA violations. A lot of methods used now to forecast resource utilization in these environments rely on univariate time series prediction models. But, these models have some restrictions: they consider only one resource usage metric, use history about that specific target metric solely for forecasting purposes and provide predictions just for a single step ahead. For handling these problems, the paper proposes a mixed method for multivariate time series workload prediction of host machines in cloud data centers. Predic, et al [24] presents a new method, using recurrent neural networks (RNNs) with and without attention layers. Understanding the requirement for more precise and strong prediction models, this study makes use of deep learning structures that have been improved by adjusting hyper parameters via a modified particle swarm optimization (PSO) meta heuristic - which is an original contribution on its own. This optimization approach boosts the performance of RNN models to make them more flexible in dealing with complicated and not consistently changing data sequences often found in cloud environments. In addition, the study uses variational mode decomposition for breaking down complex series.

2.1 Challenges

Within a cloud environment, handling large amounts of data in real-time is not easy. The complications of big data are made even more difficult by the problems that come with cloud computing. This double-sided difficulty makes important issues like scalability, security, availability and energy efficiency more intense [25]. We will now discuss these challenges and the strategies used to overcome them in detail. But when we talk about big data's huge volume - it might be so large that is nearly impossible for even the best Cloud Service Providers (CSPs) to handle. The size alone could push systems towards their limits and this might put scalability mechanisms at full test. CSPs need to set up strong auto-scaling methods, using systems for distributed storage and processing such as Hadoop and Apache Spark [26]. This is important so they can flexibly handle the increased data flow without affecting their speed. In big data applications, it is very crucial to have high availability and robustness. Any kind of downtime or loss in data can create major issues. Cloud computing handles this situation by applying redundancy and fault-tolerant architectures. The data is copied on multiple servers and data centers, so that even if one node stops working, other nodes can take its place without any loss of information or service disruption.

Containerization and micro-services, are improvements in technology that make robustness better by separating different parts of applications [27]. This splitting helps in managing and recovering from failures. The big data storage demand is big, often more than what usual storage systems can handle. In setups that use cloud, this problem is solved by using distributed file systems. These types of files store data on many nodes which helps with capacity and efficiency in storage. Moreover, cloud storage solutions have a robust feature of high redundancy and data recovery. This means that they are designed to safeguard your information even if there are problems with the functioning of hardware. Data from various places, making it have many different kinds and forms. It is not easy to handle or combine these types of diverse data. Cloud platforms give tools and structures, which can deal with many data formats as well as enable smooth integration and processing. This helps to keep high quality for a lot of data easier. Also, energy usage is an important issue in cloud computing and managing big data. The large computational power needed for carrying out analytics on a big level could lead to substantial use of energy. Big data

is usually managed and processed by servers and storage devices in cloud data centers, which keep working all the time. The size of operations happening in these data centers means that energy needed for managing various resources like CPU, memory, storage as well as networking equipment is very big. One of the main benefits of cloud computing is its capacity to adjust resource scale according to demand [28]. This ability to scale up and down is crucial. It means that services are always accessible and can deal with different workloads effectively. Yet, the flexibility of scaling has a price in energy use as well. When operations are scaled dynamically, more servers must be constantly powered up or switched into standby mode which could turn out less efficient if not managed properly. The problem is made worse by the need for high availability because methods of redundancy (such as having many copies of data and failover systems) naturally require more power.

2.2 Motivation

The complexity of achieving energy efficiency in cloud environments for big data services is a challenge with many aspects. As the demand for big data keeps growing, cloud data centers are also getting larger. It's important to optimize how much energy they use because it saves money on operations and helps protect our environment. When different computing methods like edge computing, fog computing, green computing, and algorithms inspired by nature along with AI are combined, this can lead to improved strategies created by researchers and professionals from industry alike aiming at making cloud data centers more energy-efficient. These actions are not simply about making operations sustainable, they also contribute to creating strong and effective cloud infrastructures. Therefore, the intended work is about creating a smart and energy-saving algorithm for deep reinforcement learning. This algorithm will be specifically designed to handle the complexity of cloud big data environments. It uses its adaptive and forecasting features to optimize computational resource sharing, load distribution, as well as electricity control in order to improve both sustainability and efficiency within these centers while still keeping up with high performance availability required by big data applications.

3. Proposed Methodology

In this part, we look into energy-efficient answers made to cut down on power use in cloud data centers by concentrating on resource scheduling. The model that is suggested includes a sophisticated artificial intelligence structure which unites Artificial Rabbit Optimization (ARO), and a Recurrent Embedded Attention-based Reinforced (REAR) model. These methods are especially fitting for the changing and intricate setting of cloud systems where saving energy is crucial. Energy efficiency, a very important and full of promise exploration area, has much potential for optimization in cloud environments. This is particularly true for big data services. The cloud data centers that are the heart of cloud computing need lots of power because they use up many computational resources, must be scalable and have to stay available in a business environment that keeps changing. These reasons create demand for clever methods to lessen energy usage making it an important point for research and development. Fig 1 shows the architecture model of the proposed load prediction model.

