

A MULTI-MODEL REGRESSION APPROACH FOR PREDICTING RESOURCE ALLOCATION EFFICIENCY IN IOT-DRIVEN 6G NETWORKS

Hussain AlSalman

Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

ABSTRACT

Enabling healthcare services over emerging Sixth Generation (6G) networks and Internet of Things (IoT) introduces a strict requirement for timely and reliable allocation of medical resources. Prediction of resource allocation efficiency based on rule-based or manual policies often fails to be adaptive to heterogeneous demands and dynamic conditions of IoT networks. To address this challenge, a multi-model regression-based approach is proposed to predict the efficiency of resource allocation for optimizing the MR infrastructures of IoT and 6G networks. The approach consists of data pre-processing, exploratory data analysis, multi-model regression learning, and operational factors interpretation. First, the dataset is loaded and non-informative identifier attributes are removed to reduce noise and improve generalization. Correlation analysis is performed through a heat map plot of numerical features to identify features that are strongly related to the target variable. Extensive experiments are conducted on a publicly available dataset to evaluate the proposed approach according to a number of performance metrics, such as the root mean square error (RMSE), determination coefficient (R-squared), and mean absolute error (MAE). Experimental results showed that the best regression model of proposed approach attains the highest prediction performance compared with other models and state-of-the-art work. In addition to predictive superiority, interpretation of best model's outputs regarding to throughput and utilization of the network is reported to show the association between predicted efficiency, network speed, and utilization status, which will help to design an actionable plan for deploying intelligent allocation policies.

KEYWORDS

Medical Resource Allocation, Internet of Things (IoT), Sixth Generation (6G) Networks, Multi-model Regression, Coefficient Determination

1. INTRODUCTION

Globally, with the rapid advancement of technical and informatics, healthcare systems are undergoing drastic shifts as a result of the integration of advanced computing technologies, wireless communications and sensors, customized within clinical and complex clinical environments [1]. One of the most prominent aspects of this transformation is the Healthcare Internet of Things (IoT) application, which is called the Medical Internet of Things. (IoMT) [2]. This technology has enabled real-time collection and sharing of diagnostic tools, wearable sensors and connected and connected hospital systems. The IoMT provides patient monitoring, prediction analysis, and remote therapy interventions, resulting in remarkable improvements in operational efficiency, patient outcomes, and accessibility of health care systems and services [2]. As the size and complexity of IoT networks expand, traditional health systems face significant challenges in processing the enormous amounts of data that produces quick responses to the resources to be provided to these networks [3].

Recently, as the Internet-related devices have increased, the development of the 5th (5G) and 6th (6G) generations networks have opened the door to new horizons and address the existing challenges [4]. Through the 6G network, unique opportunities in the field have been available, as it enables a large number of devices to be connected and provide highly reliable communications with the lowest level of delay, as well as the provision of safe and smart services to the parties [5]. These features are essential in the healthcare sector, as they are closely related to patient safety, continuity of services and quality of care provided. However, the presence of high-performance networks alone is not sufficient to ensure the effectiveness of health operations. Decisions are still required on how to allocate resources in cases of uncertainty, using various indicators such as patient health status, wireless and sensor data, service priorities, and the status of linked networks and devices. All these prove that the allocation of medical resources is an appropriate option for a data-based decision-making task. Some studies have demonstrated the growing role of intelligent resource allocation in wireless and IoT-enabled networks. Specifically, a deep reinforcement learning approach has been used in Massive Multiple-Input Multiple-Output (MIMO)-Non-Orthogonal Multiple Access (NOMA) systems to minimize computing complexity of dynamic resource allocation and preserve high throughput in a variety of channel conditions [6]. Furthermore, energy efficiency and service performance have significantly improved with Quality-of-Service (QoS)-aware load balancing algorithms for 5G-enabled IoT sensor networks, demonstrating the need of achieving a balance between usage efficiency and QoS requirements in connected environments [7].

In the healthcare sector, the IoT systems are still facing a range of restrictions and limitations. One of these constraints is the reliance of many of these systems on seemingly fixed structures and low control speed with centralized decision-making processes, making these systems unsuitable for dealing with rapidly changing clinical needs and extreme emergencies [5]. Usually, the allocation of medical resources is managed in real time, such as the reservation of intensive care beds and the preference for access to cameras or wireless communications, through manual procedures or specific instructions. These methods cannot or can adapt enough, and it is difficult to expand them, especially in periods of crowded networks and the increase in incoming requests. Moreover, resource allocation and resource management strategies usually lack reliability of basic networks, which may vary significantly in hospital crowded environments [8]. Healthcare providers may experience ineffective or late reactions if there are no flexible and effective predictive models for resource allocation and management, which may have negative impacts on patient care and health. Machine learning provides outstanding tools for understanding customization patterns based on previous data and current context [9]. Machine learning models can detect non-linear relationships between the variables that control and affect the efficiency of customization, reducing the need to develop complex and ineffective policies manually. The results of the models can also be evaluated using standard forecasting metrics, with operational factors such as the case of use and association with network variables and network speed [10].

This study focuses on a machine learning methodology based on a data set related to the allocation of medical resources of IoT sensors, in order to anticipate the efficiency of resource exploitation. This research follows a distinct approach, including: removing unhelpful features of learning, analysing statistical characteristics and relationships. Explanation and clarification of predictions through the study of relationships and the use of analytical graphs. The results indicate that data-driven models, especially regression-based learning, can help improve the consistency and efficiency of the allocation decision-making process in future healthcare settings that utilize IoT technology. The main contributions of this research study are summarized in the following points.

- Proposing a multi-model regression approach for predicting medical resource allocation efficiency of IoT-enabled 6G networks that consists of data pre-processing, data analysis, model learning, and operational factors interpretation.
- Conducting correlation analysis through visualizing a heat map plot to understand feature relationships and identify features that are strongly related to the target variable.
- Benchmarking multiple regression models, including Linear Regression (LR), Ridge (R), Random Forest (RF), and Gradient Boosting (GB) on a public dataset and under a consistent 80:20 train-test split and 5-fold cross-validation procedure.
- Evaluating adopted models with the performance evaluation metrics, such as root mean square error (RMSE), determination coefficient (R-squared), and mean absolute error (MAE), and selecting the best regression model based on the minimum value of RMSE metric that prioritizes the lower large error risk.
- Validating learning generalization on a holdout test set and supporting the experimental results with some diagnostic plots of residual distribution for predicted vs. actual resource utilization efficiency.
- Adding an interpretation steps of model outputs for analytical intuitions of networks' operational factors through relating predicted resource utilization efficiency to network speed and utilization status, enabling actionable plan for designing intelligent allocation policies.

The rest of this paper is structured as follows: Section 2 reviews the methods and frameworks of related studies. Section 3 outlines the proposed approach in detail. Section 4 includes the experimental analysis, findings, and discussion. Section 5 offers the conclusions and summarizes the possible future research directions for the proposed work.

