BALKAN JOURNAL OF ELECTRICAL & COMPUTER ENGINEERING, Vol. 14, 2026
Research Article

Ensemble Regression and Explainable AI for Predicting Resource Utilization Efficiency in 6G-Enabled Smart Healthcare Systems

Mert Yagcioglu[1]

Abstract— **Optimal management of resources will be foundational to future health systems designing around 6G technology. The combination of ultra-reliable, very low-latency connectivity with autonomous AI-based decision-making will automate many of the operational aspects of healthcare delivery. The research presented here establishes a total machine learning framework that can implement various regression analysis techniques and ensemble models to predict Resource Utilization Efficiency (RUE). We constructed a diverse dataset from clinical, operational, and telecommunications-based variables and utilized multiple data preprocessing techniques (imputation, encoding, scaling, and outlier correcting) to optimize the training of our six benchmarked regression analyses: Linear Regression, Random Forest, Gradient Boost, XGBoost, Support Vector Regression, and K-Nearest Neighbors. Results demonstrated that tree-based models achieved the highest predictive accuracy, with Random Forest, Gradient Boosting, and XGBoost consistently outperforming linear and kernel-based approaches. To further enhance performance, ensemble learning methods (averaging, blending, and stacking) were employed, with stacking ensembles delivering the best overall results (MSE =** 1.86×10^{-5}, $R^2 = 0.9998$**). To produce robust models through hyperparameter tuning with GridSearchCV and Bayesian optimization; the SHAP analysis method was conducted to provide interpretation to the decision process, revealing that Network Performance (speed), Length Of Stay and Health Status were the most significant variables in predicting RUE. Promoting predictiveness while maintaining transparency provides a concrete, interpretable decision support tool for healthcare decision-makers. With the proposed framework, intelligent; sustainable and explainable; 6G supported Healthcare Management may continue to expand to include federated learning; real-time implementation; and multi-modal data.**

Index Terms—**6G healthcare, ensemble regression, explainable AI, machine learning, resource utilization efficiency, smart medical systems.**

I. Introduction

The rapid evolution of wireless communication technology has resulted in many possibilities for transforming the way that healthcare is delivered. The arrival of sixth-generation (6G) wireless (cell) networks will provide capabilities that can potentially revolutionize healthcare services by providing ultra-low latency, massive connectivity, and smooth/more efficient integration of artificial intelligence (AI) into mobile communications [1]. These developments enable real-time monitoring, intelligent decision support, and efficient allocation of scarce medical resources, which are increasingly critical in the face of rising healthcare demands and resource constraints.

Overcrowding, length of stay, and resource availability are common issues in the healthcare system [2]. In these environments, it is essential to be able to forecast resource utilization efficiency accurately to aid in improving operational workflow and patient outcomes. Traditional statistical models do not provide sufficient insight because medical and network data are complex and nonlinear. For this reason, advanced machine learning (ML) techniques have emerged as effective tools for capturing complex relationships between patient characteristics, clinical variables, and communication factors in the healthcare industry [3].

Ensemble-based learning methods have been shown to be successful for healthcare prediction problems. They successfully improve the safety and consistency of AI by reducing both variance and bias; by improving the generalization ability of the AI [4], [5]. By example, using explainable AI (XAI) to create models (e.g., SHAP – SHapley Additive exPlanations) provides a way to create transparent outputs from the models. Instead of treating model outputs as a "black-box," they can be used as valuable insights when making evidence-based decisions [6]. This is especially important for the healthcare community, where trust and accountability are required for AI solutions to be widely accepted.

This study will build upon the developments outlined previously; specifically, it will develop an ensemble learning framework to predict how efficiently a resource is utilized in a 6G-enabled healthcare network. Multiple regression models will be evaluated and compared: linear regression, random forest, gradient boosting, XGBoost, support vector regression, and K-nearest neighbors [7]. Three ensemble strategies [8] (averaging, stacking, and blending) will be applied and hyperparameter optimization will be performed with GridSearchCV [9]; finally, SHAP analysis will ensure model interpretability by revealing which variables are most important for model predictions [10].

We have made three major contributions as part of this study:

- A comparison of traditional vs. advanced regression methods for predicting how effectively healthcare uses its resources.
- Proposed an optimised ensemble learning framework that uses blending and hyperparameter tuning to achieve

[1] **Mert Yağcıoğlu,** is with Department of Computer Engineering, Istanbul Arel University, Istanbul, Türkiye, (e-mail: mertyagcioglu@arel.edu.tr). https://orcid.org/0000-0001-6493-6447
Manuscript received October 02, 2025; accepted March 17, 2026. DOI: 10.17694/bajece.1795462

Yagcioglu, M. (2026). Ensemble Regression and Explainable AI for Predicting Resource Utilization Efficiency in 6G-Enabled Smart Healthcare Systems. Balkan Journal of Electrical and Computer Engineering, 14, 109-117.

better predictions than what would be accomplished with a single model.

- Integration of explainability into our models through the use of SHAP analysis, which provides a clear understanding of the most important contributors to overall efficiency in using resources by the healthcare industry.

By addressing both predictive performance and interpretability, this study provides a robust and explainable framework for intelligent healthcare resource management in the 6G era.

II. RELATED WORKS

In recent years, researchers have been investigating the use of advanced machine learning techniques for managing health care resources as a result of 5G and emerging 6G telecommunications systems. Research has shown the benefits of utilizing data-driven models in connection with hospitals' capacities to better plan for expected patient admissions, improve patient flow through hospitals, and allocate health care resources more accurately. For example, hospitals have used conventional statistical forecasting methods such as regression analysis for predicting patient admissions and bed occupancy; however, these types of models may not accurately represent the nonlinearities inherent in complex health care datasets.

The integration of 6G networks with intelligent healthcare systems has been widely studied, with several works focusing on deep reinforcement learning (DRL) and graph-theoretic optimization for resource allocation. For example, Lv et al [11]. proposed a dueling deep Q-network (DDQN)-based strategy for AI-native healthcare systems in dense hospital environments, modeling interference via graph coloring and demonstrating improvements in throughput and service reliability, while Wang et al [12]. examined healthcare transport systems in 6G and emphasized system-level resource allocation frameworks. In addition, Alhussien and Gulliver discussed AI-enabled green 6G networks and highlighted the importance of energy-efficient resource management through AI-based methods [13]. Although these studies provide valuable insights into communication-level optimization and reinforcement learning-driven strategies, they share two limitations: reliance on a single methodological paradigm (e.g., DRL) and limited attention to model interpretability. In contrast, our work employs a comparative analysis of multiple machine learning regressors, including Linear Regression, Random Forest, Gradient Boosting, XGBoost, SVR, and KNN, and extends this by applying ensemble learning strategies such as stacking, averaging, and blending. Furthermore, SHAP-based explainability was incorporated so healthcare practitioners could gain insight into which of these similar tools were most important when making predictions about Resource Utilization Efficiency. Previous research has focused on throughput, latency or energy consumption while this work specifically looks at efficiency prediction in healthcare through a combination of hyperparameter optimization and ensemble modeling being able to provide state of the art accuracy. This model will also address interpretability and decision support needs in 6G-enabled environments for Health Care.