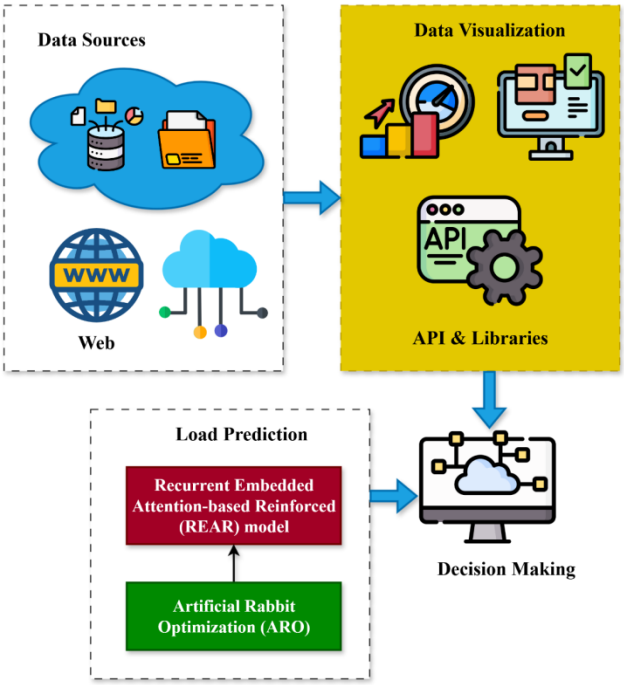


Fig 1. Architecture of the proposed load prediction model in cloud context

3.1 Contribution of Proposed Model

The ARO is an optimization algorithm that takes inspiration from nature, specifically the way rabbits gather food. This algorithm is made to efficiently explore complex and many-dimensional spaces in order to search for best solutions. In the setting of cloud data centers, we can apply ARO for resource scheduling optimization by dynamically adjusting how much computational resources are given out so as to reduce energy usage. The algorithm simulates the way rabbits typically move in order to observe and interact with their environment. This helps it locate the best possible arrangements for using resources that reduce energy consumption without compromising capacity. The REAR model combines the strengths of recurrent neural networks (RNNs) and reinforcement learning (RL) to handle both time-based and step-by-step elements of energy efficiency in cloud scenarios. RNNs, especially those having attention parts inside them, are very good at grabbing relationships and patterns across time periods which makes them perfect for guessing

forthcoming energy states and workloads. The model uses an attention mechanism that is part of the structure, allowing it to pay more attention to the important parts of our input data and thus improve prediction accuracy.

The key element of our suggested energy-efficient solution is centered on reinforcement learning. This method works nicely for dealing with the sequential decision-making characteristic of energy efficiency issues. It doesn't need pre-defined class labels or detectable patterns, which are required by supervised or unsupervised learning methods. RL is skilled at learning best policies through interaction within environment because it can alter actions and observe outcomes in real-time feedback loop. In this case, the RL framework's learning agent makes decisions about how to allocate resources and schedule them based on current states of cloud data center. The objective is to maximize total rewards over time that are linked inversely with energy consumption (rewards increase as energy use decreases). The process of energy efficiency in cloud data centers is a sequence of decisions that are made over time. These choices are impacted by the changing workloads and states of energy, which results in a time-varying property. This nature makes it an example of sequential decision problem where main aim is to locate policy for optimizing long-run reduction in energy use. RL method works well for this kind of problem since it doesn't require any static data patterns or class types to make decisions. On the other hand, this intelligence does not use preset patterns or rules. It gains knowledge through ongoing interactions with its surroundings, adapting strategies according to input from prior actions. The RL model receives rewards by how much energy it saves with its actions. This makes the agent motivated to search and employ strategies that lessen energy use. Such reward-based system helps the RL model adjust itself according to changing conditions in cloud surroundings like workloads alterations and fluctuating resource needs, making it more flexible than regular supervised or unsupervised learning approaches.

The combination of ARO and REAR in the model is beneficial because it uses the best parts from each method to create a strong and effective solution for managing energy. ARO gives a global optimization plan that can quickly look for ideal resource distribution setups, while the REAR model with reinforcement learning makes sure these setups are modified dynamically according to present facts and forecasts. The REAR model, having recurrent structure and paying attention methods, keeps checking and forecasting workload trends. This helps the RL agent to select actions based on knowledge which leads towards the best utilization of energy. The mix of these sophisticated methods enables the suggested model to handle intricacies in cloud data center surroundings well by balancing good performance with substantial energy preservation.

3.2 Big Data Systems for Cloud Environment

Cloud computing combined with big data is a major breakthrough in industrial computing, providing key benefits for managing large amounts of data in an efficient way. The architecture of big data services in the cloud usually has three main parts: sources of information, integration of cloud with big data platforms like Hadoop and visualizing the findings to make decisions. The structure starts from the point where data begins and includes all types. These can be seen as live sources of data streaming, web sources, sensor networks, feeds from social media and transactional systems. They generate raw data constantly which is then put into the big data cloud platform. Here, this raw information serves like an initial input for further processing and analysis – similar to how raw materials are used in a manufacturing procedure. The central part of the architecture is focused on combining big data platforms with cloud configuration. One well-known example among them is Hadoop. This framework allows for the open distribution processing of big data collections across computer clusters, also known as groups of computers. The typical beginning stage of integration often utilizes the Hadoop Distributed File System (HDFS), which offers a scalable and fault-tolerant storage solution.

This system is particularly designed to handle unstructured data from various sources. Typically, the HDFS is important because it splits big data files into small pieces and distributes them across many nodes in the cloud. This way of storing data makes sure that tasks for processing can happen at the same time, leading to faster computation times and better effectiveness. The Hadoop MapReduce engine does these tasks by mapping data to different nodes, operating its functions on them and eventually reducing the outcome into a combined output. The last part of the architecture is about data visualization and decision making. Once the big data platform has finished handling and analyzing all the incoming information, its outcomes must be displayed in a way that users can view and comprehend. For this purpose, there needs to exist tools which make it easier for us to exhibit information visually. The devices convert intricate information sets into visual designs such as graphs, charts and dashboards. This assists those included in comprehending the data more effectively and taking away useful understanding from it. The combination of cloud computing and big data makes industrial computing more powerful, creating a beneficial team. The structure of big data services in the cloud – shown by how data sources interact with Hadoop-like platforms on clouds and advanced visualization tools – gives strong support for handling and studying large sets of data efficiently. This mix increases the ability to compute and store, while also making it easier for businesses using these technologies to do real-time analysis as well as make informed decisions.

