



## OPEN ACCESS

## EDITED BY

H. Birkan Yilmaz,  
Boğaziçi University, Türkiye

## REVIEWED BY

Pradnya Kamble,  
K. J. Somaiya Institute of Technology, India  
Arvind Kumar Pandey,  
ARKA JAIN University - University Campus, India

## \*CORRESPONDENCE

Silvia Gaftandzhieva,  
✉ sissiy88@uni-plovdiv.bg

RECEIVED 15 May 2025

ACCEPTED 25 June 2025

PUBLISHED 21 July 2025

## CITATION

Rayyis AM, Maftoun M, Khademi M, Arslan E and Gaftandzhieva S (2025) Optimizing 5G resource allocation with attention-based CNN-BiLSTM and squeeze-and-excitation architecture. *Front. Commun. Netw.* 6:1629347. doi: 10.3389/frcmn.2025.1629347

## COPYRIGHT

© 2025 Rayyis, Maftoun, Khademi, Arslan and Gaftandzhieva. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Optimizing 5G resource allocation with attention-based CNN-BiLSTM and squeeze-and-excitation architecture

Anfal Musadaq Rayyis<sup>1</sup>, Mohammad Maftoun<sup>1</sup>,  
Maryam Khademi<sup>2</sup>, Emrah Arslan<sup>3</sup> and Silvia Gaftandzhieva<sup>4\*</sup>

<sup>1</sup>Department of Computer Engineering, Technical and Engineering Faculty, South Tehran Branch, Islamic Azad University, Tehran, Iran, <sup>2</sup>Department of Applied Mathematics, South Tehran Branch, Islamic Azad University, Tehran, Iran, <sup>3</sup>Department of Computer Engineering, Faculty of Engineering, KTO Karatay University, Konya, Türkiye, <sup>4</sup>Faculty of Mathematics and Informatics, University of Plovdiv Paisii Hilendarski, Plovdiv, Bulgaria

**Introduction:** The swift advancement of computational capabilities has rendered deep learning indispensable for tackling intricate challenges. In 5G networks, efficient resource allocation is crucial for optimizing performance and minimizing latency. Traditional machine learning models struggle to capture intricate temporal dependencies and handle imbalanced data distributions, limiting their effectiveness in real-world applications.

**Methods:** To overcome these limitations, this study presents an innovative deep learning-based framework that combines a convolutional layer with squeeze-and-excitation block, bidirectional long short-term memory, and a self-attention mechanism for resource allocation prediction. A custom weighted loss function addresses data imbalance, while Bayesian optimization fine-tunes hyperparameters.

**Results:** Experimental results demonstrate that the proposed model achieves state-of-the-art predictive accuracy, with a remarkably low Mean Absolute Error (MAE) of 0.0087, Mean Squared Error (MSE) of 0.0003, Root Mean Squared Error (RMSE) of 0.0161, Mean Squared Log Error (MSLE) of 0.0001, and Mean Absolute Percentage Error (MAPE) of 0.0194. Furthermore, it attains an R2 score of 0.9964 and an Explained Variance Score (EVS) of 0.9966, confirming its ability to capture key patterns in the dataset.

**Discussion:** Compared to conventional machine learning models and related studies, the proposed framework consistently outperforms existing approaches, highlighting the potential of deep learning in enhancing 5G networks for adaptive resource allocation in wireless systems. This approach can also support smart university environments by enabling efficient bandwidth distribution and real-time connectivity for educational and administrative services.

## KEYWORDS

wireless networks, resource allocation, 5G, deep learning, attention mechanism, CNN-BiLSTM, squeeze-and-excitation

# 1 Introduction

Advancements in data communication have revolutionized wireless networks, driving an exponential rise in connected devices. By 2030, global data traffic is projected to grow over 20,000 times. While smartphones remain dominant, wearable and smart devices are also increasing, shaping a more mobile, interconnected society with soaring connectivity and traffic demands (Kamal et al., 2021). Against this backdrop, the field of telecommunications is undergoing rapid transformation with the emergence of fifth-generation (5G) technology, which has ushered in a new era of connectivity characterized by ultra-fast data speeds, minimal latency, and vast potential to revolutionize industries (Pathak et al., 2024).

By enabling groundbreaking innovations such as autonomous vehicles, remote surgeries, and immersive virtual reality, 5G is reshaping communication and connectivity on an unprecedented scale. However, as the demand for higher data throughput and more stable connections continues to grow, the implementation of 5G networks requires a reimagined network architecture and efficient resource management to address the increasing traffic and complexity (Khani et al., 2024). The dynamic and diverse nature of 5G presents significant challenges in efficiently distributing network resources like bandwidth, frequency spectrum, and computing power (Pourmabboubi and Tabrizchi, 2024).

Traditional resource allocation methods, which often rely on static or heuristic-based approaches, struggle to meet the real-time demands and fluctuations of 5G networks. These conventional techniques lack the flexibility needed to adapt to evolving network conditions and application requirements. Consequently, more advanced and intelligent solutions are essential for dynamically managing resources, ensuring optimal performance, and maintaining high-quality service (QoS). Maximizing the capabilities of 5G networks is critical to fully harness their transformative potential and meet the evolving needs of a connected world (Tayyaba et al., 2020; Ly and Yao, 2021).

As with previous generations, the increasing number of connected devices and rising data traffic pose challenges for 5G networks, including real-time decision-making and security concerns. Many emerging applications relying on 5G demand high data rates and ultra-low latency, particularly in real-time scenarios. AI technologies offer powerful solutions for handling complex, unstructured problems that involve vast amounts of data, making them highly effective in designing and optimizing 5G networks (Wang et al., 2020).

Machine Learning (ML) as AI technology enhances existing network infrastructure by optimizing resource allocation. Effective real-time decision-making ensures users receive services based on their current needs, reducing resource wastage. The vast amounts of data generated by 5G networks should be leveraged to enhance network performance. Since ML techniques rely on data, their effectiveness depends on the volume and type of data available. It is crucial to consider the application's requirements and the desired accuracy when selecting an ML approach (Bouras and Kalogeropoulos, 2022).

AI-driven resource allocation for 5G networks faces several challenges. Many approaches, including convex optimization,

CNNs, LSTMs, and reinforcement learning, assume static conditions or require large training datasets. High computational complexity limits real-time adaptability in dynamic environments like vehicular and industrial IoT networks. Most models optimize only specific aspects, lacking a multi-objective framework. Reinforcement learning methods minimize energy consumption but suffer from slow convergence and extensive fine-tuning. Traffic prediction models still have high error rates (up to 25%), reducing allocation reliability. Scalability is another issue, as current methods struggle in large-scale, heterogeneous 5G networks, especially in edge computing and SDN environments. Furthermore, AI models often fail to integrate real-world constraints, neglecting unpredictable user behavior, congestion, and varying application demands. Despite their promise, these limitations hinder AI's practical deployment in 5G resource allocation. This gap highlights the need for more robust, adaptive frameworks that can seamlessly integrate with existing network infrastructures and dynamically respond to real-time fluctuations. Moreover, the reliance on extensive datasets for training AI models poses a significant barrier, as acquiring and processing such data in real-time is often impractical. The effectiveness of ML techniques is inherently tied to the quality and quantity of data, and in rapidly changing 5G environments, this dependency can lead to suboptimal performance. Additionally, the trade-off between computational overhead and real-time decision-making remains unresolved, particularly in scenarios requiring ultra-low latency, such as autonomous driving or remote healthcare (Kamal et al., 2021; Pathak et al., 2024; Khani et al., 2024; Pourmabboubi and Tabrizchi, 2024; Tayyaba et al., 2020; Ly and Yao, 2021; Wang et al., 2020; Bouras and Kalogeropoulos, 2022; Zhao, 2023; Hassannataj Joloudari et al., 2024; Efunogbon et al., 2025; Kamruzzaman et al., 2024).

To address these challenges, our study proposes a hybrid model with an attention mechanism, which significantly improves prediction performance and resource allocation efficiency. Our approach not only enhances adaptability and computational efficiency but also achieves the lowest error rates compared to previous studies, ensuring superior performance in dynamic 5G environments. In the sequel, we organize the main contribution of this study as follows:

- We introduce an optimized hybrid deep learning model combining CNN with Squeeze-and-Excitation Block, BiLSTM, and an attention mechanism to enhance predictive performance in 5G resource allocation. This approach captures both local spatial dependencies and long-range temporal correlations, enabling more accurate demand forecasting.
- A custom weighted loss function is incorporated to handle imbalanced resource demand scenarios, emphasizing rare and extreme values to improve robustness in dynamic 5G network conditions.
- Hyperparameter tuning using Bayesian Optimization via Optuna ensures optimal model performance, refining key parameters such as learning rate, dropout rate, and BiLSTM structure.