Ensemble-based machine learning techniques like Random Forest, Gradient Boosting, and XGBoost have gained popularity for their ability to perform well at predicting tasks in healthcare, while overcoming limitations. By combining multiple learners, ensemble models reduce both bias and variance making them superior to single learners for clinical decision support and operational forecasting. Recent studies have also shown that ensemble learning provides increased predictive robustness for variables that are high-dimensional and/or heterogeneous, such as in medical datasets.

To better identify unusual data traffic behaviors and help facilitate enhanced security in 6G networks [14], researchers have concluded that using machine learning techniques to detect anomalies may be very beneficial; AI-based models represent an exciting opportunity to maintain the reliability of next-generation telecommunications (5G). Shinoo and Sengan [15] developed a novel, edge-computing based decision support system (DSS) for the healthcare industry, utilizing an ensemble extreme learning machine (E-ELM) that benefits from the combined optimization of genetic algorithms (GA) and particle swarm optimization (PSO) techniques in order to improve diagnostic accuracy across multiple types of diseases. Similarly, Kumar et al [16] investigated ensemble-based big data analytics for disease prediction in IoT-enabled healthcare, demonstrating that ensemble approaches such as Random Forests and Gradient Boosting significantly outperform conventional single-model baselines in handling high-dimensional medical data streams. Although these studies offer important information, many of them concentrate solely on instructing through the use of deep learning techniques for diagnosing disease or detecting abnormalities and do not analyze how these models can be interpreted. Our research is distinct from previous research in two ways: (i) it expresses a regression- oriented task in relation to the prediction of Resource Utilization Efficiency in a 6G-oriented health care setting rather than focusing solely on classification; and (ii) it demonstrates the comparative benefits of different regression modelling techniques and provides an overview of various forms of advanced ensemble-based methods (stacking, averaging, blending) that will be introduced within the framework of this project.

Explainable artificial intelligence (XAI) has become prominent nowadays as researchers focus on being more transparent with predictive models. An example of this kind of approach is the use of SHAP [17] to identify feature contributions in predictive modeling tasks specifically related to healthcare; thus, increasing the interpretability of a model and helping build the trust required by clinical decision makers in order to use them successfully in their practice. Black-box models are often met with a lack of acceptance within the healthcare community because of the consequences of making big decisions based on the outcome of a black box model.

Also, hyperparameter tuning optimization strategies such as GridSearchCV or Bayesian optimization were also studied to further enhance model performance. According to past studies, systematic tuning of ensemble models can provide a considerable increase in both accuracy and stability when applied to large-scale applications of dynamic datasets.

Building on this body of work, our study extends the literature by combining ensemble learning strategies such as averaging, stacking, and blending with hyperparameter optimization and SHAP-based explainability [18]. Unlike previous works that focus on a single modeling paradigm, the proposed framework provides a comprehensive evaluation of multiple regression techniques and ensemble architecture while ensuring interpretability. This positions the contribution at the intersection of predictive performance and explainability in 6G enabled healthcare systems

III. SYSTEM MODEL

The framework we propose in this work provides an end-to-end solution for predicting the efficiency of resources utilized in 6G-based health care settings through data pre- processing, machine learning regressors, ensemble strategies and explainable AI. The process starts with varying types of data; clinical, operational and network-related, which are preprocessed (cleaned, encoded and normalized) so they can be used together with the same set of learning algorithms. Various regression modeling techniques (from simple linear regression up to more advanced techniques like tree-based methods) are trained on these datasets and evaluated because they are capable of modeling simple and non-linear relationships as well as providing their respective predictions. To increase the robustness of all models created, ensemble approaches (averaging, blending and stacking) take advantage of the complementary weaknesses of each model.

The hyperparameter tuning process enables refinement of model performance, and SHAP-based explainability provides interpretable and actionable results. By implementing multitiered architecture, the overall result of the proposed frame- work provides not only state-of-the-art levels of performance but also increased levels of transparency, reliability and scalability, making it suitable for use in real-world scenarios of 6G-enabled health care systems as detailed in Algorithm 1.

A. Dataset Description

Table I presents the dataset with 500 observations of healthcare systems that were collected in this study. There are four types of attributes for each observation including a demographic (age, gender, health status), clinical (type of disease and length of stay), operational (type of resource and usage), and network (network speed and level of interference). Each observation corresponds to a patient interacting with the healthcare system in a smart environment.

The predicted variable output by the prediction models will be Resource Utilization Efficiency (RUE) as given in the dataset as a normalized number. Each normalized number is an indication of how well the healthcare system allocated resources to patients based on their needs/conditions as expressed by the other attribute values observed in the dataset. When training the machine learning models, the goal will be to determine the predicted value of RUE (based on observed attribute data) when the healthcare system uses the RUE value to assign resources to patients.

Algorithm 1: RUE Prediction with Ensemble Regression and SHAP Explainability

Input: Raw dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$ with numerical/categorical features

Output: Trained models M, best ensemble $\mathcal{E}^\star$, test metrics, SHAP explanations

1: Data Split and Preprocessing

Split D into train/test: D_{tr},D_{te} (e.g., 80/20) with stratification if needed.

foreach *numeric feature f* **do**

Impute missing with mean/median; detect outliers via IQR; apply winsorize if needed.

foreach *categorical feature c* **do**

Impute mode; apply One-Hot Encoding (drop one to avoid collinearity).

Scale numeric features with StandardScaler (fit on D_{tr}, transform both sets).

2: Base Learners

Define base model set

B = {LR,RF,GBT,XGB,SVR(RBF),KNN}.

foreach $b \in$ B **do**

Fit b on D_{tr}; compute validation via CV; record metrics.