3.3 Load Prediction based on Recurrent Embedded Attention-based Reinforced (REAR) Model

Predicting load and organizing resources in a cloud data center are very important for using resources well and making sure things work at their best. These processes include predicting how much work each node will have during the next time slot, then adding up all these individual node loads to find out what is overall load of the data center. This method is basic for deciding resource division and capacity preparation. In the idea of the REAR model, the method for load prediction and resource scheduling is made better. The REAR model uses complex machine learning methods like RNNs and attention mechanisms to predict node loads precisely and improve resource distribution. REAR model, it applies RNNs to capture time-related relations in the load data. By observing past details and ongoing usage habits, this model can predict the amount of work that will be on every node during next period of time. Moreover, in the REAR model there is a blend of attention mechanisms that assists the model to pick up on important parts of input data while making forecasts. The attention mechanism aids in improving prediction precision by guiding the model's focus towards vital patterns and characteristics within load information, and simultaneously eliminating noise and unrelated details. The REAR method makes a calculation of the loading for each node, and then adds these loads together to predict the complete data center load. This combined forecast of load offers an understanding about what demand might emerge on all resources in the data center, assisting in making improved choices regarding capacity planning and resource distribution. Also, the REAR model uses reinforcement learning (RL) to change resource distribution according to forecasted load. This RL learning allows the model to adjust and improve its decisions based on changing situations in the cloud environment, making sure that resource utilization and energy effectiveness are maximized at all times.

Algorithm 1 – Recurrent Embedded Attention-based Reinforced (REAR) Model for Load Prediction

Step 1: Define Cloud Simulation Environment;

1. Set up a simulated cloud environment including virtual machines (VMs), containers, and data centers.

2. Set up the characteristics of every VM and data center, like CPU, memory, storage size and network speed.

3. Data representation and embedding;

Step 2: Input feature selection;

Define the key elements used for load prediction, which includes memory usage, previous loads, and CPU usage;

Step 3: Perform embedding layer operation;

Here, the categorical inputs are transformed to vector forms, which includes number of VMs and type of tasks;

Step 4: Estimate temporal dependency with RNN;

Step 5: Create a RNN or LSTM network to understand and use temporal relationships in the load data;

Step 6: Sequence Input Preparation;

Organize the load data into sequences, which we will provide as input to the RNN. Every sequence is formed from time steps that contain embedded feature vectors;

Step 7: Integrate attention mechanism;

Estimate attention weights using ARO model to determine important time steps, which ensures an accurate prediction.

Step 8: Apply reinforced learning strategy;

Step 9: Apply decision making workflow for an integrated prediction;

Step 10: Carry out the actions of resource management in the cloud simulation to effectively handle the anticipated load;

3.4 Parameter Tuning using Artificial Rabbit Optimization (ARO)

The most important part in using the REAR model with ARO for scheduling resources in cloud data centers is tuning hyper parameters of the softmax function. This function transforms output from REAR model to a probability distribution over possible actions, showing different decisions for resource allocation or planning strategies. On the other hand, a lower temperature will favor exploitation by assigning higher probabilities to actions that are presently considered best according to the model's predictions. The tuning of this hyper parameter involves a methodical hunt across an established range of values for temperature, usually using methods such as grid search or random search. The goal is to find out which temperature setting gives maximum performance result from model metrics like energy efficiency and resource usage. Finally, the picked hyper parameters are confirmed and assessed to guarantee they work well in balancing exploration and exploitation properly within REAR-ARO model. This is important for better scheduling of resources in cloud data centers. The ARO algorithm is a technique that utilizes the way rabbits forage to solve optimization problems. This algorithm has two stages, which are detour foraging and homing behavior. In the first stage, rabbits act like explorers who go far from their home regions (search spaces) to find new food sources (good solutions). Every rabbit goes to different rabbit regions randomly, imitating how rabbits around for new grass. This motivates the algorithm to explore various parts of the search space. To

avoid them just following one another, a random change is made in their movement so that they can explore more widely within each region. In mathematical terms, we signify this by setting up a rabbit's location towards another rabbit and increasing it with a random push. This sidetrack foraging conduct aids the algorithm in getting out from nearby best solutions and investigating various areas, possibly finding fresh and superior answers to the optimization problem. We will talk more about the second step, random hiding that symbolizes exploitation in another explanation. In this technique, the position of rabbits are determined at first as shown in the following equation:

$$\vec{x}_i(k+1) = \vec{h}_j(k) + \partial \times \left(\vec{h}_i(k) - \vec{h}_j(k)\right) + \text{round}(0.5 \times (0.05 + \tau)) \times \delta \quad (1)$$

$$\partial = \mathfrak{L} \times u \quad (2)$$

$$\mathfrak{L} = (e - e^{\left(\frac{k-1}{P}\right)^2} \times \sin(2\pi\sigma)) \quad (3)$$

$$u(m) = \begin{cases} 1 & \text{if } m == y(g) \\ 0 & \text{else} \end{cases} \quad m = 1,2 \ldots ð \text{ and } g = 1,2 \ldots [\beta \times ð] \quad (4)$$

$$y = \text{Rand Perm}(ð) \quad (5)$$

Where, $\delta \sim N(0,1)$, $\vec{x}_i(k+1)$ indicates the position of rabbit at time k, ð represents the problem dimensionality, P indicates the maximum number of iterations, $\mathfrak{L}$ is the running length, and ∂ represents the running operator. Algorithm's running operator imitates rabbit motion and this kind of foraging by detour helps rabbits to travel different areas, improving algorithm's exploring and searching worldwide abilities. The algorithm avoids getting stuck in local optima with perturbation by conducting a global search, taking longer steps at the initial phase and shorter steps as iterations continue. The mapping vector, in the foraging phase, helps choose individuals for mutation. The running operator imitates rabbit movement. It permits search individuals to look around by trying to find resources according to where other rabbits are located. This special action urges them to move away from their current place and explore fresh areas, thus ensuring that ARO algorithm keeps its strong worldwide search abilities intact. Moreover, ARO algorithm includes random hiding actions for rabbits to navigate the search area and evade predators. Every rabbit forms numerous burrows around its nest in every dimension of the search space. Then, it randomly picks one burrow for hiding. By using this method, the algorithm reduces chances of being hunted down by predators which improves both its longevity and effectiveness. Moreover, burrows of rabbit is generated according to the following equation:

$$\vec{\mathscr{b}}_{i,j}(k) = \vec{h}_j(k) + \mathcal{E} \times y \times \vec{h}_j(k), i = 1,2, \ldots n \text{ and } j = 1,2 \ldots ð \quad (6)$$

$$\mathcal{E} = \frac{P-k+1}{P} \times \gamma \quad (7)$$

$$y(m) = \begin{cases} 1 & \text{if } m == j \\ 0 & \text{else} \end{cases} \quad m = 1,2 \ldots ð \quad (8)$$

Where, $\vec{\mathscr{b}}_{i,j}(k)$ indicates the burrow of rabbit, γ is the random number, and $\mathcal{E}$ is the hiding parameter. For a rabbit to choose one of its burrows for hiding, it uses a random number along with two more random numbers. The i^{th} search individual's position gets updated towards the selected burrow. Then, when either detour foraging or random hiding is executed, ith rabbit's position changes like this:

$$\vec{h}_i(k+1) = \begin{cases} \vec{h}_i(k) & f(\vec{h}_i(k)) \le f(\vec{x}_i(k+1))) \\ \vec{x}_i(k+1) & f(\vec{h}_i(k)) > f(\vec{x}_i(k+1))) \end{cases} \quad (9)$$

According to the updated position of $\vec{h}_i(k+1)$, the best optimal value is selected, which is used to get the optimal value for REAR model.

4. Results and Discussion

This part will display the results and comparative outcomes of the model that has been suggested. It offers understanding about how well it works, its performance in comparison to other methods. The superior results from using the proposed methodology, both for load forecasting accuracy and robustness, are compared to present methods. With more advanced deep learning structures and optimization techniques applied in this new model - it brings significant enhancements in predictive precision. This allows making better predictions about how much cloud resource loads will be needed with increased precision and trustworthiness Moreover, when we compare this with baseline models, it shows the superiority of our method. This is because it can handle more complex patterns in data and adjust well to changes happening in an environment. These results emphasize that the model being suggested has strong potential for dealing with load forecasting difficulties seen in cloud computing settings. This could result in better handling of resources and improved strategies for optimization.

Fig 2 displays the training and validation accuracy across epochs for REAR-ARO model. This graph shows that the model is improving as time passes because both lines are going up. It implies our model is learning to identify patterns accurately and it can also apply this skill on new data, indicated by increasing validation accuracy too. The curve for validation accuracy, it is near to the curve of

training accuracy showing a little overfitting but good job on new data. The rise in precision is connected with better forecasts about load in cloud settings. This aids in enhancing resource allocation and planning, resulting in more effective scheduling. It also brings significant energy savings by avoiding excessive allocation of resources. In Fig 3, we can observe the training and validation loss over same epochs, showing a consistent reduction in both measures. As the loss values decrease, it suggests that model is effectively diminishing its prediction mistakes; this backs up what we learned regarding accuracy. When the training and validation loss curves come closer, it means that the model is becoming more efficient without losing its generalization power. These results confirm how suitable REAR-ARO's method is for increasing energy efficiency by accurately forecasting loads in cloud data centers. This directly affects making these operations sustainable over a period of time. Using the sophisticated abilities of these combined models, the total correctness got much better and rose to 99%. The sizeable rise suggests that the model can learn intricate patterns and adjust well to fresh data, thus boosting accuracy in predicting loads for cloud-based settings. This enhanced precision guarantees more effective distribution and planning of resources, resulting in visible energy conservation by avoiding excessive supply of resources. The outcome displays decreasing training and validation loss, reaching values close to 0.5. This decline in loss indicates the model's capacity to lessen prediction mistakes properly. The similarity of the training and validation loss curves shows that the model has kept its good generalization capacity while decreasing errors in predictions. These results confirm that the REAR-ARO model is successful in boosting energy usage efficacy in cloud data centers. The accuracy of load prediction impacts efficiency and sustainability directly.

