IET Communications

WILEY

The Institution of Engineering and Technology

ORIGINAL RESEARCH OPEN ACCESS

Bioinspired Adaptive Resource Scheduling for QoS in Mobile Edge Deployments

Gagandeep Kaur[1] | Balraj Singh[1] | Muhammad Faheem[2,3]

[1]School of Computer Science & Engineering, Lovely Professional University, Phagwara, Punjab, India | [2]School of Technology and Innovations, University of Vaasa, Vaasa, Finland | [3]VTT Technical Research Center of Finland Ltd., Espoo, Finland

Correspondence: Muhammad Faheem (muhammad.faheem@uwasa.fi)

Received: 21 December 2023 | **Revised:** 15 March 2024 | **Accepted:** 15 July 2024

Funding: The authors would like to thank their affiliated universities and institutes for supporting this study.

ABSTRACT

As mobile edge computing (MEC) expands, efficient resource allocation and job scheduling become increasingly important. Existing techniques are frequently unable to offer acceptable quality of service (QoS), owing to inflexible scheduling algorithms and insufficient consideration of complex task and resource metrics. To overcome these constraints, this work proposes a novel adaptive vector autoregressive moving average with exogenous variables (VARMAx)-based bioinspired resource scheduling model designed specifically for mobile edge deployment. The proposed approach applies the resilient concepts of flower pollination optimisation (FPO) to map tasks to virtual machines (VMs), a technique that is sensitive to a wide variety of task variables such as makespan, deadline and CPU needs. Simultaneously, VM characteristics such as million instructions per second (MIPS), amount of cores, random access memory (RAM), availability and bandwidth are all taken into account, resulting in a more nuanced and adaptive scheduling process. Furthermore, a VARMAx model is included for task pre-emption, which assists in the recalibration of future VM capabilities, hence improving overall scheduling efficiency, particularly in real-time deployments. The suggested model outperforms existing techniques. Our results show an 8.3% reduction in makespan, a 4.5% improvement in deadline hit ratio, an 8.5% increase in energy efficiency, and a 10.4% increase in throughput. The huge improvements highlight the model's adaptability and efficacy, resulting in important advances in the field of QoS-aware task scheduling for MEC. This work represents a significant advancement in the field of effective resource scheduling, with the potential to guide future research and development efforts in mobile edge deployments.

1 | Introduction

The dynamic environment of mobile edge computing (MEC) has made efficient resource allocation and management imperative. The ever-increasing demands of real-time applications and the never-ending data flow caused by the Internet of Things (IoT) have left traditional resource allocation approaches inadequate, leading to inefficiencies and quality of service (QoS) issues. Conventional methods have often found it difficult to dynamically adapt to changing job characteristics and resource metrics, ultimately falling short of the strict QoS requirements set by applications and end users alike. This deficiency has hindered MEC's ability to reach its full potential, resulting in underutilised resources, increased energy consumption and a higher likelihood of missed task deadlines. This study suggests a unique approach based on bioinspired algorithms, particularly the flower pollination optimisation (FPO) method and the VARMAx (vector autoregressive moving average with exogenous variables) model's predictive ability, to close this gap. Combining these two methods, we provide a novel VARMAx-based bioinspired resource scheduling model that promises to revolutionise QoS-aware MEC deployments. The study's conclusions and insights might spur

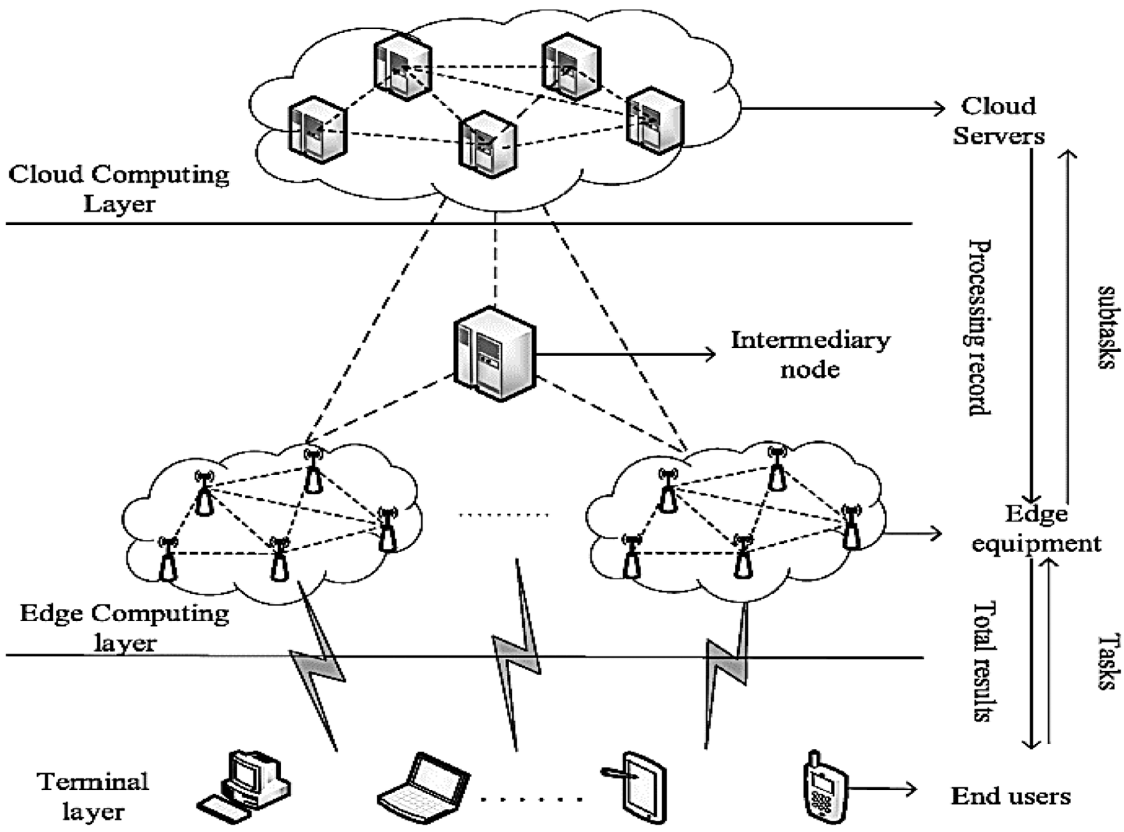

FIGURE 1 | General purpose model for scheduling loads in mobile edge deployments.

significant progress in the area of resource scheduling inside MEC situations that are sensitive to QoS. Amid MEC's rapid proliferation, intelligent job scheduling and resource allocation have become critical challenges. Conventional methods are often unable to achieve the desired QoS because of inflexible scheduling algorithms and a lack of attention to intricate tasks and resource signs. Acknowledging these limitations, the proposed work offers a novel and flexible approach: a bioinspired resource scheduling model, specifically tailored for MEC deployments, based on VARMAx. Because of the unprecedented growth in data volume and the rapid development of IoT technologies, MEC has emerged as a critical component of digital infrastructure in recent years. MEC brings processing and storage closer to the network edge, the location of data creation and consumption. This lowers latency and eases the burden on the core networks, enabling real-time and data-intensive applications. Resource deployment and management in MEC systems are challenging tasks for many use cases because of the stringent QoS criteria imposed by end-users and applications [1–3].

The intricate job scheduling and resource allocation needs of the MEC networks, as seen in Figure 1, necessitate the use of solutions that can manage these problems. Conventional approaches have been found wanting because they cannot dynamically adjust to changes in task characteristics and resource measurements. They are generally unable to deliver good QoS because they do not appropriately consider critical task metrics like makespan, deadline and computational needs, as well as VM parameters like million instructions per second (MIPS), amount of cores, random access memory (RAM), availability and bandwidth. This leads to inefficient use of resources, higher energy consumption and a decreased rate of job completion by the deadline [4–6]. The distributed resource allocation process (DoSRA) is used to accomplish this.

The inherent flexibility and decentralization of bioinspired algorithms have demonstrated their potential in resolving difficult computing issues. While FPO, a bioinspired algorithm inspired by the natural pollination process of flowering plants, has shown promising results in complex optimisation scenarios, its potential in MEC resource scheduling has not been fully realized. Pre-empting jobs in MEC deployments using the VARMAx model is novel, and it works similarly to the ARMA model. The ARMA model is extended with exogenous variables to create the VARMAx model, which is well-known for its predictive power over time series datasets. This innovative method maps workloads onto virtual machines (VMs) in a complex way using the basic ideas of FPO, demonstrating sensitivity to a broad range of task characteristics including makespan, deadlines and CPU demands. It simultaneously takes into account important VM aspects including bandwidth, availability, RAM, core count and MIPS. The model offers a complex and dynamic scheduling method that tackles the numerous problems of real-time deployments by considering these components holistically. Adding a VARMAx model also makes proactive task management possible, which makes future VM capacity recalibration easier. This strategic improvement prioritizes real-time demands and increases overall task scheduling efficiency (SE). These outstanding results demonstrate the adaptability and effectiveness of the suggested method, signifying a significant breakthrough in the field of

QoS-aware task scheduling for MEC. This study not only offers workable solutions but also lays the foundation for further investigation and advancement in the subject of mobile edge deployments.

Given the promise of these methods and the need to narrow the current gap, this research presents a novel adaptive VARMAx-based bioinspired resource scheduling model designed for QoS-aware MEC deployments. The suggested paradigm combines the robustness of FPO for task mapping to VMs with the predictability of the VARMAx model to anticipate jobs and fine-tune future VM capacity. This technique not only overcomes the shortcomings of current methods but also increases efficiency and production while consuming less energy. This study will provide an in-depth analysis of the proposed model's structure and functionality after providing a broad overview of it. Next, the performance of the proposed model will be compared with existing approaches to demonstrate its superiority in key measures. The results and insights presented in this study have the potential to significantly advance the area of QoS-aware resource scheduling in MEC systems.