- We evaluate the proposed model using comprehensive regression metrics, including MSE, MAE,  $R^2$  score, MAPE, and EVS, demonstrating superior performance over baseline models while comparing it with traditional ML models and previous studies.

The rest of the paper is organized as follows: [Section 2](#) presents the background and related works on resource allocation prediction in 5G networks. [Section 3](#) introduces the proposed predictive framework and its core components. [Section 4](#) evaluates the model's performance in terms of throughput regression metrics, comparing different ML techniques. Finally, [Section 5](#) concludes the study and outlines future research directions.

## 2 Related works

This section presents a comprehensive review of recent advances in resource allocation for 5G networks, with a particular emphasis on the pivotal role of artificial intelligence techniques. Given the complexity and dynamic nature of 5G environments, AI-driven methods have become essential for optimizing network performance, including latency reduction, energy efficiency, and spectral utilization. To provide a clear and structured overview, this review is organized into thematic subsections, each focusing on specific AI approaches such as ML, DL, reinforcement learning, and architectural frameworks tailored for heterogeneous and vehicular 5G networks. These subsections highlight the most relevant and impactful research contributions in the field.

### 2.1 Machine learning and deep learning techniques for 5G resource allocation

[Tayyaba et al. \(2020\)](#) pioneered an ML-powered resource management framework for 5G vehicular networks, harnessing the capabilities of Software-Defined Networking (SDN) to revolutionize efficiency. Their methodology integrated a flow-based policy framework with two-tier virtualization to streamline radio resource allocation. The study implemented and rigorously evaluated three DL models: Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), and Deep Neural Network (DNN). Experimental results revealed that LSTM dominated with an accuracy of 99.36%, outclassing CNN (95%) and DNN (92.58%). Simulations were executed using Mininet-WiFi, where SDN controllers optimized Vehicle-to-Everything (V2X) communication by classifying traffic into safety and non-safety flows. The proposed framework dynamically recalibrated bandwidth allocation through a Cuckoo Search Optimization algorithm, ensuring peak throughput and minimized latency. The findings underscored the transformative potential of ML in elevating the efficiency of SDN-based vehicular networks, particularly in addressing real-time traffic demands and refining resource distribution.

[Shukla et al. \(2024\)](#) enhanced resource allocation strategies in 5G networks by employing ML models such as Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Multiple

Linear Regression (MLR). Their methodology involved analyzing a dataset of 5G quality services, concentrating on application type, signal strength, and latency to forecast optimal bandwidth distribution. Their findings revealed that SVM and KNN surpassed MLR in forecasting resource allocation. The R-squared values for SVM and KNN were both 0.5167, while MLR registered a negative score of  $-0.0487$ , reflecting inadequate predictive performance. Furthermore, the Mean Squared Error (MSE) for SVM and KNN stood at 0.01468, considerably lower than MLR's 0.11876, indicating superior accuracy. The Root Mean Squared Error (RMSE) figures for SVM and KNN were 0.1212, in contrast to MLR's 0.3450. These results affirmed that ML-based models enhanced bandwidth efficiency, user satisfaction, and overall network performance, establishing them as more effective than conventional methods for 5G resource allocation.

[Rashid et al. \(2024\)](#) implemented an LSTM-FCN with self-attention, achieving significant improvements in MSE and MAE. However, further enhancements with deeper models and real-world validation are needed. [Yu et al. \(2020\)](#) combined LSTM-based traffic prediction with DRL and convex optimization, optimizing energy efficiency and bandwidth utilization, yet demand prediction and integration with interactive services require improvement. [Alhussan and Towfek \(2024\)](#) utilized bGGO for feature selection alongside an ensemble model, attaining high accuracy ( $R^2 = 99.3$ ), though scalability and real-time adaptability remain challenges. These studies emphasize the necessity for more adaptive, efficient, and scalable solutions to meet the demands of dynamic 5G environments.

### 2.2 Reinforcement learning and optimization approaches in 5G networks

[Zhao \(2023\)](#) investigated resource allocation techniques tailored for edge computing and Industrial Internet of Things (IIoT) applications within 5G networks, introducing an energy-efficient strategy based on deep reinforcement learning (DRL). The proposed model aimed to minimize total energy consumption while maintaining QoS for mobile users. Key decision variables included base station-user associations and transmission power levels. Simulation outcomes revealed accelerated training and convergence rates, with convergence achieved in 500, 350, 250, and 220 steps under varying conditions, accompanied by reward values of 21.76, 21.09, 20.38, and 20.25, respectively. For IIoT settings, Zhao developed an asynchronous advantage actor-critic (A3C) --based allocation framework, utilizing hierarchical aggregation clustering (HAC) to establish base station-user connections and optimize transmission power distribution. Furthermore, Zhao implemented a deep Q-network (DQN)-based approach for ultra-high-definition video transmission in edge computing scenarios. This framework enhanced resource allocation precision, minimized computational inaccuracies, and improved operational efficiency in complex, multi-unit environments. The study underscored the effectiveness of DRL-based methods in addressing the challenges of resource allocation in 5G-enabled edge computing and IIoT systems.

Khumalo et al. (2021) created a resource management model for fog radio access networks (F-RAN) in 5G based on reinforcement learning, tackling the issue of dynamic computational resource distribution. They represented the challenge as a Markov Decision Process (MDP) and applied a Q-learning algorithm to facilitate autonomous resource distribution at fog nodes. The objective of their model was to dynamically reduce latency and enhance resource utilization. The simulation outcomes showcased the efficacy of their method, indicating that the suggested Q-learning algorithm decreased data transmission between fog nodes and the cloud by as much as 90%. Furthermore, their approach achieved a minimum latency of 3 m after 300 iterations, surpassing the performance of SARSA and Monte Carlo algorithms, which needed 450 and 550 iterations to converge at latencies of 4.4 m and 5.4 m, respectively. In addition, their model reached a peak cost efficiency of 92%, in contrast to 88% and 81% for SARSA and Monte Carlo, respectively. Their research underscored the promise of reinforcement learning in enhancing resource allocation within F-RAN.

## 2.3 Architectures and frameworks for heterogeneous and vehicular 5G networks

Cui et al. (2020) proposed an ML-driven resource allocation framework for network slicing in vehicular networks, focusing on minimizing system delay. They framed the radio resource allocation challenge as a convex optimization problem and utilized a Convolutional LSTM (ConvLSTM) model to forecast network traffic patterns. The study examined three traffic categories involving SMS, phone, and web traffic achieving average prediction error rates of 25.0%, 12.4%, and 12.2%, respectively. Leveraging the predicted traffic patterns, the authors implemented a primal-dual interior-point method to dynamically optimize slice weight allocation. Their methodology yielded a notable reduction in overall system delay, with an average decrease of 15.33 m compared to traditional methods lacking traffic prediction. Additionally, the model demonstrated robust adaptability to real-time fluctuations in vehicular network conditions by proactively adjusting slice weights based on anticipated user load distribution.

While in the study of Huang et al. (2022), the authors introduced a DL-driven cooperative resource allocation strategy for 5G wireless networks, designed to enhance resource utilization efficiency. They utilized a convolutional neural network (CNN) to forecast channel conditions and dynamically assign spectrum and antenna resources. The primary objective of their approach was to maximize system throughput while significantly reducing computation time in comparison to conventional techniques. Simulation results revealed that their DL-based framework achieved performance levels comparable to the Minimum Mean Square Error (MMSE) method and surpassed the Zero-Forcing (ZF) approach. The study showed that their model drastically decreased computational time while maintaining near-optimal resource allocation performance. Notably, their method exhibited an average performance degradation of only 1.04% relative to MMSE, while offering enhanced adaptability and faster processing speeds. Additionally, the approach consistently maintained mean square error values between ( $10^{-6}$  and  $10^{-11}$ ) across iterations, ensuring robust reliability

in 5G network scenarios. The findings underscored the potential of DL in optimizing dynamic resource allocation for wireless networks, demonstrating its effectiveness in balancing performance and computational efficiency.