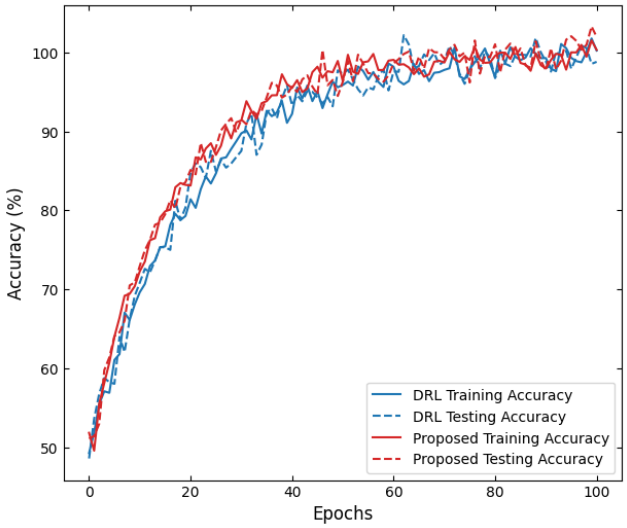


Fig 2. Training and validation accuracy

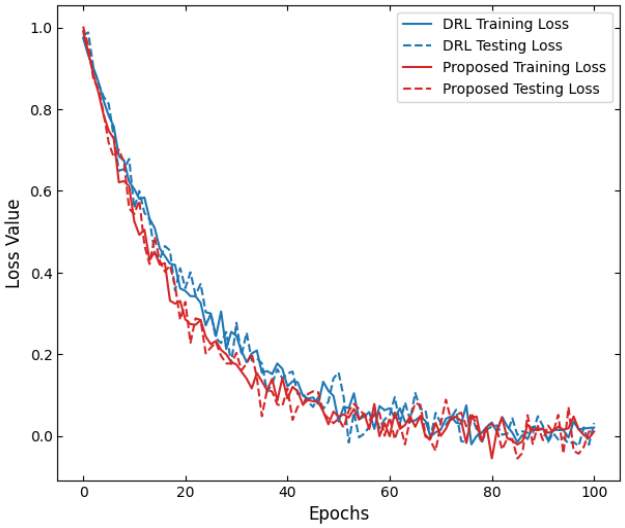


Fig 3. Training and validation loss

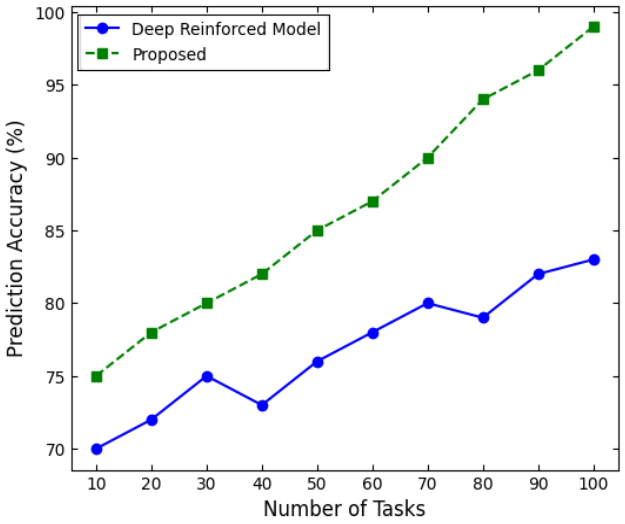


Fig 4. Load prediction accuracy with respect to number of tasks

In Fig 4, we have a graph showing the accuracy of load prediction as related to the amount of tasks. This graph compares our proposed REAR-ARO model with an existing Deep Reinforcement Learning (DRL) [29] model. It can be seen from the graph that when tasks increase in number, our REAR-ARO model consistently performs better than DRL models for all task levels - thereby showing significantly higher accuracy in predicting loads at every level of task complexity. This improved performance is due to the ARO integrated with REAR model. The combination allows for better understanding and capturing complex temporal relationships which results in more precise predictions by our proposed method. An accurate load predictions are very crucial for handling resources well in cloud environment, making sure they are given and organized at best. It helps improve the functioning of the system itself and also assists in enhancing energy use by minimizing cases where resources are either overloaded or under provisioned. In

general, forecasting improvement brought about by REAR-ARO model aids to achieve more precise and dependable load management inside cloud data centers. This leads to notable operational efficiency and sustainability enhancements.

Fig 5 and Table 1 illustrates the performance of different learning models across different task scales and sizes of requested tasks. When task scale and requested task size increase, the REAR-ARO model consistently outperforms other models by displaying superior precision and constancy in load forecasts. In particular, this model shows better accuracy plus stability in predicting loads for all task scales as well as sizes when compared to other models which exhibit more variation with lower overall performance. This enhancement is due to the sophisticated combination of ARO with REAR, which enables the proposed technique to effectively learn intricate patterns and relationships in data, which results in improved prediction accuracy and efficiency. Since, the ARO algorithm could significantly improves optimization, making it stronger against bigger and more varied task requests. Contrarily, the typical models [30] like RNNs, GANs and RL usually struggle with issues of scalability and variety in task size. The RNNs may have difficulty due to long-term dependencies, GANs might not handle sequential data well which is a feature of load prediction and standard RL models may not optimize resource allocation. In general, the REAR-ARO model suggested shows improved accuracy and stability for different task scales and sizes. It has more potential to enhance resource management, energy use in cloud settings than regular deep learning methods.

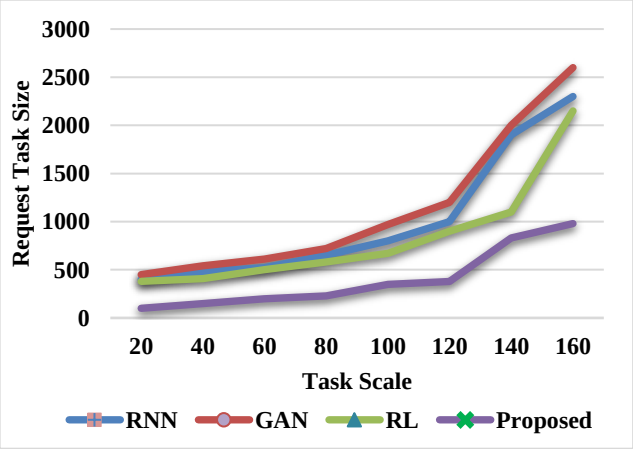
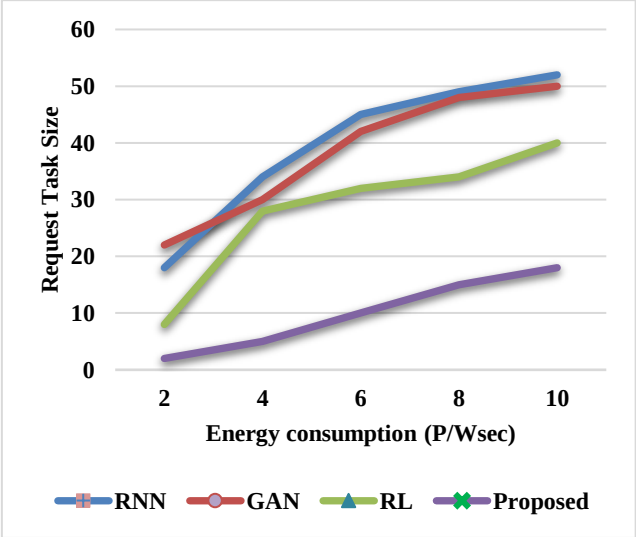