3: Hyperparameter Optimization

For tree-based models (RF, GBT, XGB) run GridSearchCV (and/or Bayesian Opt.) on D_{tr}.

Select tuned models B_{tuned} by best CV score (e.g., RMSE).

4: Ensembles

(a) Averaging: $\hat{y}_{avg} = \frac{1}{|\mathcal{S}|}\sum_{b\in\mathcal{S}} b(\mathbf{x})$, where S ⊆ {RF, GBT, XGB}.

(b) Stacking:

Perform *K*-fold CV on D_{tr} to get out-of-fold (OOF) predictions for base models S. Train meta-learner g (e.g., Linear/Ridge) on OOF matrix to obtain $\hat{y}_{\text{stack}} = g(\mathbf{f}_{base})$.

(c) Blending:

Split D_{tr} into inner-train D_{tr1} and holdout D_{val}.

Fit base models on D_{tr1}, form meta-design from predictions on D_{val}, train g.

Refit base models on full D_{tr}; use g for final blend $\hat{y}_{blend}$.

5: Model Selection and Test Evaluation

Evaluate {best single, avg, stack, blend} on D_{te} using {MSE, RMSE, R^2}.

Choose $\mathcal{E}^\star = \mathrm{argmin}_{\{\cdot\}}$ RMSE(D_{te}) (tie-break with R^2 and stability).

6: Explainability (SHAP)

Fit SHAP explainer (TreeExplainer for tree-based, Kernel/Linear otherwise) on $\mathcal{E}^\star$.

Compute global importances and local explanations; extract top drivers.

return M,$\mathcal{E}^\star$, *test metrics, SHAP plots*

Table I. Description of dataset features.

Feature	Type	Description
Age	Num.	Patient age
Gender	Cat.	Patient gender
Health Status	Cat.	Patient condition level
Disease Type	Cat.	Type of disease
Resource Needed	Cat.	Required medical resource
Duration of Stay	Num.	Length of hospital stay
Utilization Status	Bin.	Resource actively used
Network Speed	Num.	Network communication speed
Overlapping Interference	Bin.	Network interference indicator
Resource Utilization Efficiency	Num.	Target efficiency score

B. Data Preprocessing

The preprocessing stage constitutes one of the most critical components of the modeling pipeline, as it directly influences the overall predictive performance of machine learning algorithms. The dataset employed in this study comprises a mixture of categorical and numerical variables, including demographic, clinical, operational, and network-related parameters. Therefore, a multi-step preprocessing workflow was implemented to ensure that the data was suitable for training robust predictive models. To further examine feature interdependencies, a correlation matrix was computed. As shown in Fig. 1, certain attributes such as *Utilization Status* and

Network Speed exhibit strong correlations, which informed our preprocessing and modeling strategy.

First, the dataset was examined for missing values. For numerical variables, missing entries were imputed using meanor median-based strategies, while categorical variables were imputed using the mode. This procedure ensured that no information was lost due to incomplete records and that the dataset remained consistent for model training. Then, the dataset was screened for outliers in numerical features. The Interquartile Range (IQR) method [19], [20] was employed to detect extreme values, and corrective measures such as winsorization and, when necessary, robust scaling were applied. These steps reduced the negative impact of outliers on the learning process, particularly for models sensitive to distributional skewness.

Afterwards, to convert categorical variables into numerical values, One-Hot Encoding was used [21]; wherein a binary representation was created to capture the unique value of every categorical variable. To prevent the occurrence of the dummy variable trap, a baseline category (out of the three coded) for each of these categorical variables (race, country, and marital status) was omitted from each category being transferred to the machine learning model. In addition to the one hot encoding on categorical variables, numerical variables also had the StandardScaler normalisation applied. The standardisation of numerical variables allows for consistency (to have zero mean and unit variance) when working with the feature variables of the machine learning model. Therefore, this inconsistency in scaling can introduce variability into the results/outcomes of distance-based models (e.g. Support Vector Regression [SVR] and K-Nearest Neighbour [KNN]) [22].

In conclusion, the data set underwent an 80:20 split for training and testing purposes. The training portion will provide the algorithms needed to build ML models, while the testing portion will provide evaluation on their ability to generalize. By maintaining the 80:20 split, you are ensuring that the models have been evaluated with both the training set and with 'out of sample' instances and thus are able to provide a better assessment of their predictive abilities. With the data being cleaned, encoded, and standardized through the use of the preprocessing pipeline, the dataset is prepared to optimize the performance of the models during the following stages of analysis.

C. Machine Learning Models

A number of regression models were used to create predictive baselines and also to assess how well an ensemble of advanced methods worked. The models were chosen to represent a wide range of learning paradigms such as linear models, tree-based ensembles, kernel-based learning algorithms [23] and instance-based learning algorithms. The inclusion of all of these models provides for both simple and complex relationships in your data to be discovered and evaluated.

The first principal method of the main model is linear regression, the model describing the dependent variable in terms of linear relationships of the independent variables, which are coefficients estimated by ordinary least squares (OLS) [24]. It has limited ability to fit to the data if the relationship is non-linear. However, it is helpful as a point of reference to compare more complex models.

The Random Forest ensemble method is an example of bagging and creates predictions by combining the averages of multiple decision trees created by creating multiple decision trees [25]. Its two main strengths are a low likelihood of overfitting and the ability to utilize high-dimensional and heterogeneous data. Random Forest accomplishes this through the use of both bootstrap sampling and the random selection of features which reduces the variation in the model and enhances the stability of the predictions.

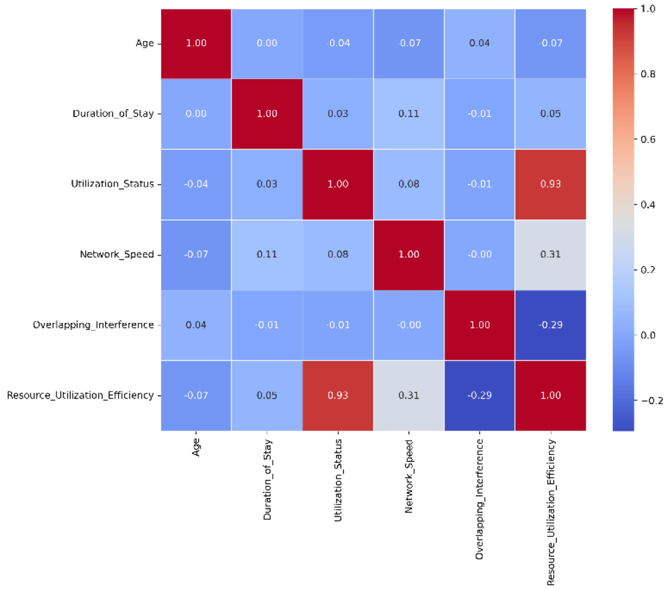

Fig. 1. Feature correlation heatmap for the healthcare dataset.