Dong et al. (2020) presented a DL-based framework designed to optimize radio resource allocation in 5G networks, with a focus on minimizing overall power consumption while meeting diverse QoS demands. Their approach utilized a cascaded neural network (NN) architecture: the first NN estimated optimal bandwidth allocation, while the second NN calculated the necessary transmit power to satisfy QoS constraints. The results demonstrated that this cascaded structure outperformed a traditional fully connected NN in maintaining QoS guarantees. Additionally, deep transfer learning was incorporated to enhance the adaptability of NNs in non-stationary wireless environments, substantially reducing the need for extensive training data. Simulations revealed that the cascaded NN framework delivered superior performance, achieving near-optimal power allocation with reduced computational complexity. The framework also exhibited a 99.98% probability of meeting Ultra-Reliable Low-Latency Communication (URLLC) QoS requirements when an additional 10% transmit power was reserved. The study confirmed that their method effectively reduced processing delays and computational overhead, enabling real-time optimization in dynamic 5G network scenarios.

Tayyaba et al., 2020 pioneered an ML-powered resource management framework for 5G vehicular networks, harnessing the capabilities of Software-Defined Networking (SDN) to revolutionize efficiency. Their methodology integrated a flow-based policy framework with two-tier virtualization to streamline radio resource allocation. The study implemented and rigorously evaluated three DL models: Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), and Deep Neural Network (DNN). Experimental results revealed that LSTM dominated with an accuracy of 99.36%, outclassing CNN (95%) and DNN (92.58%). Simulations were executed using Mininet-WiFi, where SDN controllers optimized Vehicle-to-Everything (V2X) communication by classifying traffic into safety and non-safety flows. The proposed framework dynamically recalibrated bandwidth allocation through a Cuckoo Search Optimization algorithm, ensuring peak throughput and minimized latency. The findings underscored the transformative potential of ML in elevating the efficiency of SDN-based vehicular networks, particularly in addressing real-time traffic demands and refining resource distribution.

Bashir et al. (2019) engineered a multi-tier H-CRAN architecture for 5G networks, seamlessly integrating ML to revolutionize resource allocation. Their innovative approach boosted spectral efficiency and QoS by 15%, surpassing conventional systems. The study leveraged a Q-value approximation algorithm to orchestrate intelligent resource management and mitigate interference. Simulations validated that their system enhanced energy efficiency after 400 iterations, delivering superior data rates. When benchmarked against centralized resource allocation and EE-HCRAN, their model outperformed in both spectral efficiency and throughput. The proposed framework masterfully adapted to dynamic network conditions by analyzing user behavior, slashing latency, and amplifying network scalability. Performance evaluations revealed

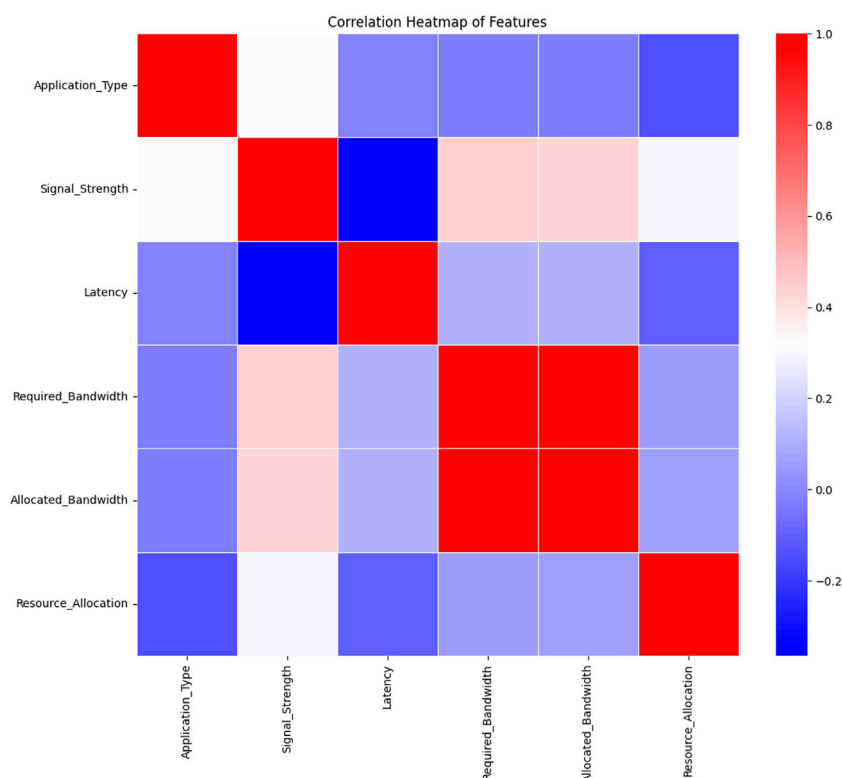


FIGURE 1

The correlation heatmap shows the relationships between features and the target variable.

that spectral efficiency escalated with higher SINR thresholds while power-saving mechanisms significantly eclipsed existing models. Furthermore, their system achieved optimal resource allocation through ML-driven decision-making, guaranteeing uninterrupted service across heterogeneous networks.

In summary, recent studies have explored various ML approaches for resource allocation in 5G networks, each with distinct methodologies and challenges. By considering all aspects of other studies, our model integrates CNN, BiLSTM, and attention mechanisms to enhance 5G resource allocation via capturing spatial and temporal patterns. A custom weighted loss function handles imbalanced demand, while Bayesian Optimization via Optuna fine-tunes key parameters.

### 3 Materials and methods

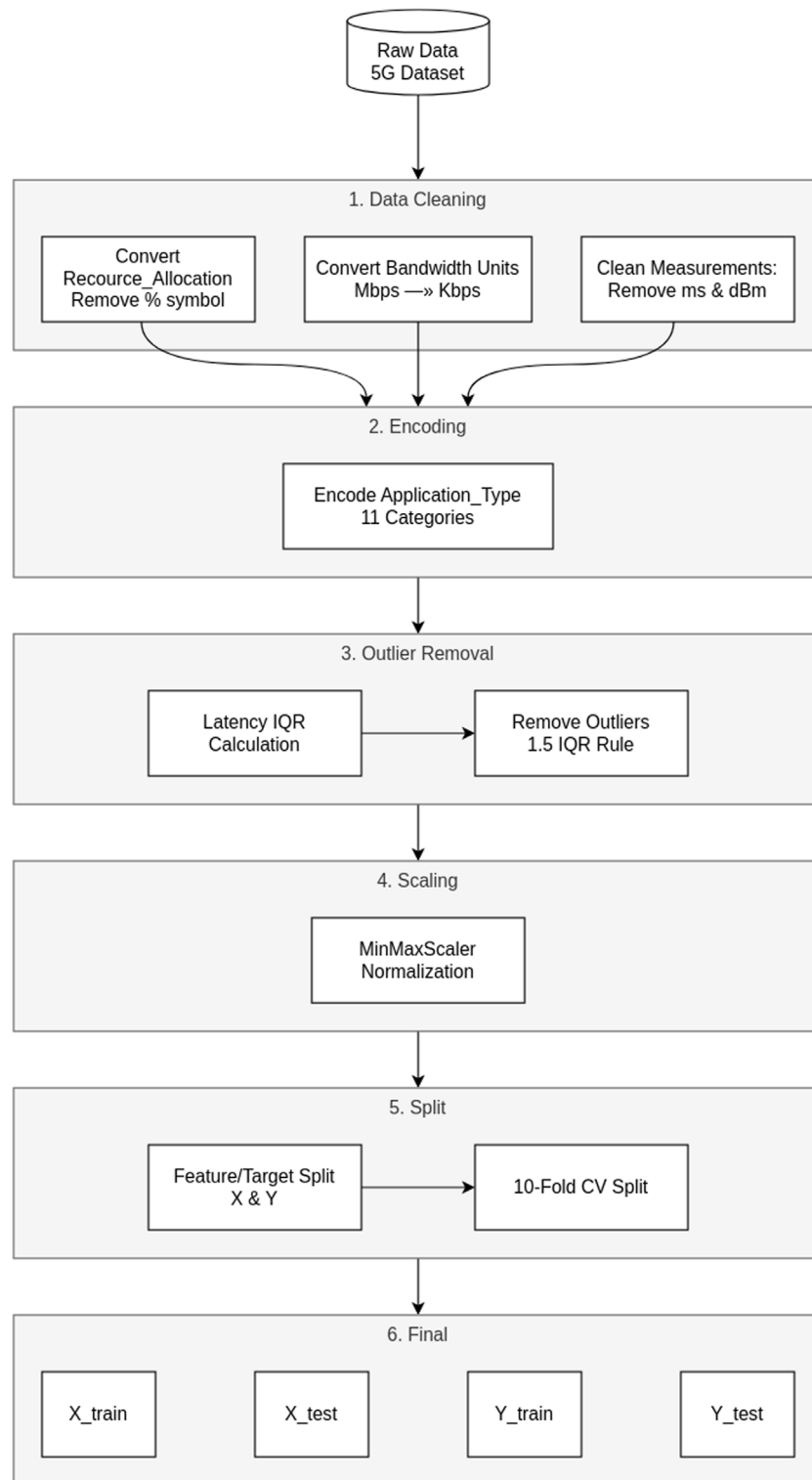
This study utilizes a 5G Resource Allocation Dataset from Kaggle, containing various features related to network performance. Data preprocessing includes handling missing values, encoding categorical features, scaling numeric data, and removing outliers. Model evaluation is performed using cross-validation. Multiple regression models are explored to predict resource allocation. The proposed DL model integrates a hybrid architecture with Conv1D, BiLSTM, and attention mechanisms for feature extraction. For the proposed model, hyperparameter optimization is conducted using Optuna with Bayesian optimization to improve efficiency.