2. RELATED WORKS

The related works section deals with reviewing the frameworks and approaches of medical development in the field of IoT and 6G network, with a focus on the challenges and difficulties associated with these ongoing advancements. The significant shift in 6G networks and IoT toward connected healthcare systems is substantially dependent on medical equipment, remote monitoring and wireless sensors. This shift increases the difference and diversity of data and makes traditional medical resource allocation methods and strategies more difficult and complex to implement than smart adaptive decisions [11]. The spread of IoT usage supported by the 6G network is one of the most important expected objectives, which are achieved by the key advantages, including improving the quality of network connection and providing high data transmission speeds[12]. At the global level, a substantial standardization is needed to ensure the effectiveness of these advantages[13]. However, there are still some challenges with the design of standard protocols that achieve reliable performance of IoT-based medical services. Most previous studies have focused on specific and theoretical background of 6G network, IoT, and communication between Vehicle-to-Everything (V2X) that are widespread and adaptable[14]. The modern 6G network's methodologies and frameworks have discussed the main points of its trends and challenges in managing and utilizing resources[9]. Integrating 6G technologies with the IoT attained an efficient use of resources by predicting the expected faults in the equipment.

The 6G networks and IoT sensors have enabled real-time patient monitoring and provision of appropriate medical services to reduce unused resources and hospital services [15]. Moreover, the

logistics sector has improved the inventory tracking and management process, which has reduced the need for excess inventory and increased the efficiency of the storage space. The ability to support a large number of devices is one of the main benefits of integrating 6G networks and the IoT devices, which contributed to improving resource management in various sectors [9]. In addition, data collection and processing in low record time is one of the advantages offered by the 6G predictive service, which enhances effective resource management [9]. Alhashimi et al. [8] provided an in-depth study focused on heterogeneous networks and resource management methods in 6G network and mobile communications systems. This study reflected current knowledge and identified promising areas for future research. One of the most prominent features of this work is the comprehensive review of spectrometry and interference management techniques, which are vital to improve service quality. Shen et al. [16] also presented an innovative approach to managing wireless resources within the high-intensity IT services within 6G networks. The authors have created a comprehensive simulation platform that includes various wireless resource management techniques, allowing for effective simulation of a variety of 6G networks and IoT services.

Other studies related to 6G network structures and requirements have shown that future scenarios, including the health sector, underscore the necessity of dynamic resource management, given the diverse needs of services between a range of heterogeneous devices, such as sensors, peripherals and platforms [17]. Cloud or terminals. There is a real challenge in the IoT data processing and decision-making that occurred close to the source, reducing the delay time and easing the burden on the basic network, which reflects an increased interest in peripheral medical devices in hospitals and health centres [18]. The effectiveness of resource management is vital in distributed systems, and this topic has been a recurring focus in academic research. In the field of 6th generation networks, the effectiveness of resource management includes the discovery and identification of all available resources, as well as the selection, arrangement and distribution of appropriate resources to enhance the performance of the benefit [19]. This improvement may include a wide range of variables, including overall performance, cost efficiency, energy efficiency, data accuracy, coverage, reliability, and more. Although important studies are being conducted in various fields of computing, resource management in 6G environments is still a major challenge that requires innovative solutions. As the fifth and sixth generation networks continue to develop, the issue of resource management has become among the most researched topics, especially with the increasing reliance on learning algorithms in dynamic resource coordination [20].

Some of the works have explicitly focused on that machine learning solutions are suitable for addressing the complexity of allocation in IoT systems within 5G and 6G environments, both at the level of technology [21]. Innovative machine learning and statistical models can develop predictive tools that enable highly accurate forecasting of network load. Learning progress and its adaptation on the network design and the ability of the management system depends on adjusting the allocation of resources to respond to the expected changes [22]. This may be complicated in cases where network conditions change quickly or unexpectedly. In such situations, the assigned resources may not match the actual requirements. Sheng et al. [23] conducted a study to develop a double approach to improve the wireless coverage of the new 6G Satellite-Terrestrial Integrated Networks (STINs) through the use of large satellite groups. This approach focuses on network design analysis to meet the needs of the 6G service, which helps in planning STINs networks and improving resource scheduling to achieve better coverage performance. The sophisticated and innovative aspects of 6G technology can be challenging during implementation, and it can be difficult to manage smart resource planning in an efficient manner.

Few studies have explored learning-based resource allocation methods that focus on service quality in next-generation networks. An approach using deep reinforcement learning is proposed

for dynamic resource allocation in Massive MIMO–NOMA systems [6]. This approach achieved a reduction in computational complexity while maintaining good performance even with channel changes over time. Dey et al. [7] presented two simple efficient distributed methods in IoT-based wireless sensor networks, including a Load Balanced Greedy Cluster Assignment (LBGCA) for large-scale IoT networks and a Multi-modal Load Balanced Greedy Cluster Assignment (MLBGCA) for QoS-aware applications. Simulation findings on several deployment patterns demonstrated major improvements in load balance and significant reductions in energy consumption as compared with previous techniques. In addition, Gad-Elrab et al. [24] proposed an adaptive fog–cloud resource allocation technique based on multi-criteria decision methods for improving response time and resource utilization efficiency in latency-sensitive applications. Almadhor et al. [25] provided an artificial intelligence-based 6G-IoT that contains an adaptable medical resource allocation. They used the eXtreme Gradient Boosting (XGBoost) method to predict the utilization efficiency of resources, and achieved 0.988 of R^2 and 0.0323 of RMSE. However, there is still a need to improve the achieved values of these assessment metrics. Recent research suggests that managing resources in healthcare sector, which relies on advanced IoT and networking technologies, has become more important. There is also an increase in machine learning reliance and data-based improvement methods. Previous studies focused on theoretical models or solutions tailored to specific systems. They lack a comprehensive assessment of resource allocation efficiency prediction using the IoT-driven 6G networks data availability and unified regression framework. This study contributes to the existing literature by adopting a data-based approach and focusing on both the accuracy of prediction and practical understanding of efficient allocation to medical resources.

3. MATERIALS AND METHODS

This section describes the data and proposed approach in terms of materials and methods of the study. The study aims to analyze and predict the efficiency of allocating medical resources in a healthcare setting supported by IoT technology and 6G networks. It begins with describing the dataset, covering its attributes and target variable, and ends with explaining the approach's phases.

a. Dataset Description

To implement and assess the developed prediction approach of resource allocation efficiency for 6G networks-enabled IoT medical sensors, a public dataset, namely "IoT-Driven MR Allocation for 6G Network", available on KAGGLE platform [26]. Table 1 describes the attributes of the dataset.