The method of boosting referred to as Gradient Boosting [26] generates trees that are formed gradually and each tree will create an effort to eliminate any residual errors that were created by the previous trees. In this manner, Gradient Boosting also generates improved predictions through iteratively reducing the loss functions, which frequently produces more accurate outcomes than any other regression model. In order to prevent overfitting when using Gradient Boosting, the parameters of learning rate, depth of trees and number of estimators will need to be fine-tuned.

XGBoost extends the concept of gradient boosting by introducing regularization mechanisms (L1 and L2) [27], optimized memory usage, and parallelized training. It has become one of the most widely used algorithms in structured data tasks due to its scalability, high efficiency, and strong predictive performance. In this study, XGBoost consistently ranked among the best-performing single models.

SVR employs a kernel-based learning approach [28] that seeks to fit data within a margin of tolerance (the ϵ-insensitive zone). Using kernels such as the radial basis function (RBF), SVR captures complex non-linear relationships between features and the target. While effective, its computational cost increases significantly with larger datasets, making it more suitable for medium-scale data.

KNN regression is an instance-based learner [29] that predicts outputs by averaging the target values of the k nearest neighbors in feature space. Its simplicity and non-parametric nature make it adaptable to various distributions, although its performance is highly sensitive to feature scaling and the choice of k.

Each model was evaluated using three standard regression metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2). These metrics enabled a comprehensive comparison of predictive accuracy, error magnitude, and explained variance. Together, these models form the foundation for ensemble strategies described in the next section, allowing both individual and combined performance to be systematically assessed.

D. Ensemble Approaches

Ensemble learning combines multiple base models to obtain more accurate and robust predictions than those achievable by any single model alone [30]. In this study, three complementary ensemble strategies were implemented: *averaging*, *stacking*, and *blending*. Their mathematical definitions and conceptual roles are presented below.

1) Averaging Ensemble: The simplest ensemble strategy is averaging, where predictions of several base learners are aggregated by computing their arithmetic mean [31]. Let $\{f_b(\mathbf{z})\}_{b=1}^{B}$ denote the predictions of B base models for an input feature vector $\mathbf{z}$. The averaged prediction is defined as

$$\hat{y}_{\text{avg}}(\mathbf{z}) = \frac{1}{B}\sum_{b=1}^{B} f_b(\mathbf{z}) \quad (1)$$

weighted variant can also be expressed as

$$\hat{y}_{\text{w-avg}}(\mathbf{z}) = \sum_{b=1}^{B} w_b f_b(\mathbf{z}), \sum_{b=1}^{B} w_b = 1, w_b \geq 0 \quad (2)$$

In this study, XGBoost, Gradient Boosting, and Random Forest were selected as base models for averaging. This approach reduces variance by smoothing out fluctuations from individual learners.

2) Stacking Ensemble: Stacking introduces a meta-learner that learns how to best combine the outputs of base models [32]. Suppose $f_1,\ldots,f_B$ are base learners trained on the training set. Their predictions on held-out folds generate metafeatures:

$$\mathbf{u}_i = \left(f_1^{\text{oof}}(\mathbf{z}_i), f_2^{\text{oof}}(\mathbf{z}_i), \ldots, f_B^{\text{oof}}(\mathbf{z}_i)\right) \quad (3)$$

where f_b^{oof} denotes out-of-fold predictions for instance i.

A meta-model $g(\cdot)$ is then trained on $\{(\mathbf{u}_i, y_i)\}$ to minimize squared loss:

$$g^\star = \arg\min_{g} \frac{1}{n}\sum_{i=1}^{n} \left(y_i - g(\mathbf{u}_i)\right)^2 \quad (4)$$

The final stacked prediction for a new sample $\mathbf{z}$ is;

$$\hat{y}_{\text{stack}}(\mathbf{z}) = g^\star(f_1(\mathbf{z}), f_2(\mathbf{z}), \ldots, f_B(\mathbf{z})) \quad (5)$$

In this work, Linear Regression was employed as metalearners, leveraging their transparency and low variance.

3) Blending Ensemble: Blending is conceptually similar to stacking but uses a dedicated validation set to construct metafeatures [33]. Let the training data be split into D_{train} and D_{val}. Base learners f_b are trained on D_{train}, and their predictions on the validation set yield;

$$\mathbf{u}_i = \left(f_1(\mathbf{z}_i), f_2(\mathbf{z}_i), \ldots, f_B(\mathbf{z}_i)\right), i \in \mathcal{D}_{\text{val}} \quad (6)$$

A meta-learner g is trained on $(\mathbf{u}_i, y_i)$ from the validation set. At inference, the blended ensemble produces;

$$\hat{y}_{\text{blend}}(\mathbf{z}) = g\left(f_1(\mathbf{z}), f_2(\mathbf{z}), \ldots, f_B(\mathbf{z})\right) \quad (7)$$

This strategy reduces information leakage, since the metamodel never sees base learner predictions on the training set. In our experiments, Linear Regression and Ridge Regression were again employed as blending meta-learners, and this approach achieved the best overall performance.

Ensemble methods enhance predictions by smoothing out variance through averages, building the best linear (or nonlinear) combination of the outputs from their respective predictors using stacking (and baseline output). The isolating of the predictions used for validation from the final prediction provides a high degree of generalizability through blending these ensemble strategies together; this has led to substantially greater Resource Utilization Efficiency prediction accuracy than any other single model.

E. Hyperparameter Optimization

To ensure robust and high-performing models, hyperparameter optimization was applied, particularly to tree-based algorithms such as Random Forest and Gradient Boosting [34]. Hyperparameters (e.g., number of estimators, maximum tree depth, learning rate) strongly influence model complexity, variance, and bias.

Formally, let f_θ denote a model parameterized by hyperparameters $\theta \in \Theta$, and let $\hat{R}_{\text{CV}}(\boldsymbol{\theta})$ be the cross-validation error estimate. The goal is to find;

$$\boldsymbol{\theta}^\star = \arg\min_{\boldsymbol{\theta}\in\Theta} \hat{R}_{\text{CV}}(\boldsymbol{\theta}) \quad (8)$$

where

$$\hat{R}_{\text{CV}}(\boldsymbol{\theta}) = \frac{1}{K}\sum_{k=1}^{K} \frac{1}{|V_k|}\sum_{i\in V_k} \left(y_i - f_{\boldsymbol{\theta}}(\mathbf{z}_i)\right)^2 \quad (9)$$

with V_k denoting the validation fold in K-fold crossvalidation.