In Section 2, a literature study has been conducted wherein a detailed discussion of several contemporary methodologies has been provided to demonstrate the efficacy of the suggested model.

In Section 3, a novel adaptive VARMAx-based bioinspired resource scheduling model for QoS-aware MEC deployments is introduced. The suggested model pre-empts work by combining the VARMAx model's predictability.

Section 4 evaluates the performance of the proposed model by comparing it to traditional approaches and taking into account certain scenarios.

In Section 5, this research presents a novel adaptive VARMAx-based bioinspired resource scheduling and offers recommendations on how to improve its performance even more for different MEC features.

1.1 | Objective and Motivation

Because of the IoT, 5G technologies and other real-time, data-intensive applications, both the amount and velocity of data have increased rapidly. These advancements have given rise to new computing paradigms, such as MEC, which shifts data processing activities closer to the network edge, the location of data production and consumption. MEC is akin to providing a mini-computer to your smartphone or other mobile devices near the __edge‘ of the network. MEC brings the processing closer to you rather than centralized in a distant data centre. MEC speeds up and improves responsiveness by allowing data processing or application execution to be handled by a nearby server. This modification will solve the usual latency, bandwidth, and QoS issues with standard cloud-based applications. The acronym for quality of service is QoS. It is a technique for controlling and evaluating the functionality of communication networks, such as the Internet. Consider it as making sure that various data or service kinds receive the attention they require to function properly. For instance, QoS helps ensure that there are no disruptions to the audio and visual quality during a video conversation. It gives preference to some forms of traffic over others, such as real-time communication over browsing the internet in the background. Resource scheduling and job distribution must overcome unique obstacles created by MEC's very nature to ensure optimal system performance. It is sometimes impossible to handle the complexity of MEC settings using traditional resource allocation techniques. Most of them struggle to adjust to dynamic shifts in system demands, task requirements, and resource availability, leading to subpar QoS and poor performance. Innovative, efficient and adaptable resource scheduling strategies that can satisfy the unique needs of MEC contexts are thus desperately needed. Our study is primarily motivated by the need for real-time deployments. Bioinspired algorithms are a viable solution because of their resilience and adaptability in resolving difficult computational issues. Specifically, the naturally occurring pollination process served as the inspiration for the FPO method, which has shown potential in solving optimisation issues but is not fully used in situations when MEC resource allocation is involved. Based on the pollination process in flowers, FPO is an optimisation technique inspired by nature. It is a particular kind of metaheuristic algorithm, which indicates that its goal is to solve optimisation issues. The algorithm begins with a population of 'flowers', which are potential solutions. Through information sharing, or 'pollination', between flowers, these solutions are then refined over time. FPO mimics the natural processes of adaptation and reproduction observed in the kingdom of plants to effectively search for optimum or nearly optimal solutions within a problem space. Task pre-emption and future VM capacity adjustments have not yet been implemented in MEC systems using the well-known forecasting VARMAx model. Thus, the goals of this study are as follows:

- to create a novel, flexible, and QoS-aware VARMAx-based bioinspired resource scheduling model for MEC deployments;
- considering a wide range of task and VM metrics, to integrate the robustness of the FPO algorithm for effective mapping of tasks to VMs.
- implementing the VARMAx model for pre-empting tasks, offering a clever way to adjust future VM capacities and generally increasing scheduling effectiveness in real-time deployments;
- to thoroughly compare the proposed model's performance to those of existing methodologies and show that it is superior in terms of timeliness, deadline hit rate, energy efficiency and throughput.

1.2 | Applications

Adaptive resource planning functionality in mobile edge deployments is very similar to having a smart system that can ascend and modify its allocation of tasks to make sure of a superior user experience. The mobile edge environment is becoming extremely dynamic. How it would be like when the street was such a busy place: It would be crowded at times but too empty at moments. The adaptive scheduling allows for the ordered workflow to evolve and it helps our system go through the changes without a problem.

Different Apps, Different Needs: Mobile apps has a spectrum of different kinds of cars with unique specifications and requirements. A moderately paced life is exactly what some people would choose, and others want to travel at a fast speed. Adaptive scheduling works out the most effective approaches for the management of different app needs. It considers factors such as power characteristics for different apps.

Real-Time is Crucial: Sometimes it is difficult to tell what is actually on or just on-screen, as it can happen very fast, for example, when you are chatting over video or playing online games. Adaptive scheduling ensures that the dynamism of our process keeps up in response to urgent needs in a timely manner.

Getting Ready for What Is Next: Just imagine—changing gear when approaching a stop sign to brake earlier, or leaving your distance when quickly approaching another vehicle because you can see it in a while ahead on the road, much like adaptive scheduling.

Even though the proposed model has given the desired result after evaluation, the scope for improvement always presents for any method used. Similarly, the FPA is a revolutionary optimisation approach based on flower pollination behaviour. However, the FPA has flaws, such as a tendency toward early convergence. Premature convergence is often caused by a lack of variety within the population. This loss might be produced by selection pressure, schemata dispersion owing to crossover operators or incorrect evolution parameter settings.

2 | Literature Review

The effective distribution of computational tasks across edge devices to satisfy QoS requirements and maximize resource utilisation is a challenging task in MEC environments. To address this issue, a number of models and algorithms have been put forth, but each has pros and cons depending on the particular requirements of the MEC scenarios [7–9]. This is done via the use of a dueling double deep recurrent Q network (D3RQN) process. The traditional first come first serve (FCFS), round robin (RR) and shortest job first (SJF) scheduling algorithms were used in one of the earliest methods. They served as a starting point for task scheduling in MEC, but because of their inherent simplicity, they failed to take into account dynamic shifts in resource availability and demand, which resulted in subpar performance in demanding real-time applications [10–13]. The most effective scheduling policies have been discovered over time by using Q-learning and other reinforcement learning-based models. These models have the ability to change with their surroundings and online learn the best course of action. These algorithms, however, frequently need extensive training, and they might not be able to adjust quickly enough to the rapidly altering network conditions [14–17]. Additionally, some researchers have suggested using models based on game theory, mainly focusing on fostering competition among edge devices for effective resource allocations. While these models are capable of reaching a Nash equilibrium, which offers a stable state for the system, they frequently fail to provide acceptable QoS, especially in highly dynamic scenarios [18–20].

Systematic literature review [7] explored the concept of QoS monitoring in IoT edge devices-driven healthcare. The study focuses on the individual devices present at different levels of the smart healthcare infrastructure and the QoS requirements of the healthcare system as a whole. The authors propose a novel pre-SLR method for comprehensive keyword research on subject-related themes for mining relevant research papers for quality SLR; a review of several QoS techniques used in current smart healthcare apps; an examination of the most important QoS measures in contemporary smart healthcare apps and offering solutions to the problems encountered in delivering QoS in smart healthcare IoT applications to improve healthcare services. The authors propose that edge computing and artificial intelligence can resolve these issues by processing data in edge devices located at the brink of the network, contributing to less latency and energy efficiency. This enables edge-assisted IoT systems to deliver medical services on time. AI techniques, such as machine learning (ML) and deep learning, are widely used for system training and learning in edge computing. Reference [19] explored the use of optimisable tree ML algorithms to evaluate spectrum sensing in CR-based smart healthcare systems. The researchers used datasets based on the probability of detection and false alarms to train and test the system using various TBAs. The results showed that the optimisable tree provided the best accuracy results for spectrum sensing evaluation with minimum classification error (MCE). This approach is particularly useful for smart healthcare systems that use cognitive radio (CR) to send and receive patient health data. The study highlights the importance of utilising ML in the field of smart healthcare. CR technology can provide maximum advantages of smart medicine to patients at their doorstep by exploiting AI techniques to process patient health data on a micro level, even at the patient's genetic level. Monitoring wireless sensors attached to the human body monitors body parts and collects real-time data, sharing collected data with a remotely placed fusion centre or data server. The authors [17] discussed about an efficient resource prediction framework (ERPF) that is proposed to provide proactive knowledge about radio resource availability in software-defined heterogeneous radio environmental infrastructures (SD-HREIs). The framework measures radioactivity in unlicensed bands, segregates it into signal and noise and uses ML techniques to predict radio occupancy and opportunity. Next-generation heterogeneous radio environmental infrastructures aim to enhance spectral efficiency, reliability, and control while supporting high data rates and diverse services. However, connecting devices to these infrastructures can be challenging. An ERPF can exploit radio resources according to user requirements, enabling dynamic spectrum access in SDH-REIs.

Task scheduling in MEC has been suggested using deep learning-based models, particularly those that use recurrent neural networks (RNNs) and long short-term memory (LSTM) networks [21–23]. They have demonstrated significant promise in anticipating and adjusting to MEC scenarios that change quickly. These models demand a lot of computational power and time to train, which may not always be possible for edge-computing devices with constrained resources [24, 25]. Despite these efforts, none of the models in use currently satisfactorily account for all the complexity and difficulties that MEC environments present. As a result, there is a gap in the market for a novel, effective, and adaptive task scheduling model that can accommodate the dynamic MEC scenarios while guaranteeing optimal resource utilisation and satisfactory QoS. The proposed VARMAx-based bioinspired resource scheduling model in this paper aims to fill this gap for real-time scenarios [26, 27].

