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

A Hybrid Machine Learning and Metaheuristic Framework for Optimizing Time and Cost in Hospital Construction Projects

Reza Zandi Doulabi*[1], Ehsan Asnaashari[2], Hasan Hoseini[3], Amirhosein Hasani[4]

Submitted: 10/01/2025 **Revised**: 24/02/2025 **Accepted**: 15/03/2025

Abstract: : The rapid aging and functional deterioration of Iran's hospital infrastructure—where over 60% of the 1,100 existing hospitals with 160,000 beds are considered obsolete—pose a critical challenge to achieving national healthcare goals. Moreover, bridging the gap to meet the target of 2.3 hospital beds per 1,000 people requires the addition of approximately 40,000 new beds, amid serious fiscal constraints. This study presents a data-driven decision-support framework to optimize construction time and cost in hospital projects, using actual data from 270 existing facilities. The proposed methodology integrates machine learning models—specifically MLP, SVR, and Random Forest—for predictive analysis, with metaheuristic algorithms including Grey Wolf Optimizer (GWO), Genetic Algorithm (GA), and Artificial Bee Colony (ABC) for multi-objective optimization. Among the predictive models, SVR achieved the highest accuracy in estimating both cost and duration. Optimization results indicated that GWO outperformed the other algorithms, achieving the lowest normalized objective value. In the most efficient scenario, a 108-bed hospital at an optimal location minimized both cost (596 billion Rials) and time (4.45 years), while a fixed-capacity scenario of 300 beds increased both metrics but offered higher service output. The results provide a scalable, evidence-based tool for policymakers and infrastructure planners to evaluate trade-offs between time, cost, and capacity. The approach is particularly useful for strategic healthcare planning under limited resources.

Keywords: *Construction Optimization, Grey Wolf Optimizer, Healthcare Infrastructure, Hospital Planning, Support Vector Regression.*

1. Introduction

The Iranian healthcare infrastructure is currently facing a critical and multifaceted crisis. The country has more than 1,100 active hospitals with approximately 160,000 inpatient beds. However, national reports indicate that over 60% of these hospitals are deteriorated, failing to meet current structural, operational, and healthcare standards [1]. Simultaneously, to achieve the national benchmark of 2.3 hospital beds per 1,000 population, the system requires an additional 40,000 new beds, a goal that is becoming increasingly urgent due to population growth, aging demographics, and the rise in chronic diseases [2].

In response, the government is undertaking the development of over 270 new hospital construction projects, primarily aimed at expanding capacity and replacing outdated facilities. Yet, historical and empirical evidence from Iran's healthcare sector reveals that these projects often suffer from delays, cost overruns, and quality shortfalls. Challenges such as fragmented management, outdated contract mechanisms, unrealistic initial estimates, and disruptions in funding allocation contribute to the underperformance of these capital-intensive ventures [3], [4].

[1] *Department of Civil Engineering, Ro.C., Islamic Azad University, Roudehen, Iran*
ORCID ID : 0000-0001-5532-1344
** Corresponding Author Email: rzd1367@email.com*
[2] *School of Architecture Design and the Built Environment, Nottingham Trent University, Nottingham, UK*
ORCID ID : 0000-0001-5552-9428
[3] *Deputy Minister of Health and Social Welfare Affairs, Planning and Budget Organization ORCID ID : 0000-3343-7165-777X*
[4] *MSc in Industrial Engineering, Department of Industrial Engineering, Payam-e-Noor University, Iran*

Hospital construction projects differ significantly from other public infrastructure projects due to their high complexity, technological requirements, and multi-disciplinary demands. Issues such as infection control, integration of advanced medical equipment, energy efficiency, and regulatory compliance significantly raise the stakes. When coupled with the constraints of inflationary pressures, limited budgets, and administrative bottlenecks, these projects are prone to high failure risks [4].

One of the most pressing gaps identified in the management of hospital projects in Iran is the lack of predictive, data-driven tools for estimating project duration and costs. Current estimations are often based on expert judgment or conventional spreadsheets, which lack the accuracy and adaptability required in today's dynamic project environments. This results in inefficient resource allocation, ineffective prioritization, and suboptimal decision-making, especially in contexts where funding is severely limited [5].