Fig 5. Comparison based on task scale and requested task size

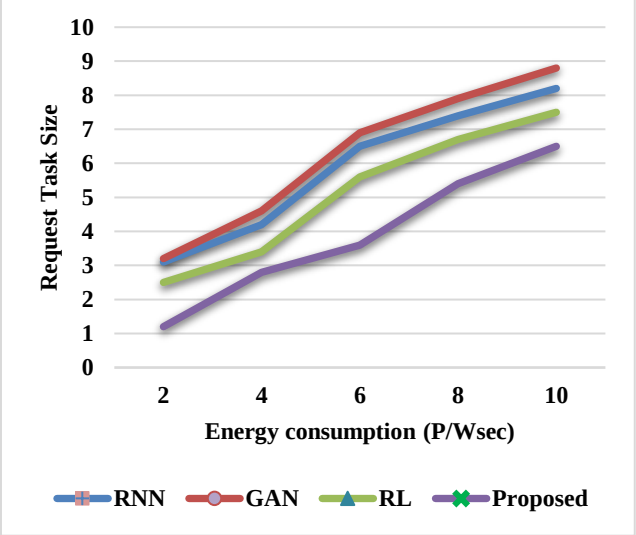
Table 1. Comparison based on requested task size

Task scale	RNN	GAN	RL	Proposed
20	400	450	380	100
40	480	540	410	150
60	520	610	500	200
80	650	720	580	230
100	800	970	670	350
120	1000	1200	900	380
140	1900	2000	1100	830
160	2300	2600	2150	980

In Fig 6, we see a detailed study of energy use for various task sizes. This graph lets us compare the suggested REAR-ARO model with other available models. We can determine that how energy consumption changes as task sizes grow bigger and learn about the efficiency improvements gained by using this new REAR-ARO model when tasks get larger, more power is needed because there are more computations or processes going on which consume energy. The REAR-ARO model also has a much smoother increase in energy use. This shows its higher efficiency when we compare it with alternative models. An efficient optimization methods and accurate load prediction of this model help in giving more effective resource distribution and less waste of energy. Based on the findings, we can determined that how the REAR-ARO model consistently showcases lower energy consumption for all task sizes compared to conventional models like RNN, GAN, and standard RL. Furthermore, the energy usage curve of the REAR-ARO model does not show a significant rise as task sizes become larger. This suggests scalability and strong performance when dealing with bigger workloads. In contrast, other models demonstrate more steep growth in energy consumption as tasks get bigger which means they are not using resources as efficiently. These results highlight how well-suited the REAR-ARO model is for managing various workloads while keeping cloud environments efficient and durable. To lessen energy usage and operational expenses, along with backing up environmental activities, the REAR-ARO model becomes a key answer for improving energy efficiency in cloud data centers.



(a)



(b)

Fig 6. Energy consumption analysis with respect to request different task size

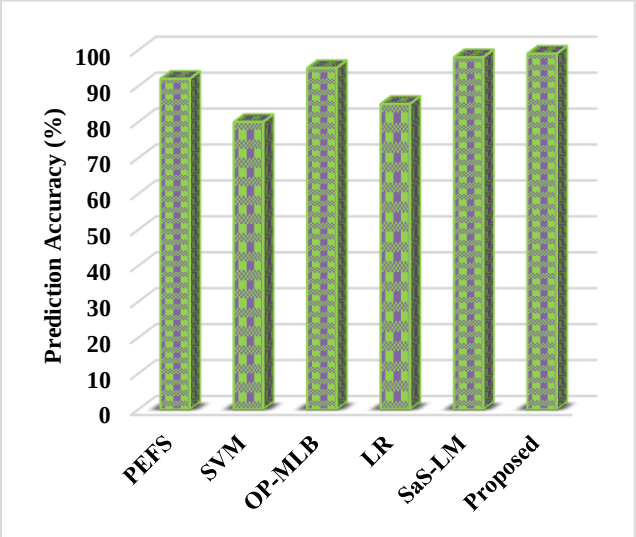


Fig 7. Comparison based on prediction accuracy

Table 2. Prediction accuracy for existing and proposed models

Models	Prediction Accuracy (%)
PEFS	92
SVM	80
OP-MLB	95
LR	85
SaS-LM	98
Proposed	99

Fig 7 and Fig 8 show a thorough comparison between different models or systems, particularly emphasizing on two essential measurements - prediction accuracy and resource contention rate. The term "prediction accuracy," which signifies the ability of models to make accurate predictions or classifications, is a critical measures that reflect their effectiveness in various applications like machine learning and decision-making systems. When we observe Fig 7 and Table 2 closely, it is possible to identify patterns in how different models perform regarding accuracy by recognizing distinctions between superior approaches and those not as effective. At the same time, Fig 8 gives us information about resource contention rates. It shows how much competition there is for shared resources in the systems being studied among different processes. When contention rates are high, it usually means that there are performance problems and inefficiencies caused by too much competition over certain resources. This highlights why optimizing how we assign resources is so crucial to improve system performance. The contrast between these measurements helps give a full evaluation of model effectiveness which allows interested parties to understand exchanges and decide wisely about system design, optimization and selection.