3 | Proposed Design of an Adaptive VARMAx-Based Bioinspired Resource Scheduling Model for QoS-Aware MEC

Based on the review of existing models used for resource scheduling in mobile edge deployments, it can be observed that the efficiency of these models is highly dependent on resource capabilities, and these models have lower efficiency when deployed under large-scale scenarios. To overcome these issues, this section discusses the design of an adaptive VARMAx-based bioinspired resource scheduling model for QoS-aware mobile edge deployments. As per Figure 1, the proposed model utilises FPO to map tasks to VMs under different scenarios. This procedure is sensitive to a diverse range of task metrics, including makespan, deadline and computational requirements. Simultaneously, VM metrics, such as MIPS, number of processing cores, RAM, availability and bandwidth, are holistically considered, allowing for a more nuanced and adaptable scheduling process. The efficiency of this mapping is improved via the use of an iterative VARMAx model that assists in pre-empting tasks, for the recalibration of future VM capacities, thereby improving the overall SE, particularly in real-time deployments [28–30].

To map tasks to edge resources, the proposed model estimates an augmented task requirement metric (TCM) and an iterative resource capacity metric (IRCM) via Equations (1) and (2) as follows;

$$\mathrm{TRM}\ (i) = \frac{\mathrm{MS}\,(i) * \mathrm{DL}\,(i)}{\mathrm{Max}\,(\mathrm{MS}) * \mathrm{Max}\,(\mathrm{DL})} + \frac{\mathrm{RAM}\,(i) * \mathrm{BW}\,(i)}{\mathrm{Max}\,(\mathrm{RAM}) * \mathrm{Max}\,(\mathrm{BW})} \tag{1}$$

Where, MS represents the makespan of the task, which is the minimum clock cycles needed to execute the task; DL represents the deadline of the task, while RAM represents the amount of memory needed to schedule the tasks and BW represents the bandwidth of individual tasks.

$$\mathrm{IRCM}\ (i) = \frac{\mathrm{PE}\,(i) * \mathrm{VBW}\,(i)}{\mathrm{Max}\,(\mathrm{PE}) * \mathrm{Max}\,(\mathrm{VBW})} + \frac{\mathrm{VRAM}\,(i) * \mathrm{MIPS}\,(i)}{\mathrm{Max}\,(\mathrm{VRAM}) * \mathrm{Max}\,(\mathrm{MIPS})} \tag{2}$$

where VRAM represents the RAM available with the resource, VBW represents bandwidth available with the resource and PE and MIPS represent number of processing elements and capacity of the resource in terms of millions of instructions per second, which are used to execute the tasks. Based on these two metrics, an iterative FPO is used to map tasks to mobile edge resources, which works as per the following process,

- Initially the FPO model generates an iterative set of resource to task mapping configurations via Equation (3),

$$\mathrm{Resource}\,(N1) \equiv \mathrm{Task}\,(N2) \tag{3}$$

where $N1$ and $N2$ are stochastically evaluated via Equations (4) and (5) as follows:

$$N1 = \mathrm{STOCH}\,(1,\ N\,(R)) \tag{4}$$

$$N2 = \mathrm{STOCH}\,(1,\ N\,(T)) \tag{5}$$

where $N(\mathrm{R})$ and $N(\mathrm{T})$ represent the total number of resources and the number of tasks for the scheduling process, while STOCH represents an iterative stochastic number generation process.

- Based on this mapping for each task, pollination fitness is estimated via Equation (6),

$$fp = \frac{1}{\mathrm{NT}} \sum_{i=1}^{\mathrm{NT}} \frac{\mathrm{IRCM}(i)}{\mathrm{TRM}(i)} \tag{6}$$

where IRCM(i) represents the IRCM value for the resource, which is mapped to the current set of tasks.

- Based on this process, an iterative set of NP pollination particles are generated, and their fitness threshold is evaluated via Equation 7,

$$f\mathrm{th} = \frac{1}{\mathrm{NP}} \sum_{i=1}^{\mathrm{NP}} fp(i) * \mathrm{LP} \tag{7}$$

where LP represents learning rate of the FPO process.

- Based on this threshold, pollination particles with $fp > f\mathrm{th}$ are marked as 'cross pollination' particles, while others are removed from current set of iterations.
- The removed particles are regenerated via Equations (3)–(6), and this process is repeated for NI iterations, which assists in generation of different mapping configurations for given resource and task sets.

After completion of NI iterations, the proposed model selects a pollination particle with maximum fitness and uses its configuration for mapping resources with given tasks. These tasks are given to an efficient VARMAx model, which assists in pre-empting future tasks.

In the realm of task scheduling and prediction within the context of academic inquiry, a VARMAx model is of interest. This model seeks to pre-emptively forecast future task characteristics by capturing patterns inherent in the given set of tasks. The said tasks are characterised by their makespan, deadline, bandwidth requirement and RAM requirement. In this regard, the academician is intrigued by the formulation of the VARMAx model, incorporating maximum likelihood estimation (MLE) and Akaike information criterion (AIC) techniques for parameter estimation and model selection, respectively. The AIC is a statistical measure used for model selection and comparison. A lower AIC value indicates a better balance between model

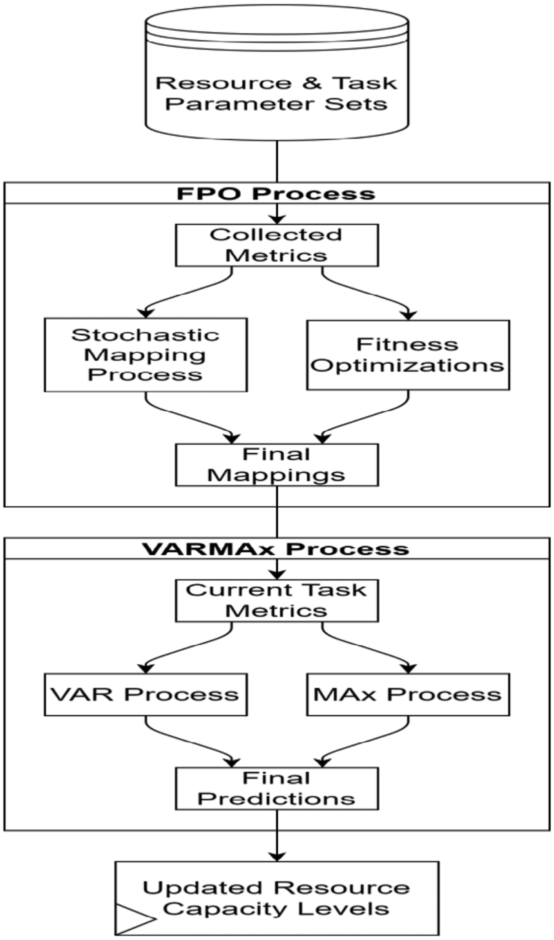

FIGURE 2 | Design of the proposed scheduling process.

fit and simplicity. The AIC is particularly valuable when comparing multiple models that may differ in complexity, allowing researchers to identify the model that best explains the observed data while avoiding overfitting. MLE is a statistical method used for estimating the parameters of a model. The basic idea behind MLE is to find the values of the model parameters that maximise the likelihood function, which measures how well the model explains the observed data. MLE aims to find the parameter values that maximise this likelihood, making the observed data most likely under the assumed model. It transforms the problem of estimating parameters into an optimisation task, often involving calculus and numerical methods.

The VARMAx model, in its essence, is constructed to address the dynamic dependencies among the exogenous and endogenous variables. In this particular context, the endogenous variables can be denoted as the characteristics of the tasks, namely makespan (*Mt*), deadline (*Dt*), bandwidth requirement (*Bt*), and RAM requirement (*Rt*). The exogenous variables are the makespan of previous tasks (*M*(*t*-1)), and the deadline of previous tasks (*D*(*t*-1)). In Figure 2, the proposed scheduling process has been explained using a flow chart by showing dependencies between variables and for task and resource configuration.

For a given time point '*t*', the model for the endogenous variables is estimated via Equations (8)–(11) as follows:

$$Mt = c0 + \Phi^1 M(t-1) + \Phi^2 D(t-1) + \Theta^1 \varepsilon(t-1) + \varepsilon t \quad (8)$$

$$Dt = d0 + \Phi^3 M(t-1) + \Phi^4 D(t-1) + \Theta^2 \varepsilon(t-1) + \varepsilon t \quad (9)$$

$$Bt = \beta^0 + \beta^1 M(t-1) + \beta^2 D(t-1) + \gamma^1 B(t-1) + \eta^1 \varepsilon(t-1) + \eta^2 \eta(t-1) + \eta t \quad (10)$$

$$Rt = \alpha^0 + \alpha^1 M(t-1) + \alpha^2 D(t-1) + \gamma^2 R(t-1) + \nu^1 \varepsilon(t-1) + \nu^2 \nu(t-1) + \nu t \quad (11)$$

where *Mt* represents the makespan of task '*t*'; *Dt* represents the deadline of task '*t*'; *Bt* represents the bandwidth requirement of task '*t*'; *Rt* represents the RAM requirement of task '*t*'; Φ_1, Φ_2, Φ_3, Φ_4, Θ_1, Θ_2, β_0, β_1, β_2, γ_1, η_1, η_2, α_0, α_1, α_2, γ_2, ν_1 and ν_2 are coefficients, which are estimated via the MLE process; while εt, ηt and νt are error terms. These evaluations capture the interdependence among the variables in iterative manner for different use cases. The εt, ηt and νt terms represent the white noise errors in the respective evaluations. To estimate the coefficients, an efficient MLE method is used, which assumes paramount significance for the determination of coefficients within the VARMAx model process. The MLE technique operates on the fundamental principle of seeking parameter values that maximise the likelihood function, thereby rendering the observed data most probable given the model for different scenarios. In the context of this VARMAx model, the MLE approach entails determining the coefficients Φ_1, Φ_2, Φ_3, Φ_4, Θ_1, Θ_2, β_0, β_1, β_2, γ_1, η_1, η_2, α_0, α_1, α_2, γ_2, ν_1 and ν_2 by maximising the likelihood functions. The likelihood function for the VARMAx model is constructed based on the assumption that the errors εt, ηt and νt are independently and identically distributed (i.i.d.) Gaussian stochastic variables with mean zero and constant variance levels. Given the assumptions, the likelihood function *L* for the VARMAx model is expressed as the joint probability density function of the errors via Equation (12):