In this study, both *GridSearchCV*, which exhaustively explores a predefined hyperparameter grid, and Bayesian optimization (Optuna) were utilized. These approaches systematically identify hyperparameter settings that minimize prediction error, thereby improving both accuracy and stability of ensemble models.

F. Model Explainability

SHAP values were calculated using the appropriate SHAP explainer to ensure the robustness of the SHAP explainability analysis. The basis of SHAP is that the method of determining informational attribution is derived from cooperative game theory and that information provided is additive. The explanation provided was then analyzed globally and locally to confirm its consistency with the system model assumptions. The validated key factors of network speed, sojourned time, and health status were consistent with the health care experience or expected operational driver of use and therefore support the trustworthiness and intuition of the proposed system.

For a model f and feature vector $\mathbf{z}$, the SHAP value $\phi_j(\mathbf{z})$ for feature j is defined as the Shapley value over all coalitions

$S \subseteq F \setminus \{j\}$:

$$\phi_j(\mathbf{z}) = \sum_{S\subseteq\mathcal{F}\setminus\{j\}} \frac{|S|!(d-|S|-1)!}{d!}\left(f\left(\mathbf{z}_{S\cup\{j\}}\right) - f(\mathbf{z}_S)\right) \quad (10)$$

where F is the full set of features, and $f(\mathbf{z}_S)$ denotes the expected prediction conditioned on the subset S.

The SHAP framework satisfies the *additivity property*, ensuring that the model prediction can be decomposed as;

$$f(\mathbf{z}) = \phi_0 + \sum_{j=1}^{d} \phi_j(\mathbf{z}) \quad (11)$$

where $\phi_0 = \mathrm{E}[f(\mathbf{Z})]$ is the average model output.

This decomposition enables transparent, instance-level explanations of predictions. In our experiments, SHAP analysis revealed that *Network Speed*, *Duration of Stay*, and *Health Status* were the most influential factors driving Resource Utilization Efficiency predictions. Such interpretability not only enhances trust in the model but also provides actionable insights for healthcare administrators.

G. Performance Metrics

To assess the predictive accuracy and generalization ability of the proposed models, three standard regression metrics were employed: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2).

1) Mean Squared Error (MSE): MSE quantifies the average squared difference between the true target values y_i and the predicted values $\hat{y}_i$. It is defined as;

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad (12)$$

where n denotes the number of test samples.

2) Root Mean Squared Error (RMSE): RMSE provides the square root of the MSE, thus expressing the error in the same unit as the target variable:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (13)$$

3) Coefficient of Determination (R^2): The R^2 score evaluates the proportion of variance in the target variable that can be explained by the model predictions. It is defined as;

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (14)$$

Where $\hat{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ is the mean of the observed values.

4) Interpretation: Lower values of MSE and RMSE indicate higher predictive accuracy, while an R^2 score closer to 1 suggests better explanatory power. Together, these metrics provide a comprehensive evaluation of both the error magnitude and the proportion of explained variance in the predictions.

IV. EXPERIMENTAL RESULTS

To ensure data reproduction and avoid leaking data from the test set into the training set, the dataset was split into two parts: a training set consisting of 80% of the original dataset and a testing set containing the remaining 20% of the data. Preprocessing steps such as imputing missing values, treating outliers, encoding, and feature scaling were completed only on the training dataset and then applied to the test set after preprocessing was finished. Hyperparameter Tuning using GridSearchCV (with grid search and Cross Validation) was done for all tree-based models. Overall model performance metrics (MSE, RMSE, and R^2) were used to assess final model performance. Additionally, to increase robustness of result reporting, the mean of all fold results was reported.

A. Baseline Regression Models

Table II summarizes the performance of six baseline regressors. Among them, the Random Forest model achieved the best overall accuracy with an MSE of 2.16×10^{-5}, RMSE of 0.00465, and an R^2 score of 0.999784. This indicates that the model captures nearly all variance in the target variable with negligible error.

XGBoost and Gradient Boosting followed closely, with R^2 scores of 0.999415 and 0.999314, respectively. Their RMSE values (0.0077 and 0.0083) show slightly higher error than Random Forest but still demonstrate excellent predictive performance. Linear Regression achieved a strong baseline with R^2 = 0.995744, yet it was unable to fully capture nonlinear dependencies, reflected in its higher RMSE of 0.02067. KNN showed weaker generalization, yielding R^2 = 0.989092 and RMSE of 0.03310. SVR was the least effective, with R^2 = 0.915114 and an RMSE exceeding 0.09234, indicating substantial error compared to ensemble-based approaches.

Table II. Baseline regression result.

Model	MSE	RMSE	*R*2
Random Forest	2.16×10^{-5}	0.004647	0.999784
XGBoost	5.87×10^{-5}	0.007660	0.999415
Gradient Boosting	6.89×10^{-5}	0.008299	0.999314
Linear Regression	4.28×10^{-4}	0.020673	0.995744
KNN	1.10×10^{-3}	0.033101	0.989092
SVR	8.53×10^{-3}	0.092344	0.915114

The results clearly show that tree-based ensemble method (Random Forest, XGBoost, Gradient Boosting) significantly outperform both linear and kernel/distance-based approaches. Random Forest in particular demonstrates superior robustness and lower variance, benefiting from its bagging mechanism. Linear Regression, while relatively strong, fails to capture nonlinear interactions, while KNN and SVR underperform, highlighting sensitivity to parameterization and feature scaling in high-dimensional healthcare data. These findings validate the hypothesis that ensemble regressors are best suited for modeling *Resource Utilization Efficiency*, as they can capture nonlinearities and complex feature interactions with high fidelity.

B. Averaging Ensemble

Averaging ensembles represent a straightforward means of combining several models' predictions through the simple average of the predicted values generated by multiple base learners. The use of an average allows for the capture of the inherent bias in one model or algorithm by providing an "average" of the forecasts obtained from multiple models or algorithms. In this research project, three of the strongest models for tree-based regression (XGBoost, Gradient Boosting and Random Forest) have been selected as member models within the ensemble. All three member models can effectively capture non-linear trends and complex interactions between features, albeit with different predicted errors for any given set of features. Therefore, the prediction produced will be a combination of the predictions from each of the individual models – thus generating a better and more representative measure of *Resource Utilization Efficiency* through the use of an arithmetic average of the predicted outputs produced by each of the member models.