### 3.1 Dataset descriptions

The 5G Resource Allocation Dataset, sourced from Kaggle, consists of 400 samples, each containing key features essential for analyzing and optimizing resource management in 5G networks. It includes attributes such as application types, signal strength, latency, bandwidth requirements including required and allocated, and AI-driven resource allocation decisions. These factors play a critical role in understanding how network resources are distributed dynamically to ensure optimal performance (Cui et al., 2020; Bashir et al., 2019). Figure 1, presents a features correlation heatmap. By examining application-specific demands, signal quality effects, and latency constraints, the dataset facilitates the development of intelligent allocation models that enhance efficiency in 5G networks. Additionally, it supports advancements in QoS optimization and real-time decision-making.

### 3.2 Data preprocessing

The data preprocessing for 5G network resource allocation prediction involves multiple steps. According to Figure 2, First, the dataset is extracted, read, and explored to detect missing values and analyze feature distributions. Categorical features, such as Application\_Type, are encoded using OrdinalEncoder for numerical compatibility. Text-based numerical values, including Resource\_Allocation (percentage), Required\_Bandwidth, Allocated\_Bandwidth, Latency, and Signal\_Strength,



Range (IQR) and removed to improve data quality. A correlation heatmap is used to understand feature relationships, ensuring that strongly correlated variables are considered during model development. Feature scaling is performed using `MinMaxScaler`, normalizing values between

TABLE 1 Hyperparameter Configuration for machine learning models.

Model name	Parameters
GBR	loss = “squared_error”, learning_rate = 0.1, n_estimators = 100, subsample = 1.0, criterion = “friedman_mse”, min_samples_split = 2, min_samples_leaf = 1, min_weight_fraction_leaf = 0.0, max_depth = 3, min_impurity_decrease = 0.0
ABR	n_estimators = 50, learning_rate = 1.0, loss = “linear”
Bagging R	n_estimators = 10, max_samples = 1.0, max_features = 1.0, bootstrap = True, bootstrap_features = False, oob_score = False, warm_start = False
ETR	n_estimators = 100, criterion = “squared_error”, max_depth = None, min_samples_split = 2, min_samples_leaf = 1, min_weight_fraction_leaf = 0.0, max_features = 1.0, min_impurity_decrease = 0.0, bootstrap = False, oob_score = False, verbose = 0, warm_start = False, ccp_alpha = 0.0
HGBR	loss = “squared_error”, learning_rate = 0.1, max_iter = 100, max_leaf_nodes = 31, max_depth = None, min_samples_leaf = 20, l2_regularization = 0.0, max_bins = 255, monotonic_cst = None, warm_start = False, early_stopping = “auto”, scoring = “loss”, validation_fraction = 0.1, n_iter_no_change = 10, tol = 1e-7, verbose = 0
SVR	kernel = “rbf”, degree = 3, gamma = “scale”, coef0 = 0.0, tol = 1e-3, C = 1.0, epsilon = 0.1, shrinking = True, cache_size = 200, verbose = False, max_iter = -1
GPR	alpha = 1e-10, optimizer = “fmin_l_bfgs_b”, n_restarts_optimizer = 0, normalize_y = False, copy_X_train = True
RFR	n_estimators = 100, max_depth = None, min_samples_split = 50, min_samples_leaf = 20, max_features = 3, random_state = 42
DTR	max_depth = 50, min_samples_split = 50, min_samples_leaf = 20, max_features = 3, random_state = 42

0 and 1 for stable learning. The dataset is then split into features and the target before applying 10-fold cross-validation using KFold, ensuring robust training and evaluation.

### 3.3 Machine learning models

After data preprocessing, which involves multiple steps, we employed nine regression models for 5G network resource allocation prediction to evaluate performance and robustness. Gradient Boosting Regressor (GBR) and Histogram-Based Gradient Boosting Regressor (HGBR) leverage boosting techniques to improve predictive accuracy. AdaBoost Regressor (ABR) enhances weak learners by adjusting weights iteratively. Bagging Regressor (Bagging R) and Random Forest Regressor (RFR) use ensemble learning to reduce variance, while Extra Trees Regressor (ETR) increases randomness for better generalization. Support Vector Regressor (SVR) captures complex patterns using kernel-based learning, whereas Gaussian Process Regressor (GPR) models uncertainty for probabilistic predictions. Lastly, Decision Tree Regressor (DTR) provides interpretable results but may overfit data. Table 1 provides a detailed overview of the hyperparameter settings.

### 3.4 Proposed model

The model incorporates Conv1D (32 filters, kernel size = 4) for spatial feature extraction, followed by a Squeeze-and-Excitation (SE) block to enhance feature importance. A three-layer BiLSTM (224 units each, tanh activation) captures temporal dependencies, while a self-attention mechanism refines feature representation. A dropout rate of 0.079 prevents overfitting. The final Dense layer outputs predictions with a linear activation function. The model is trained with a batch size of 8 using an AdamW-based optimizer (learning rate = 0.000935) for optimal convergence. The subsequent sections will delve into further details regarding the model, covering aspects such as fine-tuning, architecture, the network optimizer, and approaches to managing imbalanced distribution.

#### 3.4.1 Finding hyperparameters via Bayesian Optimization

It is essential to set the model’s hyperparameters to achieve optimal predictive performance, before training a DL model. However, relying on experience and extensive trial-and-error to determine these hyperparameters is both time-intensive and computationally demanding. Moreover, this approach does not always guarantee the best possible model performance (Jin et al., 2021). Hence, employing an efficient approach can be useful.

Bayesian optimization is an efficient method for tuning hyperparameters in deep learning. Unlike grid or random search, it builds a probabilistic model of the objective function and selects the next set of hyperparameters based on past evaluations. This is done using a surrogate model (like a Gaussian Process) and an acquisition function that balances exploration and exploitation. The surrogate model estimates the performance of hyperparameter configurations, while the acquisition function selects the most promising ones for evaluation (Shekhar et al., 2021). This approach reduces the number of costly model evaluations by focusing on promising areas in the hyperparameter space. Bayesian optimization is particularly useful for deep learning models, where training is computationally expensive (Frazier, 2018).

The Bayesian optimization process is outlined below.

- 1) For each iteration  $t = 1, 2, \dots$ , repeat:
- 2) Find the next sampling point  $x_t$  by optimizing the acquisition function over the Gaussian Process (GP):

$$x_t = \operatorname{argmax} u(x|D1: t-1)$$

- 3) Obtain a possibly noisy sample  $y_t$  from the objective function  $f$ :

$$y_t = f(x_t) + \epsilon_t$$

Tools like Optuna implement this technique, offering faster convergence to optimal hyperparameters. In 5G resource allocation, where real-time decision-making is essential, Optuna’s define-by-run API enables dynamic and adaptable hyperparameter tuning. It effectively eliminates fewer promising trials, saving valuable computation time—a key benefit for latency-sensitive 5G systems. Table 2 presents the details of the hyperparameter search space for the DL model used in this study. By leveraging insights from previous trials, Optuna prioritizes the most promising hyperparameter settings, enhancing model accuracy and reducing computational demands.

TABLE 2 Hyperparameter search space and final settings for the proposed model.