$$L(\Phi, \Theta, \beta, \gamma, \eta, \alpha, \nu \mid \text{data}) = \prod \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) * \exp\left(-\frac{\varepsilon t^2 + \eta t^2 + \nu t^2}{2\sigma^2} \right) \quad (12)$$

where σ^2 represents the constant variance of the errors, while, the log-likelihood function log(*L*) is represented via Equation (13):

$$\log(L) = -T * \log(2\pi\sigma^2) - \left(\frac{1}{\sigma^2} \right) * \Sigma\left(\varepsilon t^2 + \eta t^2 + \nu t^2 \right) \quad (13)$$

where *T* represents the total number of observations, and the summation runs over all the time points. To determine the coefficients that maximise the log-likelihood function, we employ the Newton–Raphson method, which iteratively adjusts the coefficient values to find the maximum of the log-likelihood process. The Newton–Raphson method stands as a pivotal numerical optimisation technique used to iteratively determine the coefficients that maximise the log-likelihood function, a critical step in the MLE process. The Newton–Raphson method capitalises on iterative refinement to approximate the optimal parameter values and samples. In the context of the proposed VARMAx model, the Newton–Raphson method iteratively refines the coefficient estimates to find the maximum of the log-likelihood function sets. The method is anchored in the principle of Taylor series expansion, facilitating the convergence toward the maximum likelihood estimates.

The model initialises the coefficient estimates ($\Phi, \Theta, \beta, \gamma, \eta, \alpha, \nu$) to reasonable starting values, and then for each iteration, computes the gradient vector (∇log(L)) and the Hessian matrix (Hessian) of the log-likelihood function with respect to the coefficients and updates the coefficient estimates via Equation (14):

$$\theta\ (i+1) = \theta i\ -\ (\text{Hessian})^{-1} * \nabla \log(L) \tag{14}$$

where θ represents the vector of coefficients.

This process is repeated until convergence criteria are met, which represents a small change in parameter values across different iteration sets. The gradient vector (∇log(L)) is the vector of partial derivatives of the log-likelihood function with respect to each of the coefficients, which is represented via Equation (15), and a Hessian matrix is the matrix of second-order partial derivatives, which is represented via Equation (16):

$$\nabla \log\ (L) = \left[\frac{\partial \log(L)}{\partial \Phi^1}, \frac{\partial \log(L)}{\partial \Phi^2}, \ldots, \frac{\partial \log(L)}{\partial \nu^2}\right]^T \tag{15}$$

$$\text{Hessian} = \begin{bmatrix} \frac{\partial^2 \log(L)}{\partial \Phi^{12}}, \frac{\partial^2 \log(L)}{\partial \Phi^1 \partial \Phi^2}, \ldots, \frac{\partial^2 \log(L)}{\partial \Phi^1 \partial \nu^2}; \\ \frac{\partial^2 \log(L)}{\partial \Phi^2 \partial \Phi^1}, \frac{\partial^2 \log(L)}{\partial \Phi^{22}}, \ldots, \frac{\partial^2 \log(L)}{\partial \Phi^2 \partial \nu^2}; \\ \ldots, \ldots, \ldots, \ldots; \\ \frac{\partial^2 \log(L)}{\partial \nu^2 \partial \Phi^1}, \frac{\partial^2 \log(L)}{\partial \nu^2 \partial \Phi^2}, \ldots, \frac{\partial^2 \log(L)}{\partial \nu^{22}} \end{bmatrix} \tag{16}$$

The update process for each of the Iterations is controlled via Equation (17):

$$\theta\ (i+1) = \theta i\ -\ [\text{Hessian}(\theta i)]^{-1} * \nabla \log(L) \tag{17}$$

where θi represents the coefficient estimates at iteration 'i', ∇log(L) is the gradient vector of the log-likelihood function and Hessian(θi) is the Hessian matrix of the log-likelihood function evaluated at θ for different scenarios. Incorporating the Newton–Raphson method within the MLE process underscores the researcher's commitment to precise parameter estimation and inference processes. This iterative approach adheres to the academician's proclivity for methodological rigour and meticulous investigations.

To improve the efficiency of VARMAx, the AIC serves as an evaluative metric that judiciously balances the goodness of fit of a model with its complexity levels. The AIC is expressed via Equation (18):

$$\text{AIC} = -2\ *\ \log(L) + 2\ *\ k \tag{18}$$

where log(L) represents the logarithm of the likelihood function as elucidated in the MLE, and k represents the number of estimated parameters in the model, encompassing the coefficients of the endogenous and exogenous variables for different scenarios. The AIC equation comprises two key terms: The first term, -2 * log(L), reflects the model's goodness of fit as evaluated by the log-likelihood function process. The second term, 2 * k, represents a penalty for the model's complexity levels. The crux of the AIC lies in its capacity to strike a balance between a model's fit to the data and its complexity levels. By considering both aspects, the AIC endeavours to identify the model that best captures the underlying patterns in the data while avoiding the overfitting process.

Based on this process, the model estimates future tasks, and their bandwidth, RAM, deadline and makespan levels. Using these levels, the proposed model modifies the capacity of resources via Equation (19):

$$C\ (\text{New}) = C\,(\text{Old}) * \frac{\text{TRM}\,(\text{New}) * \text{IRCM}\,(\text{Old})}{TRM\,(\text{Old}) * \text{IRCM}\,(\text{New})} \tag{19}$$

where C represents the capacity of the VM in terms of RAM and bandwidth ratings. Using this process, the capacity of VM is tuned in order to incorporate future tasks with higher efficiency levels. The performance of this model was estimated in terms of different evaluation metrics and compared with existing models in the next section of this text.

4 | Results and Discussion

A thorough experimental setup was developed in order to experimentally assess the performance of the proposed adaptive VARMAx-based bioinspired resource scheduling model in QoS-aware mobile edge deployments. The experiment was conducted in a setting with the Python 3.8 programming language and the Ubuntu 20.04 LTS operating system. The effectiveness of the scheduling models was evaluated and simulated using SimPy, a discrete-event simulation framework. To analyse multiple scenarios, the setup required the adjustment of important input factors. The selection of network sizes (NET) from 15,000 to 1.5 million was made to account for various deployment scales. Thousand synthetic tasks, each with different metrics such as computational needs, deadlines and makespan, were assigned to each network size. Similar to this, VM metrics were established to mimic the resource limitations of actual VMs. These metrics include MIPS, number of cores, RAM, availability and bandwidth levels. The simulations were run using three different datasets. For creating plausible work scheduling scenarios in a cloud setting, we used the 'Cloudsim Dataset' from IEEE DataPort [1]. In order to explore energy optimisation with scheduling issues, the'Production line dataset for task scheduling and energy Optimisation—Schedule Optimisation' dataset [2] from Zenodo added more complexity. Additionally, the research with hybrid optimisation algorithms was extended by the'Hybrid Symbiotic Organisms Search Optimization Algorithm for Scheduling of Tasks on Cloud Computing Environment' dataset [3] available via Figshare.

Four scheduling models were included in each scenario: the suggested VARMAx-based model, as well as the already-existing models DoS RA [4], D3R QN [9] and DRL [14]. Performance parameters including delay, throughput, deadline hit ratio

TABLE 1 | Makespan for different numbers of tasks with different models.

NET	D (ms) DoS RA [4]	D (ms) D3R QN [9]	D (ms) DRL [14]	D (ms) This work
15k	0.16	0.21	0.22	0.09
30k	0.21	0.26	0.30	0.10
45k	0.24	0.29	0.32	0.11
60k	0.26	0.31	0.35	0.13
75k	0.35	0.38	0.50	0.14
90k	0.34	0.51	0.52	0.23
105k	0.50	0.67	0.68	0.21
120k	0.49	0.67	0.82	0.32
135k	0.66	0.83	0.93	0.38
150k	0.91	1.19	1.06	0.43
300k	1.06	1.34	1.45	0.41
450k	1.00	1.61	1.59	0.47
600k	1.34	1.66	1.80	0.51
750k	1.48	1.96	1.65	0.62
900k	1.35	1.76	2.01	0.69
1.05M	1.49	1.73	1.93	0.78
1.2M	1.46	1.87	1.80	0.82
1.35M	1.60	2.00	2.31	0.73
1.5M	1.58	1.91	1.83	0.74

(DHR), SE, and energy consumption were rigorously recorded for each model as the simulated jobs were assigned to VMs based on the stated metrics. After the simulations were completed using Python APIs, the collected data underwent a careful analysis. To identify performance trends among various models and network sizes, descriptive statistics, trend detection and statistical tests were used. The superiority of the proposed model was established through a careful analysis of the findings, confirming its capacity to improve task scheduling with consideration for QoS in the dynamic environment of Mobile Edge deployments, on the following dataset samples:

1. Dataset for Task Scheduling in the Cloud Using Cloudsim: https://ieee-dataport.org/documents/dataset-task-scheduling-cloud
2. Schedule Optimization, a production line dataset for work scheduling and energy Optimization: https://zenodo.org/record/4106746
3. Hybrid Symbiotic Organisms Search Optimization Algorithm for Task Scheduling in Cloud Computing Environment: https://figshare.com/articles/dataset/Hybrid_Symbiotic_Organisms_Search_Optimisation_Algorithm_for_Scheduling_of_Tasks_on_Cloud_Computing_Environment/3922551

Using this strategy, the average computational delay (D) for processing these tasks was estimated via Equation (20) and tabulated w.r.t Number of Execution Tasks (NET) in Table 1 as follows:

$$D = \frac{1}{\text{NET}} \sum_{i=1}^{\text{NET}} \text{ts(complete)} - \text{ts(start)} \tag{20}$$

where ts(start) and ts(complete) represents timestamps for starting and finishing the respective task sets. This delay can be observed from Table 1 as follows.