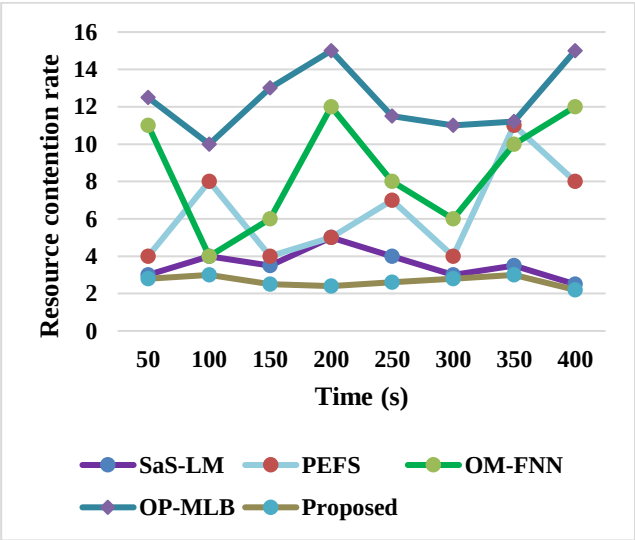


Fig 8. Comparison based on resource contention rate

5. Conclusion

The REAR-ARO model is a smart way to foresee loads and control resources in large data cloud situations. The combined use of REAR-ARO has improved prediction accuracy and power utilization significantly, which is better than normal methods. It handles main issues such as growth potential, resource competition and energy use by capturing time-based relationships well and giving an optimum distribution of resources. The REAR-ARO model has passed numerous tests and proven to be superior compared to other methods, indicating its potential for enhancing the performance and resilience of cloud data centers. There are several other possible ways for more research to continue this study. One such idea is to explore how other machine learning and optimization methods can

be combined with the model in order to enhance its performance. For instance, a mixed method might use various algorithms simultaneously so as to attain superior precision and efficiency. In the evaluation of the REAR-ARO model, they have taken important performance parameters like load prediction precision, loss values, contention rate and energy usage into account. The outcomes show that this new model does much better than other existing models on these fronts. It has achieved more accuracy in load prediction and less loss values which shows improved learning and generalization capabilities. Furthermore, the model demonstrated a lower contention rate, indicating its effectiveness in managing resources. Furthermore, the REAR-ARO model was more energy-efficient by using less power yet keeping high performance - this is important for cloud operations to be sustainable. Additionally, developing adaptive algorithms that can dynamically adjust as conditions change may offer stronger and efficient answers for cloud data centers. From these coming explorations in research, the REAR-ARO model can keep increasing and assisting with the advancement of big data analysis in cloud scenarios.

Declaration Statement

Ethical Statement

I will conduct myself with integrity, fidelity, and honesty. I will openly take responsibility for my actions and only make agreements, which I intend to keep. I will not intentionally engage in or participate in any form of malicious harm to another person or animal.

Informed Consent for Data Used

All subjects gave their informed consent for inclusion before they participated in the study. The study was conducted following the Declaration of Helsinki.

I consent to participate in the research project and the following has been explained to me: the research may not be of direct benefit to me. my participation is completely voluntary. my right to withdraw from the study at any time without any implications to me.

Data Availability

- Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.
- The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.
- All data generated or analyzed during this study are included in this published article

Conflict of Interest

The authors declare that they have no conflict of interest.

Competing Interests

The authors have no competing interests to declare that are relevant to the content of this article.

Funding Details

No funding was received to assist with the preparation of this manuscript.

Acknowledgments

I am grateful to all of those with whom I have had the pleasure to work during this and other related Research Work. Each of the members of my Dissertation Committee has provided me extensive personal and professional guidance and taught me a great deal about both scientific research and life in general.

References

[1] S. Bebortta, S. S. Tripathy, U. M. Modibbo, and I. Ali, "An optimal fog-cloud offloading framework for big data optimization in heterogeneous IoT networks," Decision Analytics Journal, vol. 8, pp. 100295, 2023.

[2] R. Rawat, "Logical concept mapping and social media analytics relating to cyber criminal activities for ontology creation," International Journal of Information Technology, vol. 15, no. 2, pp. 893-903, 2023.

[3] F. Itoo, Meenakshi, and S. Singh, "Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection," International Journal of Information Technology, vol. 13, no. 4, pp. 1503-1511, 2021.

[4] A. M. U. D. Khanday, S. T. Rabani, Q. R. Khan, N. Rouf, and M. Mohi Ud Din, "Machine learning based approaches for detecting COVID-19 using clinical text data," International Journal of Information Technology, vol. 12, pp. 731-739, 2020.

[5] S. Khetavath, N. C. Sendhilkumar, P. Mukunthan, S. Jana, S. Gopalakrishnan, L. Malliga, S. R. Chand, and Y. Farhaoui, "An intelligent heuristic manta-ray foraging optimization and adaptive extreme learning machine for hand gesture image recognition," Big Data Mining and Analytics, vol. 6, no. 3, pp. 321-335, 2023.

[6] M. Masdari, and A. Khoshnevis, "A survey and classification of the workload forecasting methods in cloud computing," Cluster Computing, vol. 23, no. 4, pp. 2399-2424, 2020.

[7] J. Bi, S. Li, H. Yuan, and M. Zhou, "Integrated deep learning method for workload and resource prediction in cloud systems," Neurocomputing, vol. 424, pp. 35-48, 2021.

[8] E. Aarthi, S. Jagan, C. P. Devi, J. J. Gracewell, S. B. Choubey, A. Choubey, and S. Gopalakrishnan, “A turbulent flow optimized deep fused ensemble model (TFO-DFE) for sentiment analysis using social corpus data,” Social Network Analysis and Mining, vol. 14, no. 1, pp. 1-16, 2024.

[9] Z. S. Ageed, S. R. Zeebaree, M. M. Sadeeq, S. F. Kak, H. S. Yahia, M. R. Mahmood, and I. M. Ibrahim, “Comprehensive survey of big data mining approaches in cloud systems,” Qubahan Academic Journal, vol. 1, no. 2, pp. 29-38, 2021.