Hyperparameter	Type	Search space	Description
n_units	Integer	32 to 256 (step = 32)	Number of units in LSTM layers
dropout_rate	Float (log)	0.01 to 0.5	Dropout rate for regularization
learning_rate	Float (log)	1e-6 to 1e-2	Learning rate for optimizer
batch_size	Categorical	{8, 16, 32, 64}	Batch size for training
num_layers	Integer	1 to 4	Number of BiLSTM layers
filters	Integer	32 to 256 (step = 32)	Number of filters in Conv1D layer
kernel_size	Integer	2 to 5	Kernel size for Conv1D layer
optimizer	Categorical	{“adam”, “adamw”, “rmsprop”, “sgd”}	Optimizer selection

### 3.4.2 Architecture

Figure 3 illustrates the architecture of the model, which is described in detail as follows.

- (1) Input Layer: The input consists of data representing resource allocation metrics in a 5G network. The dataset is preprocessed, normalized, and reshaped into a format suitable for sequential modeling. The processed input has a shape of (Independent features, 1 dependent feature) and is fed into the CNN for feature extraction.
- (2) CNN Layer with Squeeze-and-Excitation (SE) Block: A 1D convolutional layer (Conv1D) extracts local spatial patterns from the sequential data, enhancing the model's ability to detect short-term variations in resource utilization. A SE block is integrated to dynamically recalibrate feature importance by applying global average pooling and attention-based weighting to the feature channels. A MaxPooling1D layer reduces the dimensionality, and dropout is applied to prevent overfitting.
- (3) BiLSTM Layer: The extracted spatial features are fed into a multi-layer bidirectional LSTM (BiLSTM) network, which captures both forward and backward dependencies in the data. This enhances the model's ability to learn long-term temporal dependencies in resource allocation patterns. Multiple BiLSTM layers are stacked to improve representation learning.
- (4) Attention Mechanism (AM) Layer: A self-attention mechanism is applied to the BiLSTM output to assign different importance levels. This allows the model to focus on critical moments that significantly impact resource allocation. The attention-weighted features are concatenated with the original BiLSTM output, enriching the final feature representation.
- (5) Output Layer: The combined feature representation is flattened and passed through a fully connected (Dense) layer with a linear activation function, generating the final prediction of network resource allocation. The model is trained using an adaptive optimizer (AdamW), and early stopping is employed to enhance generalization and prevent overfitting.

This hybrid deep learning model effectively integrates spatial feature extraction (CNN), temporal sequence modeling (BiLSTM),

feature recalibration (SE block), and adaptive attention mechanisms, making it highly suitable for accurate and efficient 5G network resource allocation prediction.

### 3.4.3 Network optimizer

AdamW (Zhuang et al., 2022) represents an enhanced version of the Adam optimizer that separates weight decay from the gradient update process, rendering it more efficient for deep learning models, such as those used in predicting 5G resource allocation. The traditional Adam method incorporates L2 regularization by combining weight decay with the gradient update, which results in inconsistent regularization for parameters that have varying historical gradients. AdamW addresses this issue by applying weight decay independently after the gradient adjustment, resulting in improved generalization.

The AdamW update rule is:

$$\begin{aligned}
 \mathbf{m}_t &= \beta_1 \mathbf{m}_{t-1} + (1 - \beta_1) \mathbf{g}_t \\
 \mathbf{v}_t &= \beta_2 \mathbf{v}_{t-1} + (1 - \beta_2) \mathbf{g}_t^2 \\
 \widehat{\mathbf{m}}_t &= \frac{\mathbf{m}_t}{1 - \beta_1^t}, \widehat{\mathbf{v}}_t = \frac{\mathbf{v}_t}{1 - \beta_2^t} \\
 \boldsymbol{\theta}_t &= \boldsymbol{\theta}_{t-1} - \eta \frac{\widehat{\mathbf{m}}_t}{\sqrt{\widehat{\mathbf{v}}_t} + \epsilon} - \lambda \boldsymbol{\theta}_{t-1}
 \end{aligned}$$

where:

- $\mathbf{m}_t, \mathbf{v}_t$  are the first and second moment estimates,
- $\widehat{\mathbf{m}}_t, \widehat{\mathbf{v}}_t$  are bias-corrected estimates,
- $\eta$  is the learning rate,
- $\lambda$  is the weight decay factor.

AdamW addresses this issue by applying weight decay independently after the gradient adjustment, resulting in improved generalization.

### 3.4.4 Handling imbalance distribution

To address the issue of imbalanced distribution in 5G network resource allocation prediction, we employed a weighted mean squared error (WMSE) loss function. Unlike standard MSE, which treats all errors equally, WMSE assigns higher weights to

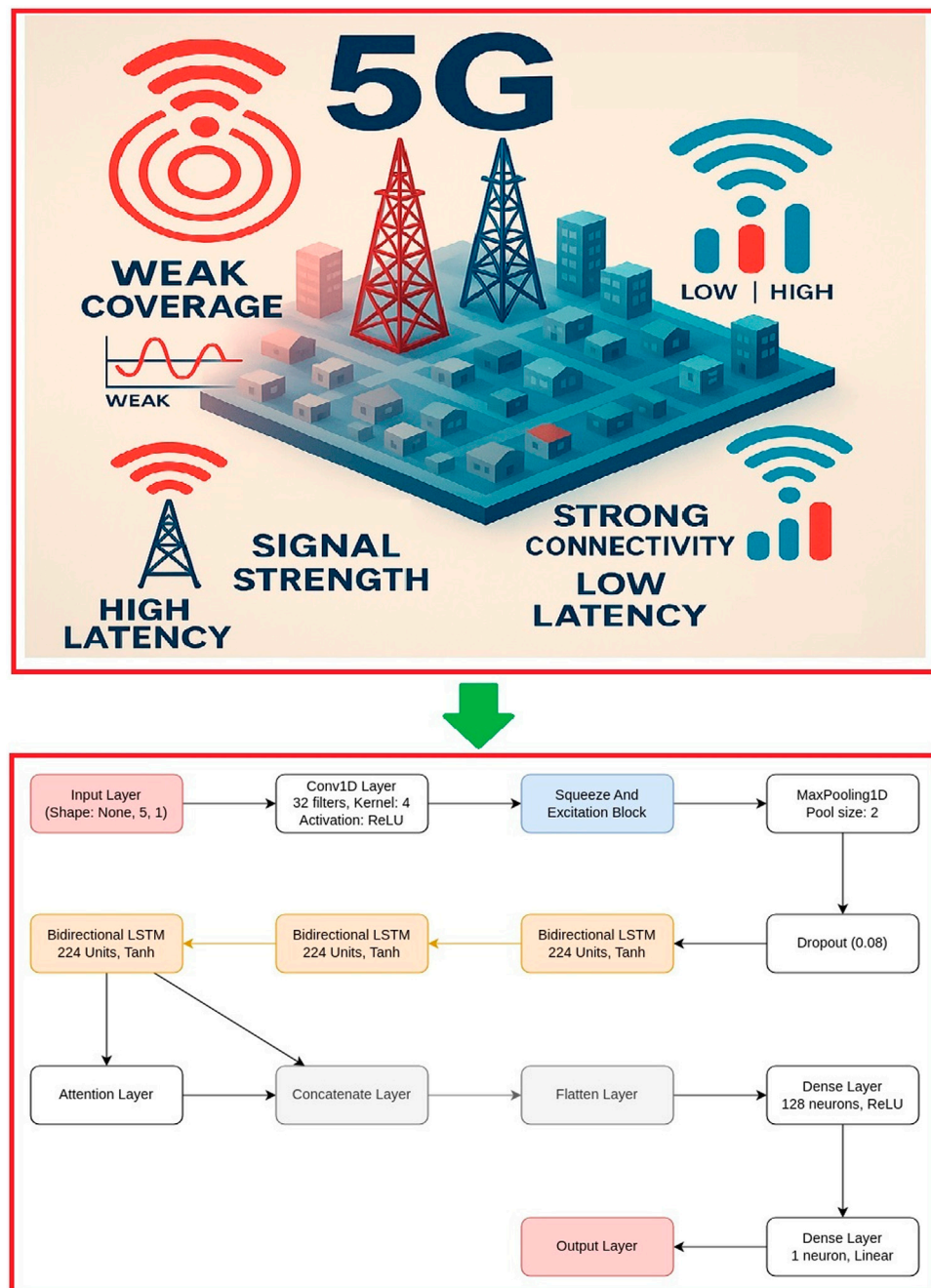


FIGURE 3  
Proposed 5G architecture for efficient resource allocation with latency and signal strength focus.

extreme values, ensuring that the model pays more attention to underrepresented or rare cases (Laxmi Sree and Vijaya, 2018).