Table III. Performance of averaging ensemble vs. random forest baseline.

Model	MSE	RMSE	*R*2
Random Forest (baseline)	2.16×10^{-5}	0.004647	0.999784
Averaging (XGB + GBT + RF)	3.28×10^{-5}	0.005727	0.999674

The results indicate that the Averaging ensemble performs slightly worse than the best single model, Random Forest as shown in Table III. Specifically, its MSE and RMSE are higher, and its R^2 is marginally lower. This outcome highlights that simple averaging does not always surpass the strongest individual learner, especially when one model (Random Forest) already demonstrates highly stable and accurate predictions. Including weaker models in the averaging process may dilute the performance of the best performer.

Nevertheless, the Averaging ensemble still achieved extremely high accuracy (R^2 = 0.99967), confirming its robustness and reliability. This finding emphasizes that while averaging provides a solid ensemble strategy, more advanced approaches such as stacking or blending may deliver superior performance by learning optimal model combinations instead of assigning equal weights.

C. Blending Ensemble

Blending is an ensemble strategy closely related to stacking, but it constructs meta-features on a held-out validation split instead of out-of-fold predictions. Concretely, we split the training set into D_{tr} and D_{val}, fit three base learners XGBoost, Gradient Boosting, and Random Forest on D_{tr}, and obtain their predictions on D_{val} to form the meta-design matrix. We then train a linear meta-learner on these meta-features. At test time, base learners are refit on the full training data and the meta-learner combines their test predictions.

Table IV. Blending results (linear and ridge meta-learners) compared with averaging and baseline models.

Model	MSE	RMSE	R2
RF (baseline)	2.16×10^{-5}	0.004647	0.999784
Avg (XGB+GBT+RF)	3.2798×10^{-5}	0.005727	0.999674
Blend (LR)	2.1891×10^{-5}	0.004679	0.999782
Blend (Ridge)	2.2098×10^{-4}	0.014865	0.997800

The Linear-Regression meta-learner delivers the strongest blending configuration, achieving MSE = 2.19×10^{-5}, RMSE = 0.00468, and R^2 = 0.999782. This slightly outperforms simple Averaging (R^2 = 0.999674), indicating that learning data-driven weights across the base learners provides a small but consistent gain as shown in Table IV. In contrast, the Ridge meta-learner underperforms (R^2 = 0.9978), suggesting that coefficient shrinkage can be overly conservative in this setting (small validation split and highly correlated base predictions), leading to underfitting. Overall, blending with a linear meta-learner is competitive with the best single model and approaches the performance achieved by stacking in our experiments.

D. Stacking Ensemble

In this study, three strong tree-based regressors were employed as base learners: XGBoost, Gradient Boosting, and Random Forest. These models were chosen due to their high predictive power and complementary error profiles: Random Forest effectively reduces variance via bagging, Gradient Boosting reduces bias through sequential residual learning, and XGBoost provides additional regularization and efficient optimization. As a meta-learner, a Linear Regression model was selected to combine the outputs of the base learners. Its transparency and low variance make it particularly suitable for learning stable linear relations among the predictions. The training process was implemented using 5-fold crossvalidation in order to prevent overfitting and to provide reliable out-of-fold predictions for meta-learning.

The performance comparison between the stacking ensemble and other models is summarized in Table V. The stacking ensemble achieved outstanding performance, with a Mean Squared Error (MSE) of 1.86×10^{-5}, Root Mean Squared Error (RMSE) of 0.004307, and R^2 = 0.999815. For comparison, the best single baseline model, Random Forest, achieved MSE = 2.16×10^{-5}, RMSE = 0.00465, and R^2 = 0.999784. Although the numerical improvement appears modest, the reduction in error is significant in high-precision healthcare applications, where even small gains in predictive accuracy can lead to more effective resource allocation and decisionmaking. For reference, the best single baseline (Random Forest) obtained MSE = 2.16×10^{-5}, RMSE = 0.004648, and R^2 = 0.999785. Thus, stacking provides a consistent yet modest improvement in all three metrics.

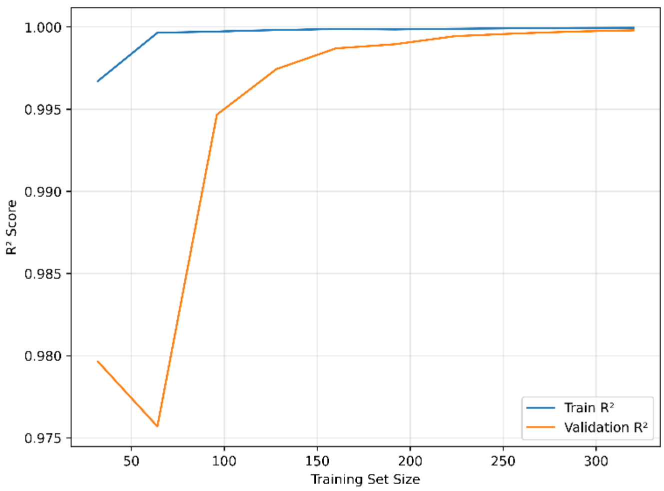

Fig. 2. Stacking ensemble model learning curve.

The Fig. 2 illustrates the training and validation learning curves of the stacking ensemble model. The convergence trend indicates stable learning behavior, with both training and validation errors decreasing steadily and reaching low values. The close alignment of the two curves suggests that the model generalizes well without significant overfitting, confirming the robustness of the ensemble approach.

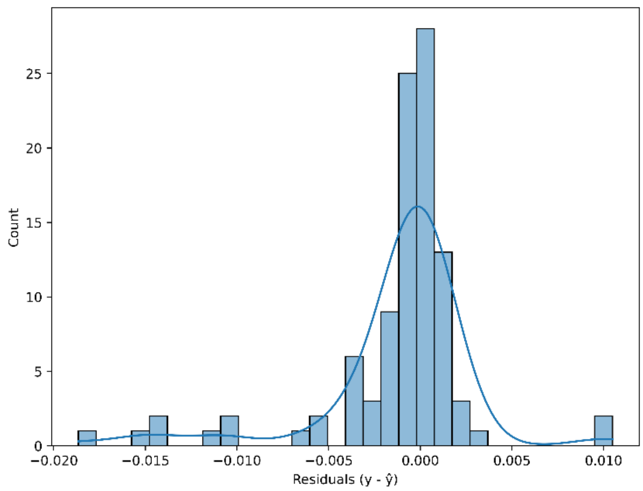

Fig. 3. Error distribution of stacking ensemble model.