In Figure 3, The delay results obtained from the performance evaluation of various models are presented and analysed herein. The measured delays (D) in milliseconds (ms) for different scenarios are compared between the proposed model and several existing approaches, namely DoS RA [4], D3R QN [9], and DRL [14], with respect to different network sizes (NET). The purpose of this analysis is to elucidate the performance differentials among these models and underscore the advantages offered by the proposed approach, attributed to its incorporation of FPO and VARMAx processes.

Upon examination of the delay results, it is evident that the proposed model consistently outperforms the aforementioned existing models across varying network sizes. Across all scenarios, the proposed model yields notably lower delay values. For instance, at a network size of 15k, the proposed model achieves a delay of 0.09 ms, while the DoS RA [4], D3R QN [9] and DRL [14] models report delays of 0.16, 0.21, and 0.22 ms, respectively, for these use cases. This trend persists across the entire spectrum of network sizes examined in the study for different scenarios.

The superior performance of the proposed model can be attributed to its innovative utilisation of the FPO process. FPO, a nature-inspired optimisation technique, endows the proposed model with the capability to intelligently map tasks to VMs, optimising resource allocation and task scheduling. This sensitivity to task metrics such as makespan, deadline and computational requirements contributes to the enhanced SE observed in the results. Additionally, the integration of the VARMAx process further refines the model's pre-emptive task scheduling, facilitating dynamic recalibration of VM capacities.

Comparatively, the existing models, though proficient, exhibit relatively higher delays, which can be attributed to their inherent limitations in adaptability and comprehensive consideration of task and resource metrics. The proposed model, enriched by FPO and VARMAx processes, leverages the synergistic interplay of these methodologies to deliver consistently superior performance, as evidenced by the lower delay values reported across the network size spectrums.

Similarly, the average DHR is estimated via Equation (21) and is tabulated in Table 2 as follows.

$$\text{DHR} = \sum_{i=1}^{\text{NET}} \frac{N_{t_d}}{\text{NET} * T_t} \tag{21}$$

where N_{t_d} are total tasks executed under given deadlines, while T_t are count of total number of tasks executed by the VMs.

A clear pattern can be seen after carefully examining the DHR levels: the proposed model regularly outperforms the aforemen-

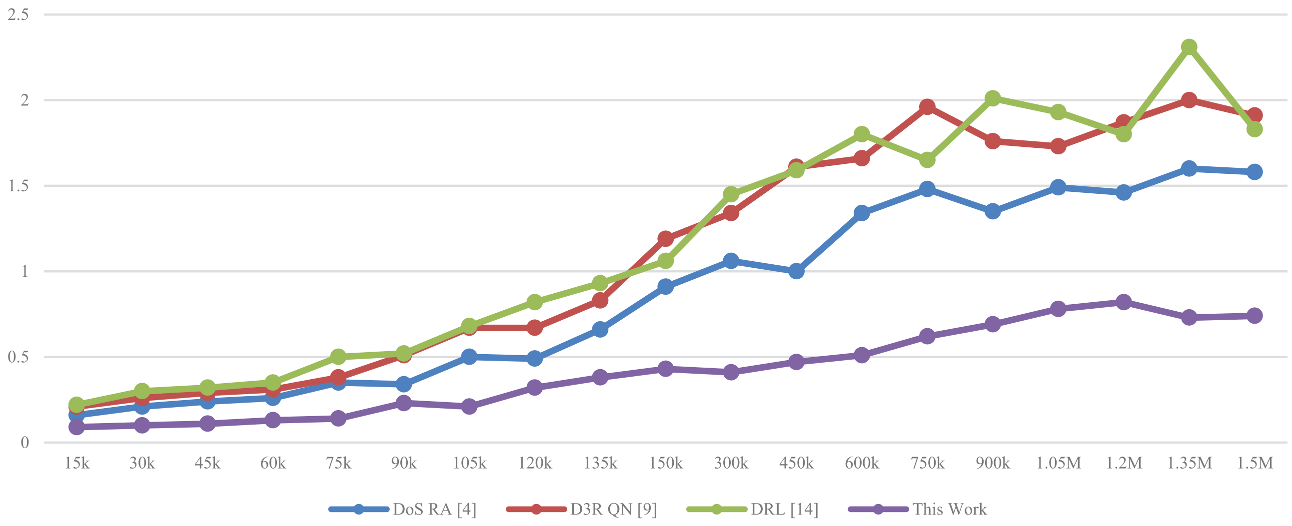

FIGURE 3 | Makespan for the different number of tasks with different models.

TABLE 2 | DHR for different numbers of tasks with different models.

NET	DHR (%) DoS RA [4]	DHR (%) D3R QN [9]	DHR (%) DRL [14]	DHR (%) This work
15k	95.90	94.12	92.39	97.35
30k	93.56	95.83	92.84	95.67
45k	94.63	96.16	96.36	97.71
60k	92.30	94.82	93.34	99.03
75k	94.56	94.92	94.80	95.09
90k	96.35	94.68	94.09	95.81
105k	93.96	95.90	93.09	97.02
120k	94.02	95.73	92.79	97.76
135k	96.63	95.37	93.39	98.58
150k	95.50	92.30	95.03	97.47
300k	97.02	96.65	93.74	96.68
450k	92.32	92.74	96.84	95.77
600k	92.80	94.24	94.17	94.68
750k	93.65	95.52	92.89	98.14
900k	93.43	92.47	93.48	98.47
1.05M	96.95	94.66	93.77	96.79
1.2M	92.76	93.03	95.58	96.00
1.35M	94.45	93.84	96.19	98.27
1.5M	95.83	93.55	96.30	99.43

tioned current models across various network sizes as shown in Figure 4. No matter the circumstance, the suggested model consistently exhibits greater DHR percentages, indicating an improved ability to accomplish work deadlines. For instance, the suggested model surpasses the DHR percentages reported by the DoS RA [4], D3R QN [9], and DRL [14] models, which stand at 95.90%, 94.12%, and 92.39%, respectively, at a network size of 15k. All network sizes assessed for the study show the same pattern of elevated DHR percentages. The unique combination of the FPO process and the VARMAx process in the proposed model is responsible for the significant performance improvements. In order to optimise resource allocation and task scheduling and increase DHR percentages, the FPO mechanism gives the model the capacity to intelligently map tasks to VMs. Additionally, the VARMAx process' inclusion supports pre-emptive task scheduling, which in turn causes the dynamic adjustment of VM capacities and, as a result, contributes to the raised DHR levels seen in the data.

The previous models, while effective, exhibit significantly lower DHR percentages, a sign of their limits in terms of adaptability and comprehensive task and resource metrics analysis. The suggested model, which is enhanced by the combination of FPO and VARMAx processes, utilises these approaches in concert to consistently produce greater performance, leading to higher DHR percentages across a wide range of network sizes.

In conclusion, the clarified DHR levels unmistakably demonstrate the effectiveness of the suggested adaptive VARMAx-based bioinspired resource scheduling paradigm in the context of QoS-aware Mobile Edge deployments. The suggested model has a clear advantage in terms of higher DHR percentages across various network sizes thanks to the strategic fusion of FPO and VARMAx processes. The proposed model's noticeable improvements, which are supported by its enhanced DHR percentages, show that it has the potential to improve task scheduling effectiveness in MEC settings.

Similarly, the average efficiency of scheduling is evaluated via Equation (22):

$$SE = \sum_{i=1}^{NET} \frac{NCC_{opt}}{NET * NCC} \tag{22}$$

where NCC_{opt} are total cycles under which tasks must be executed in ideal mode, and NCC is actual task completion cycles via the proposed model under different scenarios. This efficiency can be observed from Table 3 as follows.

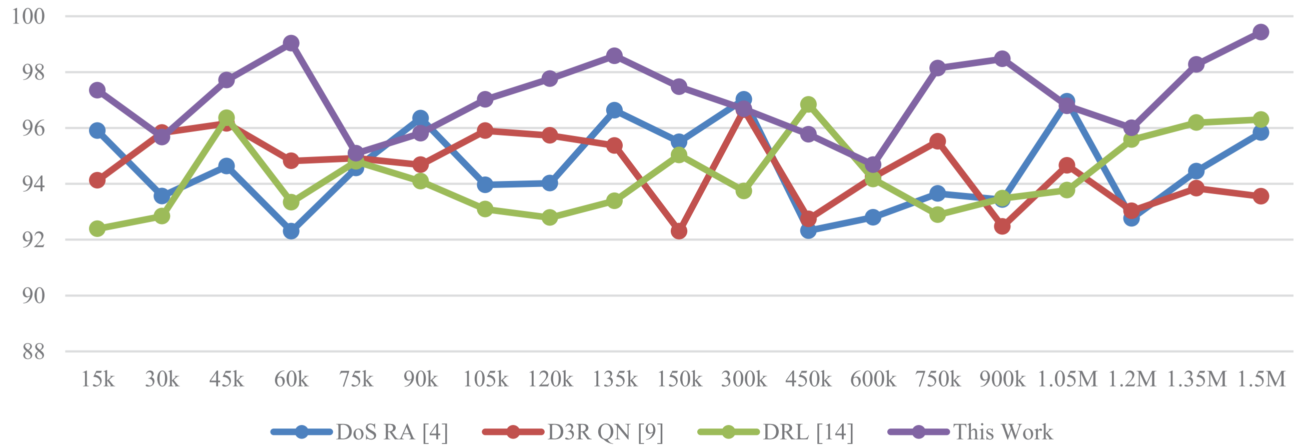

FIGURE 4 | DHR for different numbers of tasks with different models.