The WMSE function used in this model can be mathematically expressed as follows:

$$WMSE = \frac{1}{N} \sum_{i=1}^N w_i (y_i - \hat{y}_i)^2$$

Where the weight for each sample is calculated as:

$$w_i = \exp(|y_i - \bar{y}|)$$

Parameters:

- $y_i$ : True value for sample  $i$
- $\hat{y}_i$ : Predicted value for sample  $i$
- $\bar{y}$ : Mean of all true values
- $w_i$ : Weight assigned to sample  $i$
- $N$ : Total number of samples

The weighting mechanism is based on the absolute deviation of the true values from their mean, scaled using an exponential function. This allows the model to emphasize predictions where the deviation is high, improving learning in regions with scarce data points. By integrating WMSE, the

model achieves better generalization, reducing bias toward majority cases and enhancing prediction accuracy for minority samples in the highly dynamic 5G resource allocation environment.

#### Input:

Dataset: ( $X$ )(independent features),  
( $Y$ )(dependent feature)  
Hyperparameters: ( $n_{\text{units}}$ ), ( $\text{dropout}_{\text{rate}}$ ),  
( $\text{learning}_{\text{rate}}$ ), ( $\text{batch}_{\text{size}}$ ), ( $\text{num}_{\text{layers}}$ ), ( $\text{filters}$ ),  
( $\text{kernel}_{\text{size}}$ ),  
( $\text{max}_{\text{epochs}}$ )

**Output:** Optimized regression model

1. **Initialize:**
2. Define input shape from( $X_{\text{train}}$ ): ( $(n_{\text{features}}, 1)$ )
- 3.
- Split dataset into ( $(X_{\text{train}}, Y_{\text{train}})$ ),  
( $(X_{\text{test}}, Y_{\text{test}})$ ) using K-Fold cross-validation
- 4.
5. **Build Model:**
6.  $\text{inputs} \leftarrow \text{Input}(\text{shape} = (n_{\text{features}}, 1))$
- 7.
- $x \leftarrow \text{Conv1D}(\text{filters},$   
 $\text{kernel}_{\text{size}}, \text{activation} = 'relu')(\text{inputs})$
8.  $x \leftarrow \text{SqueezeExcitation}(x)$
9.  $x \leftarrow \text{MaxPooling1D}(2)(x)$ ,  
( $x \leftarrow \text{Dropout}(\text{dropout}_{\text{rate}})(x)$ )
10. **for**  $i = 1$  to ( $\text{num}_{\text{layers}}$ ) **do**
11.  $x \leftarrow \text{Bidirectional}(\text{LSTM}(n_{\text{units}},$   
 $\text{return}_{\text{sequences}} = \text{True}))(x)$
12. **End for**
13.  $\text{attention} \leftarrow \text{Attention}([x, x])$ ,
14.  $\text{concat} \leftarrow \text{Concatenate}([x, \text{attention}])$
15.  $\text{outputs} \leftarrow \text{Dense}(1, \text{activation} = 'linear')(\text{Flatten}(\text{concat}))$
16.  $\text{model} \leftarrow \text{Model}(\text{inputs}, \text{outputs})$
17. **Compile and Train:**
18. Define( $\text{WeightedMSE}$ )  $\leftarrow \text{mean}(\exp(|y_{\text{true}} - \text{mean}(y_{\text{true}})|) \times (y_{\text{true}} - y_{\text{pred}})^2)$
19. Compile model with loss =  $\text{WeightedMSE}$ , optimizer =  $\text{AdamW}(\text{learning}_{\text{rate}})$ ,
20. metrics
- 21.
- Train(model) on ( $(X_{\text{train}}, Y_{\text{train}})$ ) for ( $\text{max}_{\text{epochs}}$ ),  
 $\text{batch}_{\text{size}}(\text{batch}_{\text{size}})$ , validate on ( $(X_{\text{test}}, Y_{\text{test}})$ ),
22. apply early stopping
24. **Evaluate:**
25.  $Y_{\text{pred}} \leftarrow \text{prediction}(X_{\text{test}})$
26. Compute: MSE, RMSE, MAE,  $R^2$ , MSLE, MAPE, EVS
28. **Output:**
29. Return evaluation metrics

Algorithm 1. Attention-Based CNN-BiLSTM Model.

## 4 Results and discussion

This research employs several vital Python libraries to create and refine the deep learning model for predicting resource allocation in 5G networks. Optuna is utilized for optimizing hyperparameters, effectively identifying the most suitable model parameters. Scikit-learn offers essential evaluation metrics such as MSE, MAE,  $R^2$  score, MSLE, and explained variance in addition to ML models, facilitating a thorough assessment of performance. The deep learning framework Keras, which is built on TensorFlow, supports model implementation by integrating layers like Conv1D, BiLSTM, Attention, and Dense. Furthermore, Keras provides a range of optimizers, including Adam, SGD, Adadelta, and AdamW, to improve training efficiency. TensorFlow functions as the primary engine for computation, making high-efficiency deep learning achievable. NumPy is employed for numerical computations, managing array-based operations that are critical for preprocessing and model calculations. Finally, Matplotlib is utilized to visualize training loss curves and evaluate model performance. These libraries work together to ensure effective model training, hyperparameter optimization, performance assessment, and visualization, rendering them essential for this research.

### 4.1 Evaluation metrics

This section elaborates on the metrics used to assess the effectiveness of the regression models. Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Squared Log Error (MSLE), and Mean Absolute Percentage Error (MAPE),  $R^2$  score and Explained Variance Score (EVS) are utilized in this study (Botchkarev, 2018).

#### 4.1.1 MAE

MAE represents the average of the absolute differences between the actual and predicted values. It provides a straightforward measure of prediction accuracy and is less influenced by extreme errors.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

#### 4.1.2 MSE

MSE calculates the mean of the squared differences between actual and predicted values. Since it squares the errors, it assigns a higher penalty to larger deviations, making it sensitive to outliers.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

#### 4.1.3 RMSE

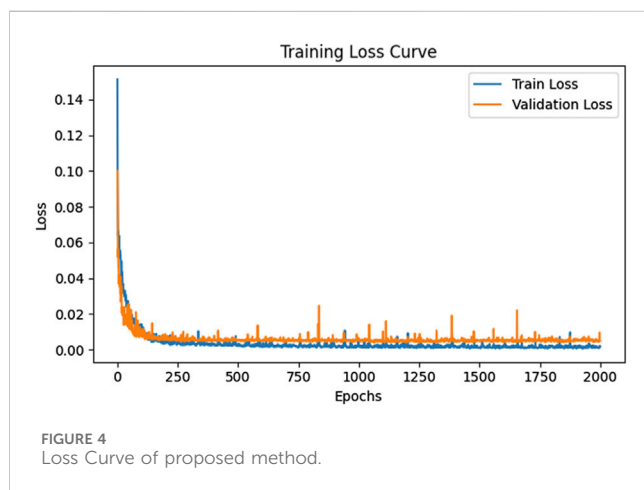
RMSE is the square root of MSE, converting the error back to the same unit as the target variable. It effectively captures large prediction errors while maintaining interpretability.

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

TABLE 3 Comparative predictive performance of different machine learning models on 5G resource allocation.

Model name	MAE	MSE	RMSE	MSLE	MAPE	R <sup>2</sup>	EVS
GBR	0.0158	0.0009	0.0298	0.0003	0.0254	0.9876	0.9884
ABR	0.0511	0.0054	0.0738	0.0020	0.0912	0.9240	0.9280
BaggingR	0.0130	0.0007	0.0272	0.0003	0.0225	0.9896	0.9902
Ridge	0.1971	0.0533	0.2308	0.0204	0.4309	0.2559	0.2809
HGBR	0.0141	0.0006	0.0239	0.0002	0.0243	0.9920	0.9926
SVR	0.0725	0.0061	0.0779	0.0024	0.1546	0.9152	0.9154
GPR	0.0126	0.0010	0.0320	0.0005	0.0250	0.9857	0.9867
RFR	0.0817	0.0108	0.1039	0.0040	0.1528	0.8492	0.8526
DTR	0.0364	0.0054	0.0732	0.0025	0.0729	0.9251	0.9273
Optimized attention-based CNN-BiLSTM (Proposed Method)	<b>0.0087</b>	<b>0.0003</b>	<b>0.0161</b>	<b>0.0001</b>	<b>0.0194</b>	<b>0.9964</b>	<b>0.9966</b>

The bold values represent the results of the proposed model.