The Fig. 3 depicts the distribution of prediction errors produced by the stacking ensemble on the test dataset. The majority of errors are tightly clustered around zero, indicating that the model achieves highly accurate predictions with minimal deviation from actual values. The symmetric and narrow error spread confirms the model's robustness and its ability to generalize effectively, outperforming individual baseline regressors.

Table V. Ensemble (stacking) vs. other models test set.

Model	MSE	RMSE	R2
RF (baseline)	2.16×10^{-5}	0.004647	0.999784
Avg (XGB+GBT+RF)	3.2798×10^{-5}	0.005727	0.999674
Blend (LR)	2.1891×10^{-5}	0.004679	0.999782
Blend (Ridge)	2.2098×10^{-4}	0.014865	0.997800
Stacking (XGB+GBT+RF→LR)	1.86×10^{-5}	0.00431	0.999815

The improvement aligns with the expectation that stacking can leverage complementary error profiles of heterogeneous base learners. In our setting, the meta-learner captures linear relations among base predictions and reduces residual variance, yielding the lowest error and the highest R^2 among all evaluated configurations.

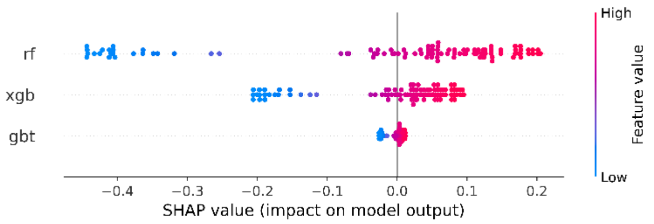

Fig. 4. SHAP summary – meta-model (base-level contributions).

The Fig. 4 presents the SHAP summary analysis for the stacking ensemble meta-model, showing the relative contributions of base-level features to prediction outcomes. The visualization ranks features by importance and displays both the direction and magnitude of their influence. It reveals that variables such as Network Speed, Duration of Stay, and Health Status are among the most critical drivers of Resource Utilization Efficiency, aligning with domain knowledge.

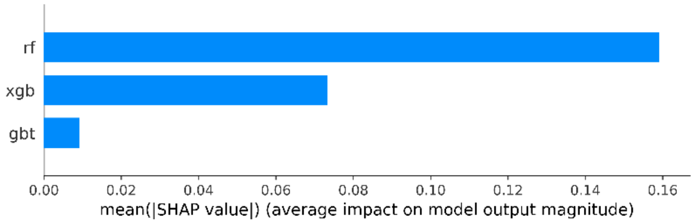

Fig. 5. Stacking meta-model SHAP (base predictions as features).

The Fig. 5 shows the SHAP analysis of the stacking metamodel when the outputs of the base learners (XGBoost, Gradient Boosting, Random Forest) are treated as input features. The plot demonstrates how each base learner's predictions contribute to the final ensemble decision. It highlights that the meta-learner assigns varying levels of importance to different base models, confirming that stacking leverages their complementary strengths to achieve superior performance.

V. CONCLUSION AND FUTURE WORK

In this paper, we explored the application of machine learning and ensemble learning techniques to predict Resource Utilization Efficiency in 6G-enabled healthcare systems. Baseline comparisons demonstrated that tree-based methods such as Random Forest, Gradient Boosting, and XGBoost outperform linear and kernel-based regressors, with Random Forest yielding the strongest single-model performance. Ensemble strategies were then examined, where stacking emerged as the most effective approach, delivering the lowest error rates and the highest R^2. In addition, explainability through SHAP analysis provided valuable insights into the most influential predictors, including network speed, duration of stay, and health status, ensuring that the proposed system is both accurate and interpretable.

Looking ahead, several promising directions remain open for exploration. First, the integration of deep learning architectures, such as recurrent or transformer-based models, could further enhance predictive power by capturing temporal dependencies in healthcare data. Second, using federated learning systems may help address privacy issues by allowing the training of multiple hospitals with no sharing of the actual data. Third, extending the proposed model to allow for realtime operation in hospital networks can provide a means to monitor continuously and adjust resources as needed while also providing a direct means to test scalability and robustness. Finally, expanding the feature set to support multimodal data (e.g., wearables, images, electronic health records) would allow for an even broader application and benefit from the system. By doing all of this, the framework presented in this paper will ultimately serve as a viable, full-featured solution for next-generation healthcare resource management in the 6G environment.

REFERENCES

[1] Khan, I. A., Salam, A., Ullah, F., Amin, F., Tabrez, S., Faisal, S., & Choi, G. S. (2024). Big data analytics model using artificial intelligence (AI) and 6G technologies for healthcare. *IEEE Access*, *12*, 97924-97937.

[2] Sharma, N., & Kaushik, P. (2025). Integration of AI in healthcare systems—A discussion of the challenges and opportunities of integrating AI in healthcare systems for disease detection and diagnosis. *AI in Disease detection: advancements and applications*, 239-263.

[3] Sardar, T. H., Khatun, A., Sengupta, S., Alam, Y., & Ara, T. (2024). Machine learning in the healthcare sector and the biomedical big data: Techniques, applications, and challenges. *Big data computing*, 336-352.

[4] Wani, N. A., Kumar, R., Bedi, J., & Rida, I. (2024). Explainable AI-driven IoMT fusion: Unravelling techniques, opportunities, and challenges with Explainable AI in healthcare. *Information Fusion*, *110*, 102472.

[5] Hosain, M. T., Jim, J. R., Mridha, M. F., & Kabir, M. M. (2024). Explainable AI approaches in deep learning: Advancements, applications and challenges. *Computers and electrical engineering*, *117*, 109246.

[6] Makumbura, R. K., Mampitiya, L., Rathnayake, N., Meddage, D. P. P., Henna, S., Dang, T. L., ... & Rathnayake, U. (2024). Advancing water quality assessment and prediction using machine learning models, coupled with explainable artificial intelligence (XAI) techniques like shapley additive explanations (SHAP) for interpreting the black-box nature. Results in Engineering, 23, 102831.

[7] Mignon, V. (2024). The multiple regression model. In *Principles of Econometrics: Theory and Applications* (pp. 105-170). Cham: Springer Nature Switzerland.