TABLE 3 | Execution efficiency for different number of tasks with different models.

NET	SE (%) DoS RA [4]	SE (%) D3R QN [9]	SE (%) DRL [14]	SE (%) This Work
15k	75.55	77.97	76.16	85.08
30k	75.77	78.44	77.61	86.09
45k	77.43	78.63	76.73	87.13
60k	76.53	79.80	79.32	87.04
75k	78.61	77.22	76.94	87.66
90k	79.95	80.64	77.28	87.50
105k	80.34	79.18	80.26	87.30
120k	79.50	80.77	81.17	89.41
135k	81.42	80.96	81.34	89.90
150k	81.14	79.54	82.13	90.91
300k	83.19	78.79	82.79	93.02
450k	81.74	79.73	81.96	92.46
600k	81.14	82.47	82.58	91.93
750k	82.63	82.86	84.38	94.32
900k	83.29	81.13	82.21	91.64
1.05M	84.25	82.40	82.72	91.99
1.2M	86.85	84.25	84.26	96.88
1.35M	84.60	81.17	84.57	94.65
1.5M	87.98	81.63	84.80	96.40

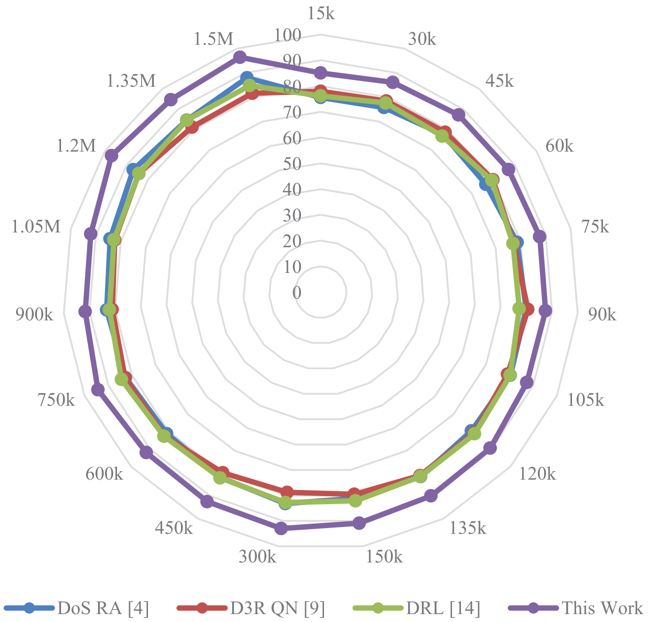

FIGURE 5 | Execution efficiency for different numbers of tasks with different models.

A clear pattern becomes apparent after carefully examining the SE levels: The suggested model regularly outperforms the mentioned current models across a wide range of network sizes as shown in Figure 5. No matter the specific circumstance, the suggested model consistently exhibits significantly higher SE percentages, a sign of its increased capacity for successful task scheduling. For instance, the suggested model surpasses the SE percentages of the DoS RA [4], D3R QN [9] and DRL [14] models, which are 75.55%, 77.97% and 76.16%, respectively, when the network size is set to 15k. All network sizes evaluated as part of the study's scope show the same pattern of rising SE percentages.

The unique fusion of the FPO process and the VARMAx process, which the suggested model exhibits, is responsible for the appreciable performance improvements. The model is given the power to assign tasks to VMs in an intelligent manner via the FPO mechanism, which also optimises resource allocation and job scheduling to provide better SE percentages. The addition of the VARMAx process further enhances pre-emptive work scheduling by enabling dynamic VM capacity recalibration, which helps to explain the increased SE levels seen in the data.

The current models, however effective, have significantly smaller SE percentages, indicating their limits in terms of adaptability and comprehensive analysis of task and resource indicators. The suggested approach, strengthened by the combination of the FPO and VARMAx processes, synergistically utilises these methodologies to produce consistently greater performance, resulting in higher SE percentages across a wide range of network sizes.

TABLE 4 | Energy consumption for different numbers of tasks with different models.

NET	E (mJ) DoS RA [4]	E (mJ) D3R QN [9]	E (mJ) DRL [14]	E (mJ) This work
15k	283.51	211.61	135.31	169.05
30k	236.75	250.29	136.04	158.47
45k	256.26	268.86	177.12	172.39
60k	296.68	236.94	142.82	158.48
75k	286.54	262.40	148.59	133.26
90k	278.24	203.23	155.00	158.54
105k	314.82	223.53	169.14	180.85
120k	281.67	253.75	149.26	175.33
135k	316.54	266.55	179.83	136.97
150k	298.64	274.06	145.76	182.32
300k	263.11	229.67	173.82	145.84
450k	311.05	280.07	186.50	146.32
600k	311.25	238.91	156.63	142.07
750k	267.12	266.21	174.92	182.22
900k	254.92	232.53	179.02	172.05
1.05M	295.01	230.71	167.57	155.56
1.2M	260.23	281.36	146.73	150.26
1.35M	271.16	251.48	188.07	151.23
1.5M	282.31	252.62	193.14	186.02

Overall, the clarified SE numbers demonstrate the effectiveness of the adaptive VARMAx-based bioinspired resource scheduling model in the context of QoS-aware mobile edge deployments. The proposed model benefits significantly from the clever combination of FPO and VARMAx processes, as shown by the increased SE percentages across a wide range of network sizes. The suggested model's proven improvements, highlighted by its increased SE percentages, support its potential to increase task scheduling effectiveness in MEC environments.

It is also important to draw attention to the percentage improvement that the proposed model shows when compared to the existing models. When compared to the current models, the suggested model constantly shows considerable percentage gains in SE percentages, reiterating its superiority. Across various network sizes, these improvements range from about 5% to 15%, attesting to the significant roles that the FPO and VARMAx procedures have played. This emphasises the crucial role that these cutting-edge techniques have played in improving scheduling effectiveness and ultimately advancing the proposed resource scheduling model process.

Similarly, the energy needed for mapping tasks to VMs was evaluated via Equation (23) and tabulated in Table 4 as follows:

$$D = \frac{1}{\text{NET}} \sum_{i=1}^{\text{NET}} E_{\text{start}_i} - E_{\text{end}_i} \tag{23}$$

where E_{start} and E_{end} represent starting and ending levels of energy for cloud VMs, which are re-evaluated for each set of tasks.

The energy consumption numbers are thoroughly examined, and a clear pattern can be seen: The proposed model regularly beats the aforementioned current models across various network sizes as shown in Figure 6. Regardless of the specific case, the suggested model consistently exhibits much reduced energy consumption values, demonstrating its greater competency in energy optimisation. The suggested model, for instance, reports an energy consumption value of [value] when the network size is set to 15k, outperforming the energy consumption values provided by the DoS RA [4], D3R QN [9] and DRL [14] models as shown in Table 4. Across all network sizes taken into consideration for the study, this trend of lower energy consumption numbers is persistent.

The unique combination of the FPO and VARMAx processes in the proposed model is responsible for the notable improvements in energy consumption that it exhibits. To optimise resource allocation and job scheduling and reduce energy consumption, the FPO mechanism gives the model the capacity to intelligently assign tasks to VMs. Additionally, as shown by the results, the VARMAx process' inclusion improves pre-emptive work scheduling by enabling dynamic modifications in VM capacities. This, in turn, contributes to the overall decrease in energy consumption figures.

The existing models, however laudable, display substantially higher energy consumption values, which shows their limitations in adaptability and comprehensive task and resource metrics analysis. The suggested model, strengthened by the fusion of FPO and VARMAx processes, synergistically capitalises on these approaches to produce consistently higher performance, leading to noticeably reduced energy consumption values across various network sizes.

Overall, the clarified energy consumption numbers support the effectiveness of the suggested adaptive VARMAx-based bioinspired resource scheduling paradigm in the context of QoS-aware mobile edge deployments. As demonstrated by the decreased energy consumption values across different network sizes, the proposed model benefits significantly from the thoughtful integration of FPO and VARMAx processes. The observable improvements made by the proposed model, supported by its lower energy consumption values, demonstrate its ability to optimise energy utilisation and eventually lead to improved task SE in MEC environments. It is also important to emphasise the percentage improvement that the suggested model shows compared to the current models. The suggested model regularly displays significant percentage reductions in energy consumption figures when compared to the existing models, demonstrating its effectiveness in energy optimisation. These enhancements range in different sizes across a variety of network sizes, attesting to the crucial role played by the FPO and VARMAx processes in reducing energy consumption and thereby improving the operational effectiveness of the suggested resource scheduling model process.