#### 4.1.4 MSLE

MSLE computes the squared difference between the logarithm of actual and predicted values. It is particularly useful when dealing

with datasets where lower values are more significant and large variations exist.

$$\text{MSLE} = \frac{1}{n} \sum_{i=1}^n (\log(y_i + 1) - \log(\hat{y}_i + 1))^2$$

#### 4.1.5 MAPE

MAPE expresses the prediction error as a percentage of the actual values, making it unit-independent. However, it can be unstable when actual values are very small.

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100$$

#### 4.1.6 R<sup>2</sup> score

R<sup>2</sup> quantifies how well the model explains the variance in actual values. It ranges from 0 to 1, with higher values indicating better predictive performance.

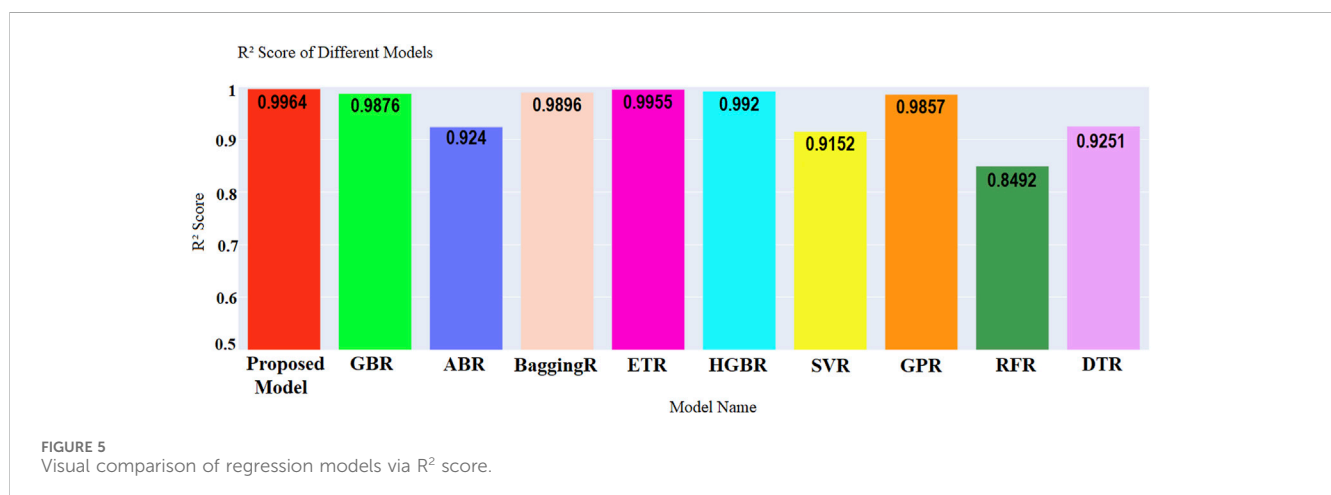


TABLE 4 Benchmarking the proposed approach against existing studies.

Study	Methods	Performances of proposed model
<a href="#">Shukla et al. (2024)</a>	KNN, SVM, MLR	SVM (MSE 0.0353, RMSE 0.1878)
<a href="#">Pathak et al. (2024)</a>	Variational Quantum Regressor (VQR) QML approach	VQR MSE 0.008, RMSE 0.02, best accuracy
<a href="#">Tikar and Patil (2024)</a>	KNN, SVM, MLR DT, GBR models	KNN (R <sup>2</sup> 0.8835, MSE 0.0085)
<a href="#">Pourmabboubi and Tabrizchi (2024)</a>	LSTM	(MSE 0.0012, R <sup>2</sup> of 0.9759)
<a href="#">Alhussan and Towfek (2024)</a>	Feature selection and ensemble model	R <sup>2</sup> 0.993
This Study	Proposed Model	(MAE of 0.0087, MSE of 0.0003, RMSE of 0.0161 R <sup>2</sup> of 0.9964)

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

4.1.1.7 EVS

EVS measures the proportion of variance in actual values that the model successfully captures. Unlike R<sup>2</sup>, it does not account for bias but provides insight into model reliability.

$$EVS = 1 - \frac{Var(y - \hat{y})}{Var(y)}$$

4.2 Prediction results

The experimental results showcase the effectiveness of different models in predicting resource allocation in 5G networks, as provided in [Table 3](#). Based on mentioned metrics, we evaluate the models' performance. Among traditional regression models, Ridge Regression exhibits the weakest performance, with significantly higher errors (MAE = 0.1971, MSE = 0.0533, RMSE = 0.2308) and a poor R<sup>2</sup> score of 0.2559. This suggests that simple linear models are insufficient for capturing the complex patterns in 5G resource allocation.

Ensemble methods such as GBR, Bagging Regressor, and HGBR perform well, with HGBR achieving the second-best results (MAE = 0.0141, MSE = 0.0006, RMSE = 0.0239, R<sup>2</sup> = 0.9920). These results indicate that ensemble techniques improve predictive accuracy, but they still fall short compared to the proposed deep learning approach.

Overall, the optimized attention-based CNN-BiLSTM model demonstrates superior performance, highlighting the advantages of deep learning with attention mechanisms in handling the dynamic and complex nature of 5G resource allocation prediction.

With more than 2.8 million trainable parameters, the proposed model architecture adeptly balances feature extraction, sequence modeling, and attention mechanisms, establishing itself as a robust solution for predicting 5G resource allocation.

The [Figure 4](#) depicts the training loss curve of the proposed optimized attention-based CNN-BiLSTM model designed for predicting resource allocation in 5G networks. The x-axis

corresponds to the number of epochs, and the y-axis displays the loss values. The training loss is represented by the blue line, while the validation loss is shown by the orange line. In the initial stages, both losses are relatively high but decline quickly during the early epochs, reflecting efficient learning. As the training continues, the losses stabilize at a low level, signifying model convergence. Slight fluctuations in the validation loss indicate minor variations in generalization performance, but the model demonstrates robust learning capabilities with negligible overfitting.

According to [Figure 5](#), our proposed method stood out, achieving an impressive R<sup>2</sup> score of 0.9964, surpassing other models. Ensemble models, including the ETR, Bagging Regressor, and HGBR, also performed exceptionally well, each scoring above 0.98. However, models like SVR, ABR, and DTR had slightly lower scores, while the RFR trailed with a score of 0.8492. These results highlight the strength of our approach in efficient resource allocation.

The findings presented in [Table 4](#) illustrate the comparative performance of different models in the context of 5G network resource allocation prediction. [Shukla et al. \(2024\)](#) utilized KNN, SVM, and MLR, with the SVM model yielding an MSE of 0.0353 and an RMSE of 0.1878, reflecting moderate predictive performance. [Pathak et al. \(2024\)](#) applied a Variational Quantum Regressor (VQR), which demonstrated superior performance compared to conventional techniques, achieving an MSE of 0.008 and an RMSE of 0.02. The authors in ([Tikar and Patil, 2024](#)) evaluated several models, with KNN achieving an R<sup>2</sup> of 0.8835 and an MSE of 0.0085, indicating enhanced predictive performance. [Pourmabboubi and Tabrizchi \(2024\)](#) employed LSTM, achieving an MSE of 0.0012 and an R<sup>2</sup> of 0.9759, showcasing robust predictive capabilities. While in [Bashir et al. \(2019\)](#), the authors used hybrid ensemble model with feature selection technique attained 0.993 R<sup>2</sup>. In contrast, our proposed demonstrates superior performance, achieving an MAE of 0.0087, MSE of 0.0003, RMSE of 0.0161, and an R<sup>2</sup> of 0.9964, representing a notable advancement over existing methods. The integration of CNN for feature extraction, BiLSTM for capturing sequential dependencies, AM for adaptive feature weighting, and SE block for feature refinement contributes to the model's enhanced performance and efficiency. These outcomes underscore the efficacy of our approach in advancing 5G network resource allocation, ensuring improved performance and scalability.

## 5 Conclusions and future directions

This research presents an optimized attention-based CNN-BiLSTM model that proficiently forecasts resource allocation within 5G networks. The model exhibits outstanding performance, showcasing an impressively low MAE of 0.0087, MSE of 0.0003, and RMSE of 0.0161, which reflects minimal forecasting errors. Furthermore, the elevated  $R^2$  score of 0.9964 and EVS of 0.9966 validate the model's capability to discern crucial trends in the dataset, rendering it exceptionally reliable for practical use. The congruence between training and validation losses indicates robust generalization abilities, ensuring the model's resilience in ever-changing network settings. Future investigations ought to delve into the realm of 6G networks alongside the incorporation of Large Language Models (LLMs). Given that 6G heralds ultra-reliable, low-latency communication and AI-centric infrastructure, these models must evolve to thrive in dynamic, densely populated settings while facilitating semantic-level resource allocation. LLMs have the potential to act as intelligent, context-sensitive decision-makers, thereby boosting the autonomy and responsiveness of systems. These advancements present thrilling prospects for developing predictive, explainable, and adaptable models tailored for next-generation wireless environments, seamlessly merging sophisticated communication paradigms with the cognitive capabilities of contemporary AI systems.