Table 1. Description of dataset attributes

#	Feature (Column)	Type	Description
1	Patient_ID	object	Unique patient identifier used for record tracking. Typically removed before modeling because it does not represent clinical/operational behavior and can cause ID-driven learning.
2	Age	int64	Patient age (numeric). Used to capture demographic differences that may affect health condition, care intensity, and expected resource demand.
3	Gender	object	Patient gender category. Can correlate with disease prevalence/risk and care pathways; encoded as categorical during preprocessing.
4	Health_Status	object	Categorical indicator of the patient's general condition (e.g., stable/critical). Useful for prioritization and

			expected resource intensity.
5	Disease_Type	object	Disease/diagnosis category. Represents clinical context that influences the type and amount of resources required.
6	Resource_Needed	object	Category describing the type of medical resource required (e.g., bed, equipment, specialist service). Drives allocation decisions and expected utilization patterns.
7	Duration_of_Stay	int64	Length of stay (numeric). Reflects time-based resource consumption and affects utilization efficiency and planning.
8	Sensor_ID	object	Unique IoT sensor ID associated with the record. Often removed for modeling for the same reasons as Patient_ID (identifier, not a causal predictor).
9	Resource_ID	object	Unique identifier of the allocated resource instance. Typically excluded to improve generalization.
10	Utilization_Status	int64	Discrete operational status (numeric-coded) indicating utilization state (e.g., low/medium/high or available/busy depending on dataset definition). Used to capture current operational load.
11	Network_Speed	float64	Continuous network performance indicator (e.g., throughput/speed). Represents connectivity quality that can impact timely allocation and efficiency.
12	Overlapping_Interference	int64	Numeric indicator of interference level. Higher interference can degrade communication performance and indirectly affect allocation efficiency.
13	Resource_Utilization_Efficiency	float64	Continuous outcome representing utilization/allocation efficiency. This is the prediction target in the regression task.

This dataset consists of two main CSV files: the medical resource allocation binary dataset and medical resource allocation dataset. The first file contains a structured binary classification dataset to model utilization allocation efficiency as a binary class target for a supervised classification (0: not allocate and 1: allocate). The second file has full medical resource allocation dataset to model utilization allocation efficiency as a continuous target for a supervised regression. This study uses the full medical resource allocation dataset. The raw file contains several attributes describing the allocation context, including patient and device related identifiers, network/service indicators, and operational status variables. The regression target is resource utilization efficiency, which represents the efficiency of utilization allocation as a real-valued outcome to be predicted from the remaining features. The feature set is a mixture of network-related indicators, such as network speed that reflect communication conditions affecting allocation efficiency, Contextual or clinical status variables, such as health status that represents patient condition categories influencing prioritization and resource demand, and operational utilization signals such as actual utilization status that represents current or observed utilization state, which is also examined against predicted efficiency in the results analysis, and additional numeric or categorical attributes provided in the dataset capturing the broader allocation and environment context.

b. Proposed Approach

The developed approach includes four successive phases aim at predicting the effectiveness of resource exploitation, based on medical information derived from the IoT data. Figure 1 depicts the procedural structure of approach's phases. The first phase begins with the reception of the data set, removing the unique identifiers, addressing the missing values, and converting the values of

categorical variables into a numerical values, in addition to standardizing the values of all dataset attributes. The second phase is dedicated to conducting an analytical exploratory of data that includes the perception of how traits are spread, the presentation of tail statistics, and the calculation of correlations to understand the operational behaviors and underlying patterns. The third phase trains and evaluates multiple regression models through a 5-cross-validation and train-test split strategies to select the best predictor of resource allocation utilization efficiency. The last phase is a post-model interpretation that inspects the relation between the key operational factors and the predictions of selected model.

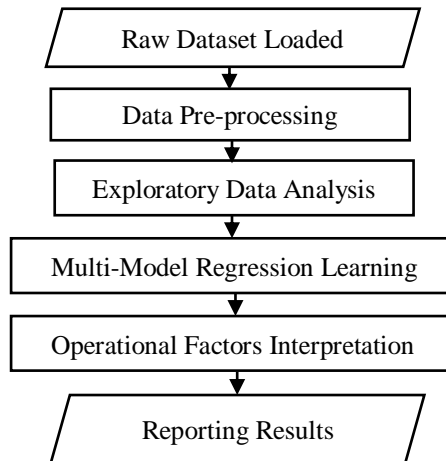


Figure 1. Flowchart of proposed approach

c. Data Pre-processing

This phase receives and processes the raw dataset, preparing it for effective learning and unbiased modeling. All identifier variables, such as Patient ID, Sensor ID, and Resource ID, have been removed because they do not represent significant clinical or operational patterns. Their inclusion could lead to memorization based on IDs rather than promoting generalizable learning. The remaining features are categorized into numerical and categorical groups to allow for tailored processing. Algorithm 1 gives the steps of data pre-processing phase.

Algorithm 1: Data Pre-processing
Input: Raw dataset X
Output: Transformed dataset X'
Begin
1. Removing identifier columns (Patient_ID, Sensor_ID, and Resource_ID).
2. Identifying numerical and categorical features.
3. Imputing numerical features using median.
4. Standardizing numerical features using Z-score standardization.
5. Imputing categorical features using the most frequent value.
6. Applying one-hot encoding to categorical features.
7. Returning the preprocessed dataset X' .
End

Numerical features are imputed with median values and standardized to maintain consistent magnitudes across all attributes using Z-score standardization, calculated as:

$$x' = \frac{x - \mu}{\sigma} \quad (1)$$

Where x is the raw value, μ is the feature mean, and σ is feature standard deviation.

Similarly, we impute the categorical features using the most frequent value and transform them by one-hot encoding method. At the category level, this processing of categorical features allows the models to successfully interpret the differences between them. The pipeline of this phase ensures that all subsequent models receive inputs that are model-suitable, clean, and reliable.

3.2.2. Exploratory Data Analysis (Eda)

The EDA phase provides an initial understanding of the operational and statistical features of the used dataset. The significant variables, such as utilization status, network speed, and resource utilization efficiency, are analyzed using descriptive statistics and percentile function formulated in Equation 2. This analysis incorporates means and tail percentiles (p95 and p99) to ensure that both extreme operational and typical conditions behavior are covered.

$$\text{Percentile}_p(X) = \inf\{x: F_X(x) \geq p\} \quad (2)$$

The equation of percentile function is determined by the value below p percent for all observations fall. A correlation heat map is created to detect linear dependencies and emphasize strong relationships between the predictors and the target variable. Further visual analyses, including scatter plots and group comparisons based on health status or disease type, provide valuable insights into how clinical and network characteristics may affect resource utilization. This phase clarifies the data structure and sets clear expectations for how the model will learn. Algorithm 2 provides the steps of exploratory data analysis phase.

Algorithm 2: Exploratory Data Analysis

Input: Feature matrix X' , target y

Output: statistical_summary, correlation_matrix

Begin

1. Computing descriptive statistics (mean, std, min, max) for all features
2. Computing tail percentiles: p95, p99 for key operational variables
3. Generating correlation matrix for numerical features
4. Visualizing correlations using heat map
5. Plotting distributions of major variables (histograms, boxplots, scatterplots)
6. returning statistical_summary, correlation_matrix

End

d. Multi-model Regression Learning

The multi-model regression learning is a key part of the proposed approach. It builds four effective regression models described as follows:

Linear Regression (LR): It describes the connection between input features and the target variable by using a weighted linear combination of the predictors. It presumes a linear relationship and reduces the total of squared differences between actual and predicted values. This method acts as a clear baseline, offering insight into how effectively allocation efficiency can be understood through basic linear relationships.

Ridge Regression (RR): It enhances linear regression by incorporating an L2 regularization term, which discourages large coefficients. This regularization technique decreases model variance and alleviates multicollinearity among features, resulting in more stable predictions. The RR is

especially beneficial when dealing with datasets that include correlated predictors, as it emphasizes generalization over precise fitting.