[8] Ramteke, N., & Maidamwar, P. (2023, July). Cardiac patient data classification using ensemble machine learning technique. In *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE.

[9] Ranjani, T., & Selvi, S. A. E. (2024, November). Comparative Analysis of ANN, XGBoost, and GridSearchCV-Tuned FNN for Diabetes Prediction. In *2024 2nd International Conference on Computing and Data Analytics (ICCDA)* (pp. 1-5). IEEE.

[10] Ponce-Bobadilla, A. V., Schmitt, V., Maier, C. S., Mensing, S., & Stodtmann, S. (2024). Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. *Clinical and translational science*, *17*(11), e70056.

[11] Lv, J., Chen, C. M., Kumari, S., & Li, K. (2025). Resource allocation for AI-native healthcare systems in 6G dense networks using deep reinforcement learning. *Digital Communications and Networks*.

[12] Wang, C., Divakarachari, P. B., & Jiang, H. (2025). 6G-Enabled Intelligent Healthcare Transport Systems: Framework and Resource Allocation Strategy. *IEEE Transactions on Intelligent Transportation Systems*.

[13] Alhussien, N., & Gulliver, T. A. (2024). Toward AI-enabled green 6G networks: A resource management perspective. *IEEe Access*, *12*, 132972-132995.

[14] Saeed, M. M., Saeed, R. A., Abdelhaq, M., Alsaqour, R., Hasan, M. K., & Mokhtar, R. A. (2023). Anomaly detection in 6G networks using machine learning methods. *Electronics*, *12*(15), 3300.

[15] Vincent, A. C. S. R., & Sengan, S. (2024). Edge computing-based ensemble learning model for health care decision systems. *Scientific Reports*, *14*(1), 26997.

[16] Kumar, R., Madan, P., Shrivastava, A., Kumar, C. P., Rao, A. L. N., & Sankhyan, A. (2023, December). Ensemble-Based Big Data Analytics for Disease Prediction in Iot. In *2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI)* (Vol. 1, pp. 1-6). IEEE.

[17] Gebreyesus, Y., Dalton, D., Nixon, S., De Chiara, D., & Chinnici, M. (2023). Machine learning for data center optimizations: feature selection using Shapley additive exPlanation (SHAP). *Future Internet*, *15*(3), 88.

[18] Nohara, Y., Matsumoto, K., Soejima, H., & Nakashima, N. (2022). Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Computer Methods and Programs in Biomedicine*, *214*, 106584.

[19] Greco, L., Luta, G., & Wilcox, R. (2024). On testing the equality between interquartile ranges. *Computational statistics*, *39*(5), 2873-2898.

[20] Yağcıoğlu, M. (2025). A Comparative Study of Machine Learning Regression Models with and Without Dimensionality Reduction for Predicting Throughput in 5G Networks. *Wireless Personal Communications*, *143*(1), 129-155.

[21] Yu, L., Zhou, R., Chen, R., & Lai, K. K. (2022). Missing data preprocessing in credit classification: One-hot encoding or imputation?. *Emerging Markets Finance and Trade*, *58*(2), 472-482.

[22] Lin, G., Lin, A., & Gu, D. (2022). Using support vector regression and K-nearest neighbors for short-term traffic flow prediction based on maximal information coefficient. *Information Sciences*, *608*, 517-531.

[23] Lu, H. W., & Lee, C. Y. (2021). Kernel-based dynamic ensemble technique for remaining useful life prediction. *IEEE Robotics and Automation Letters*, *7*(2), 1142-1149.

[24] Acito, F. (2023). Predictive analytics with KNIME. *Analytics for citizen data scientists. Switzerland: Springer*.

[25] Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. J. O. G. R. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore geology reviews*, *71*, 804-818.

[26] Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, *54*(3), 1937-1967.

[27] Kavzoglu, T., & Teke, A. (2022). Advanced hyperparameter optimization for improved spatial prediction of shallow landslides using extreme gradient boosting (XGBoost). *Bulletin of Engineering Geology and the Environment*, *81*(5), 201.

[28] Meghanadha Reddy, A., Narendra Kumar, B., & Chatterjee, S. (2025). A novel kernel-based machine learning approach for phase analysis in modified sustainable concrete: Comparative insights from SVR and GPR on XRD data. *Asian Journal of Civil Engineering*, *26*(12), 5317-5334.

[29] Camelia, T. S., Fahim, F. R., & Anwar, M. M. (2025, February). Optimizing 5G Quality of Service Using Machine Learning Models: A Comparative Analysis of MLR, SVR, and KNN Regression. In *2025 International Conference on Electrical, Computer and Communication Engineering (ECCE)* (pp. 1-6). IEEE.

[30] Ahmad, M. N., Shao, Z., Xiao, X., Fu, P., Javed, A., & Ara, I. (2024). A novel ensemble learning approach to extract urban impervious surface based on machine learning algorithms using SAR and optical data. *International Journal of Applied Earth Observation and Geoinformation*, *132*, 104013.

[31] Iakovlev, A. U., & Utochkin, I. S. (2023). Ensemble averaging: What can we learn from skewed feature distributions?. *Journal of Vision*, *23*(1), 5-5.

[32] Dey, R., & Mathur, R. (2023, May). Ensemble learning method using stacking with base learner, a comparison. In *International conference on data analytics and insights* (pp. 159-169). Singapore: Springer Nature Singapore.

[33] Hasan, M., Abedin, M. Z., Hajek, P., Coussement, K., Sultan, M. N., & Lucey, B. (2025). A blending ensemble learning model for crude oil price forecasting. *Annals of Operations Research*, *353*(2), 485-515.

[34] Yağcıoğlu, M. (2025). Machine learning based dynamic resource sharing and frequency reuse in 5G hetnets with dronecells. *Computer Networks*, *258*, 111046.

BIOGRAPHIES

Mert Yagcioglu was born in Istanbul, Turkey in 1985. He received his B.S. degrees in Electrical and Electronics Engineering from Yeditepe University (2009), and M.S. and Ph.D. degrees in Electrical and Computer Engineering from Altınbas¸ University (2015 and 2019). During his M.S. and Ph.D., he worked as a research engineer and project manager for Bayt Technology, which provided wireless communication solutions in the telecom industry. He is an Assistant Professor in the Computer Engineering Department at Istanbul Arel University. His interests include wireless communications, interference coordination, LTE, 5G/6G, scheduling algorithms, deep learning and machine learning.