## Data availability statement

The dataset presented in this study can be found at <https://www.kaggle.com/datasets/omarsobhy14/5g-quality-of-service>.

## Author contributions

AR: Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review and editing. MM: Data curation, Formal Analysis, Investigation, Software, Visualization,

Writing – original draft, Writing – review and editing. MK: Formal Analysis, Project administration, Resources, Supervision, Validation, Writing – review and editing. EA: Validation, Writing – review and editing. SG: Funding acquisition, Project administration, Validation, Writing – review and editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This paper is financed by the European Union-NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, Project No. BG-RRP-2.004-0001-C01.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Alhussan, A. A., and Towfek, S. K. (2024). 5G resource allocation using feature selection and Greylag Goose optimization algorithm. *Comput. Mater. Contin.* 80 (1), 1179–1201. doi:10.32604/cmc.2024.049874
- Bashir, A. K., Arul, R., Basheer, S., Raja, G., Jayaraman, R., and Qureshi, N. M. F. (2019). An optimal multitier resource allocation of cloud RAN in 5G using machine learning. *Trans. Emerg. Telecommun. Technol.* 30 (8), e3627. doi:10.1002/ett.3627
- Botchkarev, A. (2018). Performance metrics (error measures) in machine learning regression, forecasting and prognostics: properties and typology. *arXiv*. 14 45–79. doi:10.48550/arXiv.1809.03006
- Bouras, C., and Kalogeropoulos, R. (2022). Prediction mechanisms to improve 5G network user allocation and resource management. *Wirel. Pers. Commun.* 122 (2), 1455–1479. doi:10.1007/s11277-021-08957-4
- Cui, Y., Huang, X., Wu, D., and Zheng, H. (2020). Machine learning-based resource allocation strategy for network slicing in vehicular networks. *Wirel. Commun. Mob. Comput.* 2020 (1), 1–10. doi:10.1155/2020/8836315
- Dong, R., She, C., Hardjawana, W., Li, Y., and Vucetic, B. (2020). Deep learning for radio resource allocation with diverse quality-of-service requirements in 5G. *IEEE Trans. Wirel. Commun.* 20 (4), 2309–2324. doi:10.1109/twc.2020.3041319
- Efunogbon, A., Liu, E., Qiu, R., and Efunogbon, T. (2025). Optimal 5G network sub-slicing orchestration in a fully virtualised smart company using machine learning. *Future Internet* 17 (2), 69. doi:10.3390/fi17020069
- Frazier, P. I. (2018). A tutorial on Bayesian optimization. *arXiv*. doi:10.48550/arXiv.1807.02811
- Hassannataj Joloudari, J., Mojriani, S., Saadatfar, H., Nodehi, I., Fazl, F., Khanjani, S. S., et al. (2024). Resource allocation problem and artificial intelligence: the state-of-the-art review (2009–2023) and open research challenges. *Multimed. Tools Appl.* 83 (26), 67953–67996. doi:10.1007/s11042-024-18123-0
- Huang, D., Gao, Y., Li, Y., Hou, M., Tang, W., Cheng, S., et al. (2022). Deep learning based cooperative resource allocation in 5G wireless networks. *Mob. Netw. Appl.* 27, 1131–1138. doi:10.1007/s11036-018-1178-9
- Jin, X. B., Zheng, W. Z., Kong, J. L., Wang, X. Y., Bai, Y. T., Su, T. L., et al. (2021). Deep-learning forecasting method for electric power load via attention-based encoder-decoder with Bayesian optimization. *Energies* 14 (6), 1596. doi:10.3390/en14061596
- Kamal, M. A., Raza, H. W., Alam, M. M., Su'ud, M. M., and Sajak, ABAB (2021). Resource allocation schemes for 5G network: a systematic review. *Sensors* 21 (19), 6588. doi:10.3390/s21196588
- Kamruzzaman, M., Sarkar, N. I., and Gutierrez, J. (2024). Machine learning-based resource allocation algorithm to mitigate interference in D2D-enabled cellular networks. *Future Internet* 16 (11), 408. doi:10.3390/fi16110408
- Khani, M., Jamali, S., Sohrabi, M. K., Sadr, M. M., and Ghaffari, A. (2024). Resource allocation in 5G cloud-RAN using deep reinforcement learning algorithms: a review. *Trans. Emerg. Telecommun. Technol.* 35 (1), e4929. doi:10.1002/ett.4929

- Khumalo, N. N., Oyerinde, O. O., and Mfupe, L. (2021). Reinforcement learning-based resource management model for fog radio access network architectures in 5G. *IEEE Access* 9, 12706–12716. doi:10.1109/access.2021.3051695
- Laxmi Sree, B. R., and Vijaya, M. S. (2018). *A weighted mean square error technique to train deep belief networks for imbalanced data*.
- Ly, A., and Yao, Y. D. (2021). A review of deep learning in 5G research: channel coding, massive MIMO, multiple access, resource allocation, and network security. *IEEE Open J. Commun. Soc.* 2, 396–408. doi:10.1109/ojcoms.2021.3058353
- Pathak, P., Oad, V., Prajapati, A., and Innan, N. (2024). Resource allocation optimization in 5G networks using variational quantum regressor. 2024 Int Conf on Quantum Communications, Networking, and Computing (QCNC), 101–105. doi:10.1109/qcnc62729.2024.00025
- Pourmahboubi, A., and Tabrizchi, H. (2024). LSTM-based framework for 5G resource allocation prediction. 2024 11th Int Symp on Telecommunications (IST), 295–302. doi:10.1109/ist64061.2024.10843412
- Rashid, H. U., Chane, A. S., Awoke, L. A., and Jeong, S. H. (2024). Deep LSTM-FCN with self-attention for 5G dual connectivity uplink resource forecasting. 2024 15th Int Conf on Information and Communication Technology Convergence (ICTC), 1017–1021. doi:10.1109/ictc62082.2024.10826773
- Shekhar, S., Bansode, A., and Salim, A. (2021). “A comparative study of hyper-parameter optimization tools,” in 2021 IEEE Asia-Pacific Conf on Computer Science and Data Engineering (CSDE), 1–6. doi:10.1109/csde53843.2021.9718485
- Shukla, N., Siloiya, A., Singh, A., and Saini, A. (2024). “Xcelerate5G: optimizing resource allocation strategies for 5G network using ML,” in 2024 IEEE Int Conf on Computing, Power and Communication Technologies (IC2PCT). 5 417–423. doi:10.1109/ic2pct60090.2024.10486750
- Tayyaba, S. K., Khattak, H. A., Almogren, A., Shah, M. A., Din, I. U., Alkhalifa, I., et al. (2020). 5G vehicular network resource management for improving radio access through machine learning. *IEEE Access* 8, 6792–6800. doi:10.1109/access.2020.2964697
- Tikar, S., and Patil, R. (2024). “Machine learning methods for allocating resources in wireless communication,” in IEEE 9th Int Conf for Convergence in Technology (I2CT), 1–6. doi:10.1109/i2ct61223.2024.10544278
- Wang, C. X., Di Renzo, M., Stanczak, S., Wang, S., and Larsson, E. G. (2020). Artificial intelligence enabled wireless networking for 5G and beyond: recent advances and future challenges. *IEEE Wirel. Commun.* 27 (1), 16–23. doi:10.1109/mwc.001.1900292
- Yu, P., Zhou, F., Zhang, X., Qiu, X., Kadoch, M., and Cheriet, M. (2020). Deep learning-based resource allocation for 5G broadband TV service. *IEEE Trans. Broadcast* 66 (4), 800–813. doi:10.1109/tbc.2020.2968730
- Zhao, S. (2023). Energy efficient resource allocation method for 5G access network based on reinforcement learning algorithm. *Sustain Energy Technol. Assess.* 56, 103020. doi:10.1016/j.seta.2023.103020
- Zhuang, Z., Liu, M., Cutkosky, A., and Orabona, F. (2022). Understanding AdamW through proximal methods and scale-freeness. *arXiv*. doi:10.48550/arXiv.2202.00089