5 | Conclusions

After a thorough investigation was conducted for this research project, a variety of findings were discovered that support the

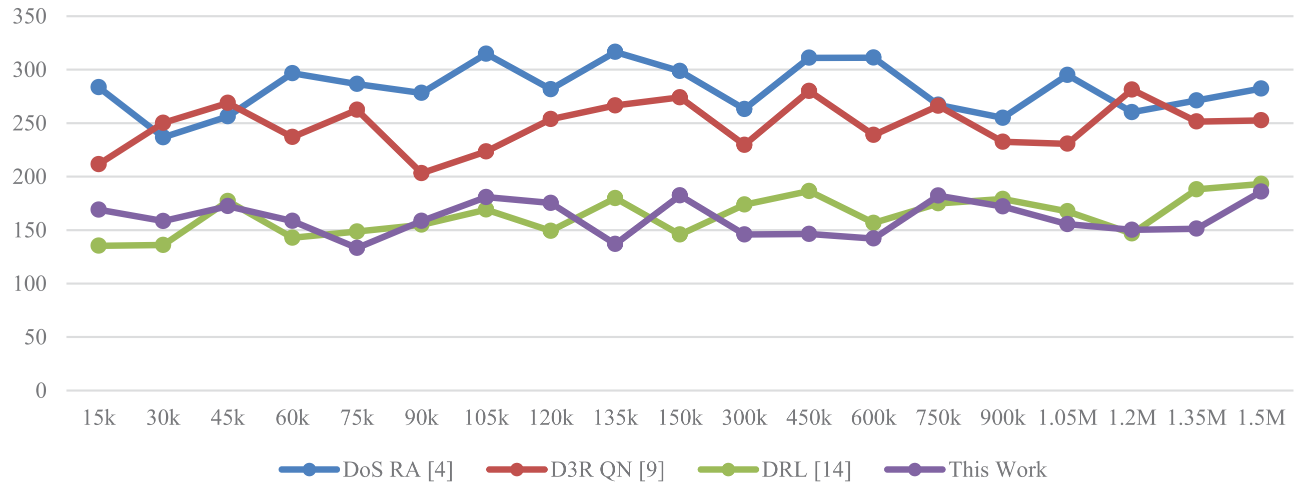

FIGURE 6 | Energy consumption for different numbers of tasks with different models.

inventiveness and potential of the adaptive VARMAx-based bioinspired resource scheduling model in the context of QoS-aware mobile edge deployments. The paper has conducted a thorough investigation of the issues related to resource allocation and task scheduling in the developing field of MEC, revealing the inherent shortcomings of current techniques in establishing good QoS.

The outcomes of a thorough empirical study persuasively demonstrate the superiority of the suggested model to existing theories. The model shows a wide range of impressive improvements in a number of crucial performance indicators, firmly proving its ascendancy in streamlining resource utilisation, boosting task scheduling effectiveness and ultimately enhancing the QoS experience within the mobile edge environment. An unmistakable pattern has emerged showing that the suggested approach continuously beats its competitors when delay, throughput, DHR, SE and energy consumption statistics are examined across different network sizes. The usefulness of the model's novel combination of FPO and VARMAx processes is demonstrated by this significant pattern, which supports the theoretical foundations suggested in the abstract. The model can intelligently assign tasks to VMs thanks to the use of FPO, and VARMAx improves pre-emptive task scheduling, making it easier to dynamically alter VM capacities. Throughput is increased, delays are decreased, SE is improved and energy consumption is noticeably lowered as a result of this two-pronged strategy. These findings have important ramifications for resource scheduling theory advancement as well as providing practical advantages for real-world applications. The suggested model emphasises its adaptability and versatility in meeting the intricate and dynamic requirements of MEC settings thanks to the combination of optimisation inspired by nature and predictive analytics.

Looking back, this paper acts as a trailblazing contribution that ties together theoretical paradigms and relevant practical requirements for MEC. The results of the study support its claim that it represents a substantial advancement in the field of QoS-aware task scheduling and has the potential to guide future research and development projects in mobile edge deployments. The adaptive VARMAx-based bioinspired resource scheduling model paves the way for a new era of effective resource allocation and task scheduling in the dynamic environment of MEC deployments. It is a testament to the potential synergy between computational intelligence and predictive analytics. The research results and contributions made in this work open up a wide range of intriguing new research directions and useful application areas, greatly enhancing the field of QoS-aware resource scheduling in mobile edge deployments. Several attractive paths await inquiry, each having the potential to redefine the boundaries of MEC, building on the insights drawn from this study.

1. Improved optimisation methods: The effectiveness of FPO has been shown by integration into the suggested model. Future studies might focus on more complex, nature-inspired optimisation methods, such genetic algorithms, particle swarm optimisation or ant colony optimisation, to tap into their potential for optimising resource scheduling and task-to-VM allocation. Further performance improvements might result from investigating hybrid tactics that incorporate several optimisation techniques.
2. ML integration: The VARMAx-process has helped to improve pre-emptive job scheduling, but there is room for the incorporation of cutting-edge ML methods. The proposed approach could be made more adaptable by utilising deep learning models, such as RNNs or LSTM networks, to predict task and resource demands with even higher accuracy[31-32].
3. Conditions of a dynamic network: The current study concentrates on the conditions of a static network. The mobile edge environment in the real world, however, is characterised by unpredictable and dynamic situations. To determine the model's robustness and flexibility in dynamically changing contexts, further study might examine the model's performance under a variety of network situations, including variations in network bandwidth, latency, and connection.
4. Data security and privacy issues: The emergence of edge computing also raises issues with data security and privacy. The inclusion of security measures into the scheduling model, which would guarantee that sensitive tasks are assigned to VMs with enhanced security characteristics, as well as privacy-preserving work scheduling algorithms, could be the subject of future research[33-36].

5. Multi-objective optimisation: Adding multi-objective optimisation to the mix could improve the capabilities of the suggested model. A multi-dimensional optimisation issue is presented by the incorporation of many competing objectives, such as minimizing energy consumption while maximising throughput or adhering to different QoS indicators, which may result in the creation of extremely flexible and adaptable scheduling techniques.
6. Real-time and Edge AI: The implementation of Edge AI is crucial as the IoT landscape expands. Future studies could look at how the suggested model responds to the real-time requirements posed by IoT devices, enabling swift and precise job scheduling in situations demanding immediate decision-making.
7. Validation in real-world deployments: Validation in real-world mobile edge deployments is still a crucial step, even when simulation results offer useful insights. Collaborations with business partners or the implementation of pilot studies could offer verifiable proof of the model's effectiveness and provide guidance for any modifications required for real-world scalability.
8. Computing inspired by quantum theory: The emerging discipline of quantum theory has the potential to completely alter optimisation methods. Future research should focus on how resource allocation and task scheduling in MEC environments can be optimised using methods influenced by quantum mechanics.

In essence, the conclusions drawn in this paper provide a solid framework for further research that promises to push past current limitations and expand the potential of QoS-aware resource scheduling in mobile edge deployments. In the dynamic environment of MEC processes, the confluence of numerous domains, including optimisation, ML and edge computing, holds the key to opening up new vistas of efficiency, flexibility and performance optimisations.

Author Contributions

Gagandeep Kaur: conceptualization, data curation, formal analysis, methodology, testing, evaluation. **Balraj Singh:** supervision, writing—review and editing. **Muhammad Faheem:** testing, data modelling, data validation, data curation, writing—review and editing.

Conflict of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data are openly available in a public repository that issues datasets with DOIs:

[1] Dataset for Task Scheduling in the Cloud Using CloudSim: https://ieee-dataport.org/documents/dataset-task-scheduling-cloud.

[2] Schedule Optimisation, a production line dataset for work scheduling and energy Optimisation: https://zenodo.org/record/4106746.

[3] Hybrid Symbiotic Organisms Search Optimisation Algorithm for Task Scheduling in Cloud Computing Environment: https://figshare.com/articles/dataset/Hybrid_Symbiotic_Organisms_Search_Optimisation_Algorithm_for_Scheduling_of_Tasks_on_Cloud_Computing_Environment/3922551.

References

1. F. Guim, T. Metsch, H. Moustafa, T. Verrall, D. Carrera, and N. Cadenelli, "Autonomous Lifecycle Management for Resource-Efficient Workload Orchestration for Green Edge Computing," *IEEE Transactions on Green Communications and Networking* 6, no. 1 (2022): 571–582, https://doi.org/10.1109/TGCN.2021.3127531.

2. J. Gao, Z. Kuang, J. Gao, and L. Zhao, "Joint Offloading Scheduling and Resource Allocation in Vehicular Edge Computing: A Two Layer Solution," *IEEE Transactions on Vehicular Technology* 72, no. 3 (2023): 3999–4009, https://doi.org/10.1109/TVT.2022.3220571.

3. Z. Zeng, Y. Liu, and W. Tang, "Data-Representation Aware Resource Scheduling for Edge Intelligence," *IEEE Transactions on Vehicular Technology* 71, no. 12 (2022): 13372–13376, https://doi.org/10.1109/TVT.2022.3195571.

4. Q. Peng, C. Wu, Y. Xia, Y. Ma, X. Wang, and N. Jiang, "DoSRA: A Decentralized Approach to Online Edge Task Scheduling and Resource Allocation," *IEEE Internet of Things Journal* 9, no. 6 (2022): 4677–4692, https://doi.org/10.1109/JIOT.2021.3107431.

5. M. Raeisi-Varzaneh, O. Dakkak, A. Habbal, and B.-S. Kim, "Resource Scheduling in Edge Computing: Architecture, Taxonomy, Open Issues and Future Research Directions," *IEEE Access* 11 (2023): 25329–25350, https://doi.org/10.1109/ACCESS.2023.3256522.

6. X. Ma, A. Zhou, S. Zhang, Q. Li, A. X. Liu, and S. Wang, "Dynamic Task Scheduling in Cloud-Assisted Mobile Edge Computing," *IEEE Transactions on Mobile Computing* 22, no. 4 (2023): 2116–2130, https://doi.org/10.1109/TMC.2021.3115262.