[10] R. Turaka, S. R. Chand, R. Anitha, R. A. Prasath, S. Ramani, H. Kumar, S. Gopalakrishnan, and Y. Farhaoui, “A novel approach for design energy efficient inexact reverse carry select adders for IoT applications,” Results in Engineering, vol. 18, pp. 101127, 2023.

[11] D. Saxena, A. K. Singh, and R. Buyya, “OP-MLB: an online VM prediction-based multi-objective load balancing framework for resource management at cloud data center,” IEEE Transactions on Cloud Computing, vol. 10, no. 4, pp. 2804-2816, 2021.

[12] M. E. Karim, M. M. S. Maswood, S. Das, and A. G. Alharbi, “BHyPreC: a novel Bi-LSTM based hybrid recurrent neural network model to predict the CPU workload of cloud virtual machine,” IEEE Access, vol. 9, pp. 131476-131495, 2021.

[13] H. A. Kholidy, “An intelligent swarm based prediction approach for predicting cloud computing user resource needs,” Computer Communications, vol. 151, pp. 133-144, 2020.

[14] I. K. Kim, W. Wang, Y. Qi, and M. Humphrey, “Forecasting cloud application workloads with cloudinsight for predictive resource management,” IEEE Transactions on Cloud Computing, vol. 10, no. 3, pp. 1848-1863, 2020.

[15] M. Junaid, A. Sohail, R. N. B. Rais, A. Ahmed, O. Khalid, I. A. Khan, S. S. Hussain, and N. Ejaz, “Modeling an optimized approach for load balancing in cloud,” IEEE access, vol. 8, pp. 173208-173226, 2020.

[16] S.-Y. Hsieh, C.-S. Liu, R. Buyya, and A. Y. Zomaya, “Utilization-prediction-aware virtual machine consolidation approach for energy-efficient cloud data centers,” Journal of Parallel and Distributed Computing, vol. 139, pp. 99-109, 2020.

[17] S. Negi, M. M. S. Rauthan, K. S. Vaisla, and N. Panwar, “CMODLB: an efficient load balancing approach in cloud computing environment,” The Journal of Supercomputing, vol. 77, no. 8, pp. 8787-8839, 2021.

[18] Z. Li, Y. Li, Y. Liu, P. Wang, R. Lu, and H. B. Gooi, “Deep learning based densely connected network for load forecasting,” IEEE Transactions on Power Systems, vol. 36, no. 4, pp. 2829-2840, 2020.

[19] G. Rjoub, J. Bentahar, O. Abdel Wahab, and A. Saleh Bataineh, “Deep and reinforcement learning for automated task scheduling in large-scale cloud computing systems,” Concurrency and Computation: Practice and Experience, vol. 33, no. 23, pp. e5919, 2021.

[20] A. H. A. Al-Jumaili, R. C. Muniyandi, M. K. Hasan, J. K. S. Paw, and M. J. Singh, “Big data analytics using cloud computing based frameworks for power management systems: Status, constraints, and future recommendations,” Sensors, vol. 23, no. 6, pp. 2952, 2023.

[21] J. Dogani, F. Khunjush, and M. Seydali, “Host load prediction in cloud computing with discrete wavelet transformation (dwt) and bidirectional gated recurrent unit (bigru) network,” Computer Communications, vol. 198, pp. 157-174, 2023.

[22] E. Patel, and D. S. Kushwaha, “An integrated deep learning prediction approach for efficient modelling of host load patterns in cloud computing,” Journal of Grid Computing, vol. 21, no. 1, pp. 5, 2023.

[23] J. Dogani, F. Khunjush, M. R. Mahmoudi, and M. Seydali, “Multivariate workload and resource prediction in cloud computing using CNN and GRU by attention mechanism,” The Journal of Supercomputing, vol. 79, no. 3, pp. 3437-3470, 2023.

[24] B. Predić, L. Jovanovic, V. Simic, N. Bacanin, M. Zivkovic, P. Spalevic, N. Budimirovic, and M. Dobrojevic, “Cloud-load forecasting via decomposition-aided attention recurrent neural network tuned by modified particle swarm optimization,” Complex & Intelligent Systems, vol. 10, no. 2, pp. 2249-2269, 2024.

[25] Y. Zhou, "Application of big data and cloud computing for the development of integrated smart building transportation energy systems," Advances in Digitalization and Machine Learning for Integrated Building-Transportation Energy Systems, pp. 223-237: Elsevier, 2024.

[26] S. Guan, C. Zhang, Y. Wang, and W. Liu, “Hadoop-based secure storage solution for big data in cloud computing environment,” Digital Communications and Networks, vol. 10, no. 1, pp. 227-236, 2024.

[27] P. Roopmathi, J. Chockalingam, and A. S. A. Khadir, “A Big Data Virtualization Role In Agriculture And Cloud Computing-Base Smart Agriculture,”

Educational Administration: Theory and Practice, vol. 30, no. 5, pp. 4565-4573, 2024.

[28] S. Kanungo, “AI-driven resource management strategies for cloud computing systems, services, and applications,” World Journal of Advanced Engineering Technology and Sciences, vol. 11, no. 2, pp. 559-566, 2024.

[29] G. Zhou, W. Tian, R. Buyya, R. Xue, and L. Song, “Deep reinforcement learning-based methods for resource scheduling in cloud computing: A review and future directions,” Artificial Intelligence Review, vol. 57, no. 5, pp. 124, 2024.

[30] H. Yuan, J. Bi, S. Li, J. Zhang, and M. Zhou, “An Improved LSTM-Based Prediction Approach for Resources and Workload in Large-scale Data Centers,” IEEE Internet of Things Journal, 2024.