Random Forest Regressor (RFR): It is an ML technique that creates several decision trees by using bootstrap samples and selecting features randomly. The final predictions are generated by averaging the results from individual trees, which helps minimize overfitting and effectively represents complex, non-linear relationships. This model is appropriate for diverse Internet of Things (IoT) and network-related data, where it is challenging to clearly define the interactions between variables.

Gradient Boosting Regressor (GBR): It creates a series of decision trees one after another, with each new tree designed to address the errors made by the previous group of trees. The model attains high predictive accuracy and a good balance between bias and variance by repeatedly minimizing a loss function. This approach is useful for recognizing subtle patterns and relationships in organized healthcare and network data.

At this phase, the built regression models are trained to predict resource utilization efficiency using a consistent and reproducible protocol. The dataset is split into an 80% training set and a 20% test set. Model evaluation is conducted using 5-fold cross-validation on the training portion. This method offers an in-depth assessment of the changes, rather than a single split. After the training stage, the models are evaluated under the same conditions for comparing their results using evaluation metrics, including the mean absolute error (MAE), mean square error (RMSE), and coefficient determination (R^2), computed using the following equations:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2} \quad (5)$$

Where y_i and \hat{y}_i are the actual, predicted and the mean values of the target variable, and \bar{y}_i is the mean value of the actual target variable.

These metrics measure average error, sensitivity to large deviations, and the proportion of variance explained. The model with the lowest RMSE is chosen as the final predictor and retrained on full training data. The selected best trained model is evaluated on testing set using the same MAE, RMSE, and R^2 metrics. The values of MAE, RMSE, and R^2 for all models on training data of cross-validation and testing data of best model is reported at this phase. This structured comparison framework guarantees that the selected model is both accurate and stable in its predictions of utilization efficiency. Algorithm 3 presents the steps of multi-model regression learning phase.

Algorithm 3: Multi-Model Regression Learning

Input: Feature matrix X' , target y
 Output: Best regression model m^* and values of MAE, RMSE, and R^2
 Begin
 1. Splitting dataset into training set (80%) and testing set (20%).
 2. Applying a 5-fold cross-validation on training data.
 2.1. Training regression models (LR, RR, RFR, and GBR).
 2.2. Evaluating models using MAE, RMSE, and R^2 .
 2.3. Selecting the best regression model m^* with minimum RMSE model.
 3. Retraining the selected best model m^* on full training data.
 4. Evaluating the selected best model m^* on testing set using MAE, RMSE, and R^2 .
 5. Returning the values of MAE, RMSE, and R^2 for all models on training data of cross-validation and testing data of best model.
 End

e. Interpretation of Operational Factors

Besides to reaching a high level of expectation for selected model performance, this interpretive phase compares the results of the model to the realistic operational conditions. It examines the association between the predicted efficiency estimates and the elements influencing the network structure, such as network speed or overlapping interference, to determine whether the model reflects the significant behavior of the system or not. The predictions of utilization efficiency are examined in relation to specific clinical health status factors to identify possible variations in resource needs among different patient groups. This informative phase converts the basic prediction into practical intuition and aids in resource allocation in healthcare data-driven decision-making that utilizes IoT technology. Algorithm 4 lists the steps of operational factors interpretation phases.

Algorithm 4: Operational Factors Interpretation
 Input: Model predictions \hat{y} , operational features
 Output: Interpretable operational insights
 Begin
 1. Plotting actual against predicted values.
 2. Analyzing relation between network speed and predictions.
 3. Comparing prediction distribution across health status.
 4. Plotting the residual errors actual and predicted values for bias.
 5. Summarizing the insights of resource allocation decisions.
 End

4. EXPERIMENTAL RESULTS AND DISCUSSION

This part reviews the practical tests conducted on the proposed method of distributing medical resources. After completing the sequence of experiments, the results obtained are detailed and analyzed based on the specified standard performance indicators. Comparisons of the performance results for all comparable models are also presented. In addition, this section describes the diagnostic statistics, which include graphs, tables and explanations to show the quality of prediction outcomes, the analysis of the residual errors, and the investigation of the relations between model predictions and the operational factors of the network. These experiments are conducted using the Google Cloud computing environment, with an NVIDIA Tesla T4 graphics accelerator with 16 GB of RAM. Moreover, the operating settings included a CPU equipped with two virtual processing units and 12 GB RAM to support training and

evaluation processes. This setup offered necessary computational power to train the models efficiently within a reasonable time cost. The experiments start by reading the full dataset and apply the data pre-processing steps that include removing identifier fields (Patient ID, Sensor ID, and Resource ID) before analysis. This is important because they act as an identifier for entities rather than true analysts. After that, scaling and one-hot encoding are performed on the desired columns. Next, the EDA stage is conducted on the raw selected columns. Figure 2 shows the efficiency scores of resource utilization target predictor across the 500 records and Figure 3 shows the histogram distribution of resource utilization efficiency values.

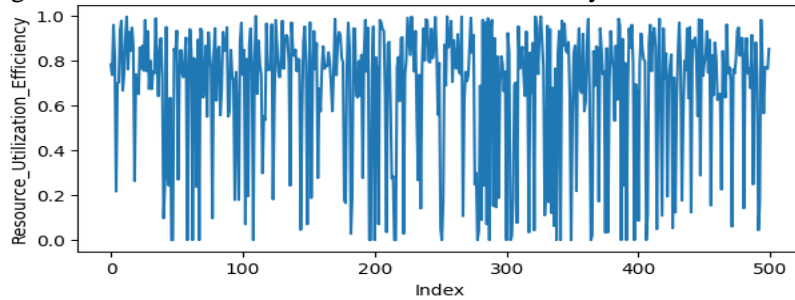


Figure 2. Efficiency scores of resource utilization target predictor.

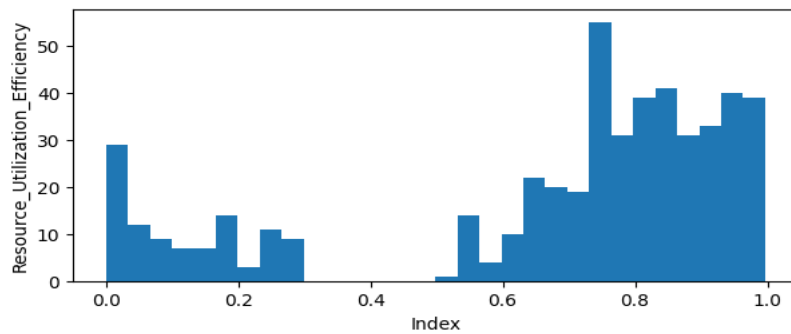


Figure 3. Histogram distribution of resource utilization efficiency values

As shown in Figure 2, the curve fluctuates strongly, with many observations in the high-efficiency region near 0.7–1.0, but also repeated sharp drops close to 0. The frequent peaks near the upper bound (close to 1) suggest that, under many conditions, the allocation process is achieving high utilization efficiency, which means good matching between needed and allocated resources. Also, in Figure 3, the distribution is not uniform. Most values are concentrated in the higher-efficiency range, which matches the high p95 and p99 values. Figure 4 visualizes the cumulative distribution function (CDF) values of resource utilization efficiency.