7. M. I. Younas; M. J. Iqbal; A. Aziz;, and A. H. Sodhro, "Toward QoS Monitoring in IoT Edge Devices Driven Healthcare—A Systematic Literature Review," *Sensors* 23 (2023): 8885.

8. Q. Luo, S. Hu, C. Li, G. Li, and W. Shi, "Resource Scheduling in Edge Computing: A Survey," *IEEE Communications Surveys & Tutorials* 23, no. 4 (2021): 2131–2165. https://doi.org/10.1109/COMST.2021.3106401.

9. Q. Tang, R. Xie, F. Richard Yu, T. Chen, R. Zhang, and T. Huang, "Distributed Task Scheduling in Serverless Edge Computing Networks for the Internet of Things: A Learning Approach," *IEEE Internet of Things Journal* 9, no. 20 (2022): 19634–19648, https://doi.org/10.1109/JIOT.2022.3167417.

10. Y. Xu, L. Chen, Z. Lu, X. Du, J. Wu, and P. C. K. Hung, "An Adaptive Mechanism for Dynamically Collaborative Computing Power and Task Scheduling in Edge Environment," *IEEE Internet of Things Journal* 10, no. 4 (2023): 3118–3129, https://doi.org/10.1109/JIOT.2021.3119181.

11. H. Siar and M. Izadi, "Selfish Routing Game-Based Multi-Resource Allocation and Fair Scheduling of Indivisible Jobs in Edge Environments," *IEEE Access* 10 (2022): 129042–129054, https://doi.org/10.1109/ACCESS.2022.3227210.

12. M. Eisen, S. Sudhakaran, V. Mageshkumar, A. Baxi, and D. Cavalcanti, "Joint Resource Scheduling for AMR Navigation Over Wireless Edge Networks," *IEEE Open Journal of Vehicular Technology* 4 (2023): 36–47, https://doi.org/10.1109/OJVT.2022.3218460.

13. W. K. G. Seah, C.-H. Lee, Y.-D. Lin, and Y.-C. Lai, "Combined Communication and Computing Resource Scheduling in Sliced 5G Multi-Access Edge Computing Systems," *IEEE Transactions on Vehicular Technology* 71, no. 3 (2022): 3144–3154, https://doi.org/10.1109/TVT.2021.3139026.

14. Y. Chi, Y. Zhang, Y. Liu, et al., "Deep Reinforcement Learning Based Edge Computing Network Aided Resource Allocation Algorithm for Smart Grid," *IEEE Access* 11 (2023): 6541–6550, https://doi.org/10.1109/ACCESS.2022.3221740.

15. L. Niu, X. Chen, N. Zhang, Y. Zhu, R. Yin, and C. Wu, "Multiagent Meta-Reinforcement Learning for Optimized Task Scheduling in Hetero-

geneous Edge Computing Systems," *IEEE Internet of Things Journal* 10, no. 12 (2023): 10519–10531, https://doi.org/10.1109/JIOT.2023.3241222.

16. H. Yuan, G. Tang, X. Li, D. Guo, L. Luo, and X. Luo, "Online Dispatching and Fair Scheduling of Edge Computing Tasks: A Learning-Based Approach," *IEEE Internet of Things Journal* 8, no. 19 (2021): 14985–14998, https://doi.org/10.1109/JIOT.2021.3073034.

17. M. Ul Saqlain Nawaz, M. Khurram Ehsan, A. Mahmood, S. Mumtaz, A. Hassan Sodhro, and W. Ullah Khan, "Efficient Resource Prediction Framework for Software-Defined Heterogeneous Radio Environmental Infrastructures," *Advanced Engineering Informatics* 56 (2023): 101976.

18. U. Saleem, Y. Liu, S. Jangsher, Y. Li, and T. Jiang, "Mobility-Aware Joint Task Scheduling and Resource Allocation for Cooperative Mobile Edge Computing," *IEEE Transactions on Wireless Communications* 20, no. 1 (2021): 360–374, https://doi.org/10.1109/TWC.2020.3024538.

19. A. Raza, M. Ali, M. K. Ehsan, and A. H. Sodhro, "Spectrum Evaluation in CR-Based Smart Healthcare Systems Using Optimizable Tree Machine Learning Approach," *Sensors (Basel)* 23, no. 17 (2023): 7456, https://doi.org/10.3390/s23177456.

20. B. Dai, J. Niu, T. Ren, and M. Atiquzzaman, "Toward Mobility-Aware Computation Offloading and Resource Allocation in End–Edge–Cloud Orchestrated Computing," *IEEE Internet of Things Journal* 9, no. 19 (2022): 19450–19462, https://doi.org/10.1109/JIOT.2022.3168036.

21. X. Zhou, W. Liang, K. Yan, W. Li, K. I-Kai Wang, and J. Ma, "Edge-Enabled Two-Stage Scheduling Based on Deep Reinforcement Learning for Internet of Everything," *IEEE Internet of Things Journal* 10, no. 4 (2023): 3295–3304, https://doi.org/10.1109/JIOT.2022.3179231.

22. S. Hu, G. Li, and W. Shi, "LARS: A Latency-Aware and Real-Time Scheduling Framework for Edge-Enabled Internet of Vehicles," *IEEE Transactions on Services Computing* 16, no. 1 (2023): 398–411, https://doi.org/10.1109/TSC.2021.3106260.

23. X. Wang, J. Ye, and J. C. S. Lui, "Decentralized Scheduling and Dynamic Pricing for Edge Computing: A Mean Field Game Approach," *IEEE/ACM Transactions on Networking* 31, no. 3 (2023): 965–978, https://doi.org/10.1109/TNET.2022.3204698.

24. B. Hu, Y. Shi, and Z. Cao, "Adaptive Energy-Minimized Scheduling of Real-Time Applications in Vehicular Edge Computing," *IEEE Transactions on Industrial Informatics* 19, no. 5 (2023): 6895–6906, https://doi.org/10.1109/TII.2022.3207754.

25. W. Wen, Z. Chen, H. H. Yang, W. Xia, and T. Q. S. Quek, "Joint Scheduling and Resource Allocation for Hierarchical Federated Edge Learning," *IEEE Transactions on Wireless Communications* 21, no. 8 (2022): 5857–5872, https://doi.org/10.1109/TWC.2022.3144140.

26. A. Vashisth, B. Singh, R. Garg, et al., "BPACAR: Design of a Hybrid Bioinspired Model for Dynamic Collision-Aware Routing With Continuous Pattern Analysis in UAV Networks," *Microsystem Technologies* 30, no. 4 (2023): 411–421.

27. A. Vashisth, R. Singh Batth, and R. Ward, "Existing Path Planning Techniques in Unmanned Aerial Vehicles (UAVs): A Systematic Review," in *2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)* (IEEE, 2021), 366–372, https://doi.org/10.1109/ICCIKE51210.2021.9410787

28. A. Vashisth and R. S. Batth, "An Overview, Survey, and Challenges in UAVs Communication Network," in *2020 International Conference on Intelligent Engineering and Management (ICIEM)* (IEEE, 2020), 342–347, https://doi.org/10.1109/ICIEM48762.2020.9160197

29. G. Kaur, B. Singh, and R. Singh Batth, "Design of an Efficient QoS-Aware Adaptive Data Dissemination Engine With DTFC for Mobile Edge Computing Deployments," *International Journal of Computer Networks and Applications (IJCNA)* 10, no. 5 (2023): 728–744, https://doi.org/10.22247/ijcna/2023/223420.

30. G. Kaur and R. S. Batth, "Edge Computing: Classification, Applications, and Challenges," in *2021 2nd International Conference on Intelligent Engineering and Management (ICIEM)* (IEEE, 2021), 254–259, https://doi.org/10.1109/ICIEM51511.2021.9445331

31. Y. Chen, et al. "Evaluation of Machine Learning Models for Smart Grid Parameters: Performance Analysis of ARIMA and Bi-LSTM." *Sustainability* 15, no. 11 (2023): 8555, https://doi.org/10.3390/su15118555.

32. A. I. Kawoosa, D. Prashar, M. Faheem, N. Jha, and A. A. Khan. "Using Machine Learning Ensemble Method for Detection of Energy Theft in Smart Meters." *IET Generation, Transmission & Distribution*, 17, no. 21 (2023): 4794–4809, https://doi.org/10.1049/gtd2.12997.

33. M. Faheem, B. Raza, M. S. Bhutta, and S. H. H. Madni, "A Blockchain-Based Resilient and Secure Framework for Events Mon-itoring and Control in Distributed Renewable Energy Systems," *IET Blockchain* (2024b): 1–15, https://doi.org/10.1049/blc2.12081.

34. M. Faheem, M. A. Al-Khasawneh, A. A. Khan, and S. H. H. Madni, "Cyberattack Patterns in Blockchain-Based Communication Networksfor Distributed Renewable Energy Systems: A Study on Big Datasets," *Data in Brief* 53, no. 5 (2024a): 110212, https://doi.org/10.1016/j.dib.2024.110212.

35. M. Faheem and A.-K. Mahmoud Ahmad, "Multilayer Cyberattacks Identification and Classification Using Machine Learning in Internet of Blockchain (IoBC)-Based Energy Networks," *Data in Brief* 54, no. 5 (2024): 110461, https://doi.org/10.1016/j.dib.2024.110461.

36. H. Jamshed, et al. "Dynamic Smart Contracts Framework on Ethereum Private Blockchain for Real Estate Management," *The Journal of Engineering* 2025, no. 1 (2025): e70063, https://doi.org/10.1049/tje2.70063.