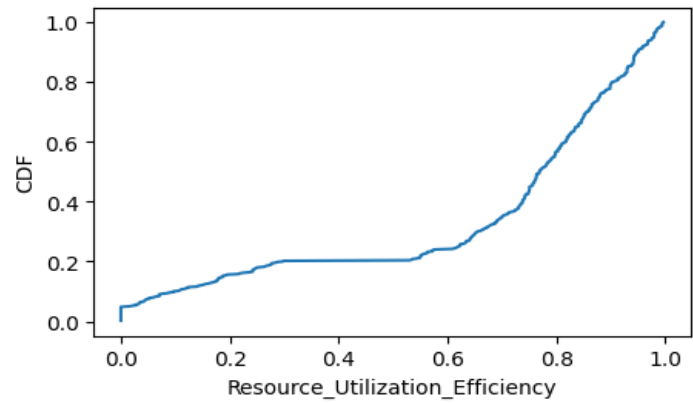


Figure 4. Cumulative distribution function (CDF) values of resource utilization efficiency

The CDF indicates that most of the utilization efficiency is high, with the upper end nearing perfect efficiency (approximately 1.0). However, a smaller group of records shows low efficiency, which lowers the overall average. Figure 5 presents the network speed variation with values spanning a wide range roughly from low teens up to near 100. Figure 6 shows the histogram distribution of network speed values. Also, Figure 7 gives the CDF values of network speed.

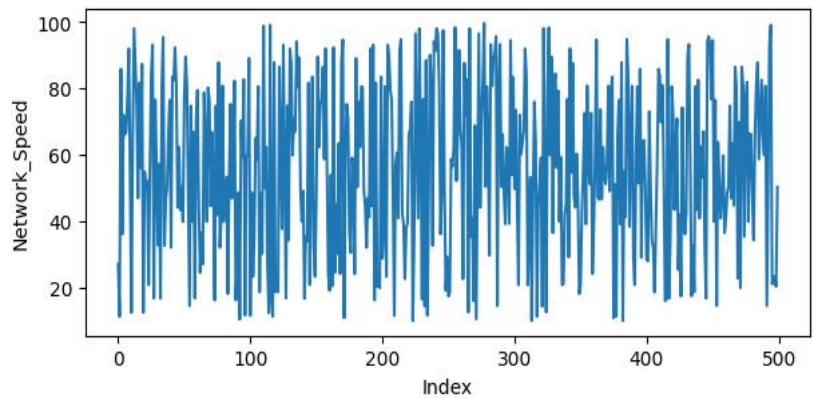


Figure 5. Network speed variation with values spanning from low teens up to near 100

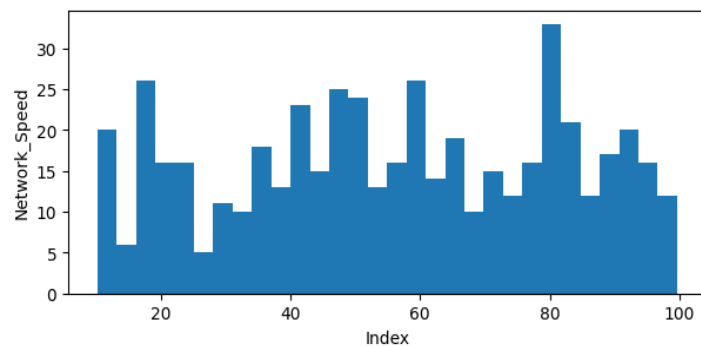


Figure 6. Network speed variation with values spanning from low teens up to near 100

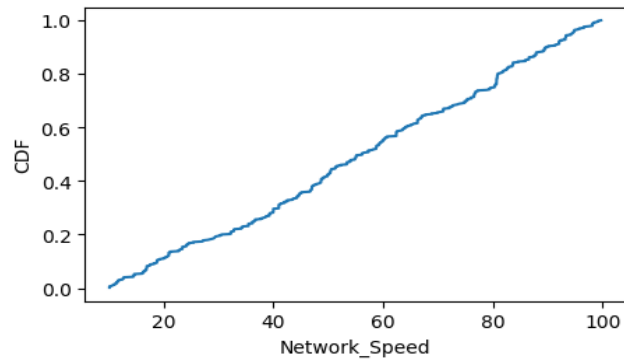


Figure 7. Network speed variation with values spanning from low teens up to near 100

From Figure 5, we can see that the significant differences indicate that network conditions are diverse, with some records reflecting slow or problematic network states and others indicating fast or optimal conditions. This variation is important because network speed usually serves as an indicator of capacity and quality of service. When the speed is low, the system may have difficulty coordinating and using resources effectively. Conversely, when the speed is high, coordination tends to improve, leading to increased efficiency. The variation of network speed, shown in Figure 6, is helpful for learning, as it offers different operating conditions. If it is noticed later a positive relationship between network speed and predicted utilization efficiency, this distribution supports this finding by showing that speed varies significantly. In Figure 7, the CDF of network speed steadily increases across the range, showing how percentiles move from low to high speeds. To analyse the correlation between features and the resource utilization efficiency target variable, Figure 8 illustrates the pairwise Pearson correlation between numerical variables.

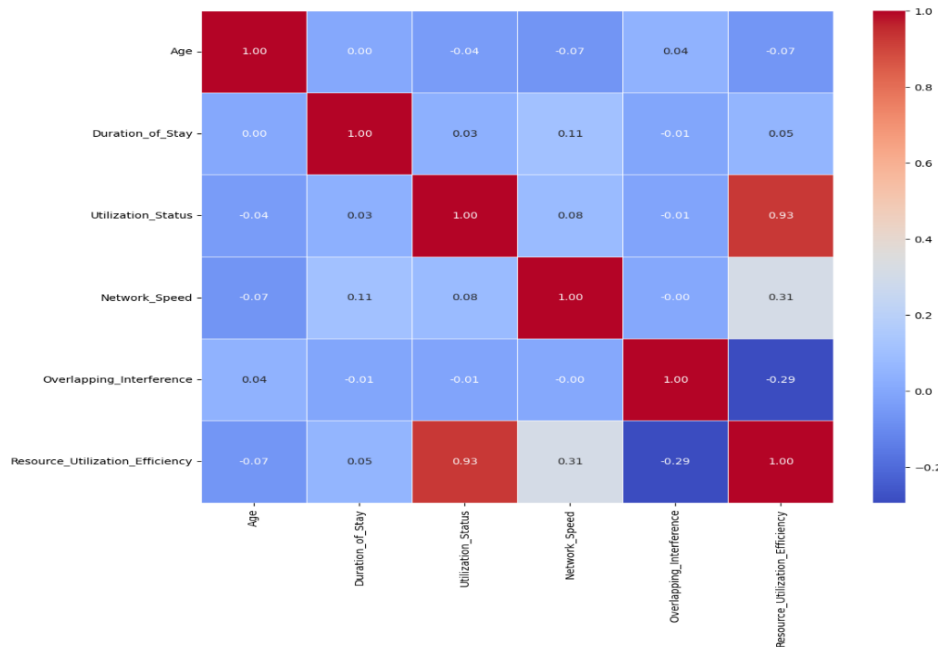


Figure 8. A correlation heat map plot of numerical features

The analysis of pairwise Pearson correlation between numerical variables given in Figure 8 has two main objectives: first, to identify factors that are closely associated with the target and could

enhance model performance, and second, to identify multicollinearity among the input variables. Input pairs that are strongly correlated (for example, with an absolute correlation greater than 0.85) might provide overlapping information. Although tree-based models typically handle this redundancy well, linear models may experience issues due to unstable coefficients. The heat map aids in understanding the data and helps in choosing models.

To generate the results of multi-model regression learning stage, the dataset is divided into training and testing sections with an 80/20 ratio. Model selection is performed using 5-fold cross-validation on the training dataset. The MAE, RMSE, and R^2 evaluation metrics are utilized to assess the four models' prediction of resource utilization efficiency based on regression results. The regression models are compared for dealing with linearity distribution of the features, in which the LR and RR are linear baselines, and RFR and GBR are non-linear ensemble methods. The model selection idea allows to choose an interpretable linear reference as well as addressing complex data feature interactions occurred in IoT-network-driven data. Table 2 presents the results of the 5-fold cross-validation conducted on the training set. The optimal model is determined by the lowest RMSE, as RMSE highlights significant differences that are especially unfavourable for decisions regarding resource allocation.

Table 2. Cross-validation performance of the compared regression models

Model	MAE	RMSE	R2
LR	0.012428	0.019995	0.995146
RR	0.012775	0.020071	0.995138
RFR	0.002418	0.004742	0.999714
GBR	0.003531	0.005685	0.999607

As given in Table 2, the LR and RR linear models demonstrate good performance ($R^2 \approx 0.995$), showing that a significant amount of the variability in the target can be accounted for by roughly linear trends. The ensemble RFR and GBR models show significant improvement. The RFR achieves the lowest RMSE of 0.004742 and the highest R^2 of 0.999714, indicating that non-linear interactions and feature thresholds are important factors in predicting resource utilization efficiency. The GBR is competitive but somewhat less accurate than RFR in this evaluation. Figure 9 displays the predicted values compared to the actual values for each model, along with a diagonal line representing perfect prediction.

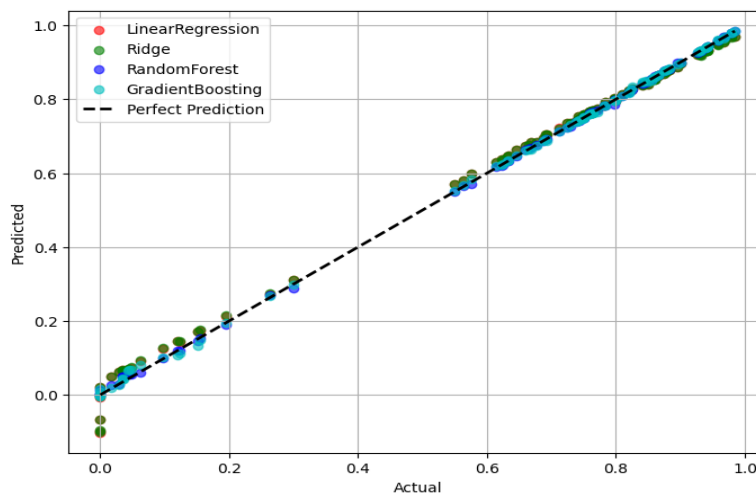


Figure 9. Predicted and actual resource utilization efficiency for all compared models

From Figure 9, we can see that the models that generate points closely grouped along the diagonal show greater accuracy and reduced systematic bias. The ensemble models in this study show a much closer alignment compared to the linear baselines, which corresponds with the better MAE/RMSE and R^2 values presented in Table 2. The Figure 9 also aids in detecting heteroscedasticity, which occurs when the magnitude of errors varies with the level of the target, indicated by an increasing spread of points in specific ranges. The RFR model is chosen as the final model based on cross-validation RMSE results. It is retrained using the entire training dataset and then assessed using the separate test set. The results on the test set (MAE ≈ 0.002302 , RMSE ≈ 0.004239 , and $R^2 \approx 0.999821$) indicate that the model generalizes well and has a low prediction error. This relation means that the model can accurately follow the expected efficiency signal for new samples, which lowers the likelihood of significant allocation mistakes. Figure 10 exhibits the distribution of residuals, which is the difference between actual values and predicted values.

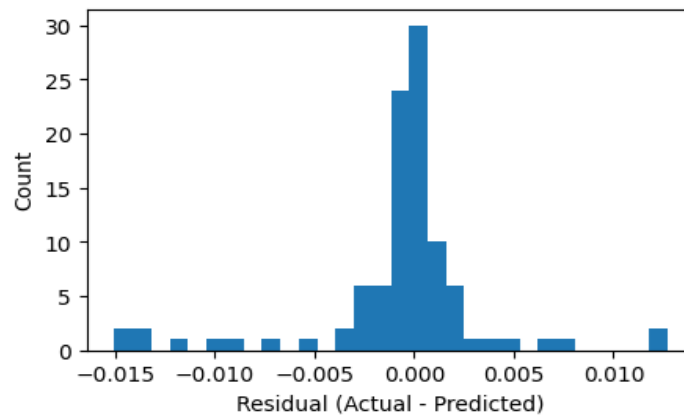


Figure 10. Residual distribution of test set for the selected RFR model

As shown in Figure 10, a residual histogram that is centred close to zero and has a narrow range suggests that the model is accurate and has little bias. Skewness indicates a consistent tendency to overestimate or underestimate in certain areas, while heavy tails suggest that there can be significant errors from time to time, which may result in ineffective allocation decisions during unusual situations. Based on the assessment results, we anticipate that the residuals will be close to zero, corresponding with the low RMSE observed in the test set. For the results of operational factors interpretation stage, Figure 12 explores the relationship between predicted resource utilization efficiency and network speed.

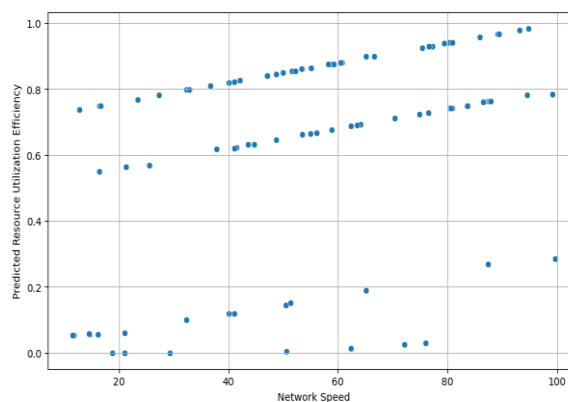


Figure 11. Scatter plot of predicted resource utilization efficiency versus network speed with Pearson correlation

As exposed in Figure 11, the positive relationship suggests that increased network speed is generally linked to higher predicted efficiency. This relationship aligns with the understanding that better network capacity can lead to more effective use of resources. The Pearson correlation reported ($r = 0.4704$) indicates a moderate relationship, implying that network speed plays a significant role in efficiency, but it is not the only factor; other contextual elements also influence the final outcome. Figure 12 shows the differences in predicted efficiency among various health-status groups.

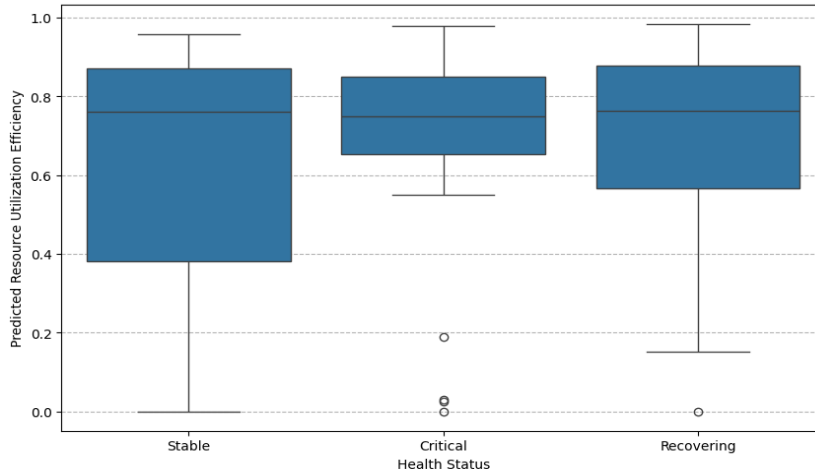


Figure 12. Boxplot of predicted resource utilization efficiency distribution by health status

In Figure 12, the variations in the median and interquartile range (IQR) among groups suggest that the context of patients may affect the patterns of allocation efficiency. This analysis helps determine if the model identifies significant differences between groups and can also reveal fairness issues if some groups consistently have lower predicted efficiency. These results encourage additional analysis that includes specific constraints related to the field, such as priority rules and clinical risk. Figure 13 examines the connection between predicted efficiency and the actual utilization status.

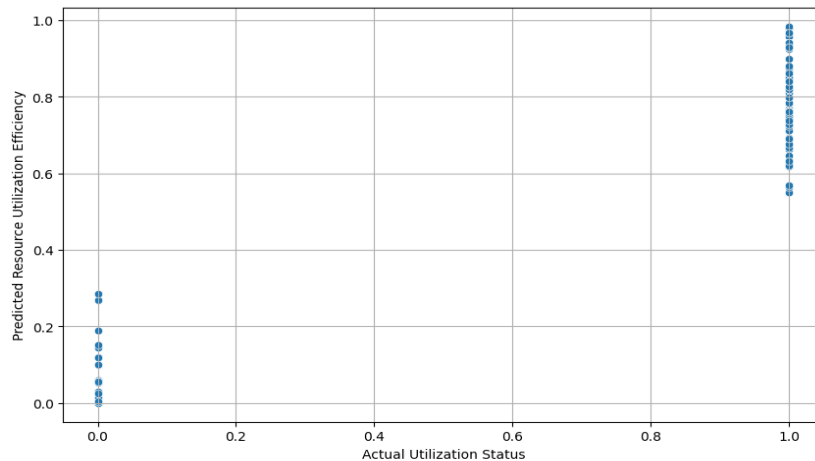


Figure 13. Scatter plot of predicted resource utilization efficiency versus actual utilization status

From Figure 13, the distinct trend, whether upward or downward, would indicate alignment between the forecasted continuous efficiency signal and the related utilization measure. Figure 13

acts as a validity test. If the predicted efficiency does not show significant changes based on utilization status, it may indicate that the model is identifying false correlations instead of relevant operational patterns. To compare our proposed approach with recent related work, Table 3

Table 3. Comparison results of proposed model against the model or recent related work.

Work	Model	RMSE	R2
Almadhor et al. [25]	XGBoost	0.0323	0.9880
This work	RFR	0.0047	0.9997

The experimental and comparison results show that ensemble regression methods, especially proposed RFR model, deliver very accurate predictions of resource utilization efficiency for IoT-6G network-driven data features. The diagnostic plots confirm this conclusion by indicating a close alignment between predicted and actual values, with the residual distribution primarily centred on zero. Moreover, the subsequent analyses show that network speed and health status context are linked to expected efficiency, providing clear information for creating and reviewing medical resource allocation policies.

5. CONCLUSIONS AND FUTURE WORK

This research work proposed an approach that uses the 6G network-based data to understand and predict how efficient medical resources are allocated in the IoT-powered healthcare environment. This research shows how machine learning techniques can determine the complex relationships between clinical and operational factors and communication networks through the use of a comprehensive approach that includes data pre-processing, exploratory analysis, multiple regression modeling, and interpretation of results. Experimental results indicate that ensemble regression methods, particularly those based on tree structures, demonstrate superior performance better than linear models in terms of MAE, RMSE and R². These models are more effective at capturing nonlinear relationships during resource allocation processes. Beyond experimental results and predictive accuracy, this study further emphasizes the importance of applied benefits and findings. Analysis of distribution, percentage, and cumulative data behavior exposes widespread inefficiency scenarios despite high utilization rates. This finding is critical for healthcare systems operating under stringent service conditions. The significant correlation between predicted efficiency and network speed further underscores the pivotal role of critical communications in effectively enhancing resource allocation. The results confirm the validity of the use of learning-based methods in modeling medical resource consumption and set a standard for future studies that benefit from public data sets. This research focuses on using regression-based methods to estimate the efficiency of utilization for the resource allocation.

In future work, there are some suggestions to enhance the proposed work. Upcoming studies can include samples for identifying different customization options to enhance prediction results by addressing the personalization issues or priority levels directly. Also, expanding the approach to include multi-objective improvements, such as balancing efficiency, delay, fairness and reliability, will provide a more accurate representation of the challenges facing healthcare environments in reality. Ultimately, studies using larger datasets or those that rely on real-time data can highlight the suitability of this approach for use in smart health resource management systems.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

ACKNOWLEDGEMENTS

“The author is grateful to the Deanship of Scientific Research at King Saud University's College of Computer and Information Sciences (CCIS) for funding this research.”

REFERENCES

- [1] V. A. Dang, Q. Vu Khanh, V.-H. Nguyen, T. Nguyen, and D. C. Nguyen, "Intelligent healthcare: Integration of emerging technologies and Internet of Things for humanity," *Sensors*, vol. 23, no. 9, p. 4200, 2023.
- [2] M. Papaioannou, M. Karageorgou, G. Mantas, V. Sucasas, I. Essop, J. Rodriguez, and D. Lymberopoulos, "A survey on security threats and countermeasures in internet of medical things (IoMT)," *Transactions on Emerging Telecommunications Technologies*, vol. 33, no. 6, p. e4049, 2022.
- [3] L. Liu, D. Essam, and T. Lynar, "Complexity measures for IoT network traffic," *IEEE Internet of Things Journal*, vol. 9, no. 24, pp. 25715-25735, 2022.
- [4] H. Mazumdar, K. R. Khondakar, S. Das, and A. Kaushik, "Aspects of 6th generation sensing technology: from sensing to sense," *Frontiers in Nanotechnology*, vol. 6, p. 1434014, 2024.
- [5] A. Kumar, M. Masud, M. H. Alsharif, N. Gaur, and A. Nanthaamornphong, "Integrating 6G technology in smart hospitals: challenges and opportunities for enhanced healthcare services," *Frontiers in Medicine*, vol. 12, p. 1534551, 2025.
- [6] P. H. An, N. Dung, T. X. Uyen, N. T. C. Nghia, and N. M. Nghia, "Deep Reinforcement Learning-Based Resource Allocation in Massive MIMO NOMA Systems," *International Journal of Computer Network & Communications (IJCNC)*, vol. 17, no. 6, pp. 1-20, 2025.
- [7] B. Dey, S. Nandi, S. Bandyopadhyay, and S. Borgohain, "Multimodal QOS Aware Load Balanced Clustering in 5G-Enabled IOT Sensor Network," *International Journal of Computer Networks & Communications (IJCNC)*, vol. 17, no. 6, pp. 21-40, 2025.
- [8] H. F. Alhashimi, M. N. Hindia, K. Dimiyati, E. B. Hanafi, N. Safie, F. Qamar, K. Azrin, and Q. N. Nguyen, "A survey on resource management for 6G heterogeneous networks: current research, future trends, and challenges," *Electronics*, vol. 12, no. 3, p. 647, 2023.
- [9] S. S. Sefati, A. U. Haq, R. Craciunescu, S. Halunga, A. Mihovska, and O. Fratu, "A Comprehensive Survey on Resource management in 6G network based on internet of things," *IEEE Access*, 2024.
- [10] M. Abd Elaziz, M. A. Al-qaness, A. Dahou, S. H. Alsamhi, L. Abualigah, R. A. Ibrahim, and A. A. Ewees, "Evolution toward intelligent communications: Impact of deep learning applications on the future of 6G technology," *Wiley Interdisciplinary Reviews: Data Mining Knowledge Discovery*, vol. 14, no. 1, p. e1521, 2024.
- [11] O. S. Albahri, A. Zaidan, B. Zaidan, M. Hashim, A. S. Albahri, and M. Alsalem, "Real-time remote health-monitoring Systems in a Medical Centre: A review of the provision of healthcare services-based body sensor information, open challenges and methodological aspects," *Journal of medical systems*, vol. 42, no. 9, p. 164, 2018.
- [12] M. Mohammed, S. Desyansah, S. Al-Zubaidi, and E. Yusuf, "An internet of things-based smart homes and healthcare monitoring and management system," in *Journal of physics: conference series*, 2020, vol. 1450, no. 1, p. 012079: IOP Publishing.
- [13] L.-H. Shen, K.-T. Feng, and L. Hanzo, "Five facets of 6G: Research challenges and opportunities," *ACM Computing Surveys*, vol. 55, no. 11, pp. 1-39, 2023.
- [14] M. Noor-A-Rahim, Z. Liu, H. Lee, M. O. Khyam, J. He, D. Pesch, K. Moessner, W. Saad, and H. V. Poor, "6G for vehicle-to-everything (V2X) communications: Enabling technologies, challenges, and opportunities," *Proceedings of the IEEE*, vol. 110, no. 6, pp. 712-734, 2022.
- [15] S. Rajak, A. Summaq, M. P. Kumar, A. Ghosh, K. Elumalai, and S. Chinnadurai, "Revolutionizing healthcare with 6G: a deep dive into smart, connected systems," *IEEE Access*, 2024.

- [16] X. Shen, W. Liao, and Q. Yin, "A novel wireless resource management for the 6G-enabled high-density Internet of Things," *IEEE Wireless Communications*, vol. 29, no. 1, pp. 32-39, 2022.
- [17] H. Amirian, K. Dalvand, and A. Ghiasvand, "Seamless integration of Internet of Things, miniaturization, and environmental chemical surveillance," *Environmental Monitoring Assessment*, vol. 196, no. 6, p. 582, 2024.
- [18] J. R. Bhat and S. A. Alqahtani, "6G ecosystem: Current status and future perspective," *IEEE Access*, vol. 9, pp. 43134-43167, 2021.
- [19] R. Gupta, D. Reebadiya, and S. Tanwar, "6G-enabled edge intelligence for ultra-reliable low latency applications: Vision and mission," *Computer Standards Interfaces*, vol. 77, p. 103521, 2021.
- [20] A. Thantharate and C. Beard, "ADAPTIVE6G: Adaptive resource management for network slicing architectures in current 5G and future 6G systems," *Journal of Network Systems Management*, vol. 31, no. 1, p. 9, 2023.
- [21] H. M. F. Noman, E. Hanafi, K. A. Noordin, K. Dimiyati, M. N. Hindia, A. Abdrabou, and F. Qamar, "Machine learning empowered emerging wireless networks in 6G: Recent advancements, challenges and future trends," *IEEE Access*, vol. 11, pp. 83017-83051, 2023.
- [22] C.-C. Shen, C. Srisathapornphat, and C. Jaikaeo, "An adaptive management architecture for ad hoc networks," *IEEE Communications Magazine*, vol. 41, no. 2, pp. 108-115, 2003.
- [23] M. Sheng, D. Zhou, W. Bai, J. Liu, H. Li, Y. Shi, and J. Li, "Coverage enhancement for 6G satellite-terrestrial integrated networks: performance metrics, constellation configuration and resource allocation," *Science China Information Sciences*, vol. 66, no. 3, p. 130303, 2023.
- [24] A. A. A. Gad-Elrab, A. S. Alsharkawy, M. E. Embabi, A. Sobhi, and F. A. Emara, "Adaptive Multi-Criteria-Based Load Balancing Technique for Resource Allocation in Fog-Cloud Environments," *International Journal of Computer Networks & Communications (IJCNC)*, vol. 16, no. 1, pp. 105-124, 2024.
- [25] A. Almadhor, M. Ayari, A. Alqahtani, A. Al Hejaili, B. Bouallegue, R. A. Juanatas, and G. A. Sampedro, "An AI and 6G-IoT enabled computational framework for intelligent medical resource allocation and adaptive personalized healthcare," *Computing*, vol. 107, no. 12, p. 234, 2025.
- [26] Ziya, "IoT-Driven MR Allocation for 6G Network Dataset," <https://www.kaggle.com/datasets/ziya07/iot-driven-medical-resource-allocation-dataset>, [Last access: December, 2025].

AUTHOR

Dr. Hussain AlSalman is a staff member at the computer science department at King Saud University in Riyadh, Kingdom of Saudi Arabia. He got his Ph.D. in Artificial Intelligence from the United Kingdom. He received his B.Sc. and M.Sc. in computer science from KSU. He worked for several years as a consultant for several companies in the private sector and institutes in the government sector in KSA. From 2009 to 2014, he chaired Computer Science Department at the College of Computer and Information Sciences at KSU. Dr. Hussain AlSalman was a member of the review board of the Saudi Computer Journal from 2004 to 2014. His main research interests include medical image processing, machine learning algorithms, neural networks, computational methods for health care monitoring, ensemble and deep learning models for medical analysis and diagnosis.