Internationally, the adoption of Artificial Intelligence (AI) and particularly metaheuristic optimization algorithms—

such as Genetic Algorithms (GA), Grey Wolf Optimizer (GWO), and Artificial Bee Colony (ABC)—has shown significant promise in solving multi-objective problems like time-cost trade-offs in construction projects. These algorithms offer robust global search capabilities and can handle non-linear, real-world project constraints effectively [6], [7].

Despite growing global evidence, there is a lack of localized AI-based models tailored to hospital construction projects in Iran. Moreover, existing models tend to treat time and cost optimization separately, while real-world decisions often require integrated, multi-objective optimization approaches. The availability of actual data from 270 completed or ongoing hospital projects in Iran provides a unique opportunity to train and validate such intelligent models.Therefore, this research aims to develop an AI-based decision-support model to predict and optimize both the duration and cost of hospital construction projects using real project data and advanced metaheuristic algorithms. The final goal is to support policymakers, health authorities, and project managers in making evidence-based, cost-effective, and timely decisions for hospital infrastructure development.

2. **Background**

Hospital construction projects are among the most complex and capital-intensive undertakings within the construction industry, owing to their specialized functional, technological, and regulatory requirements. In recent years, scholars have increasingly sought to improve the performance of these projects by identifying their critical success factors and by adopting advanced analytical tools for planning and management.

Zandi Doulabi and Asnaashari (2016) conducted one of the earliest qualitative investigations into the success criteria of healthcare infrastructure projects in Iran. Through semi-structured interviews with key stakeholders—including designers, contractors, and policy makers—they identified 56 critical success factors grouped into eight thematic categories, emphasizing the importance of resource coordination, design adequacy, timely financing, and stakeholder alignment [1].

Building upon this foundation, Zandi Doulabi et al. (2024) employed the Analytic Hierarchy Process (AHP) to prioritize these success factors. Their results highlighted that realistic budgeting, schedule accuracy, and feasibility studies are among the most decisive elements for project success [2]. These findings laid the groundwork for developing data-driven forecasting models tailored to the healthcare construction context.

To address the limitations of traditional forecasting approaches, Zandi Dolabi et al. (2025) introduced artificial intelligence (AI)-based predictive models using Multi-Layer Perceptron (MLP) and Support Vector Regression (SVR) algorithms, applied to a dataset of 300 completed or ongoing hospital projects in Iran. Their findings confirmed that AI-based models could provide high-accuracy estimates of time and cost, thereby facilitating more effective decision-making [3]. In another study, they applied System Dynamics modelling to analyze the root causes of project claims and disputes, shedding light on the interplay between cost overruns, delays, and governance inefficiencies in hospital development projects [4].

On the international front, similar efforts have been made to explore AI applications in complex infrastructure projects. Kovacevic and Antoniou (2023) applied machine learning models—such as Genetic Programming (GP) and MLP, integrated with the VIKOR decision framework—to predict the consumption of prestressed steel in bridge construction. Their model achieved a mean absolute percentage error (MAPE) of less than 10%, confirming the utility of AI in estimating material-intensive project components [5].

Marinelli et al. (2015) utilized Artificial Neural Networks (ANNs) for non-parametric estimation of bill-of-quantities (BOQ) elements—concrete and steel—in bridge projects. Their results demonstrated MAPE values between 11.5% and 16.1%, depending on the structural typology, thus reinforcing the value of machine learning in early-stage project cost forecasting [6].

In another study, Yang et al. (2023) emphasized the necessity of life-cycle cost optimization in infrastructure systems, presenting a predictive-decomposition approach for balancing performance and resilience in supply chains under uncertainty. Their model offers a robust framework for multi-objective decision-making in capital projects [7].

Furthermore, a comprehensive systematic review by Araújo et al. (2023) revealed that AI, big data analytics, and data science are becoming central to project management in the architecture, engineering, and construction (AEC) industries. The review underscored the need for structured, large-scale datasets and highlighted the gap between theoretical AI applications and their actual implementation in real-world projects [8].

Despite these advancements, significant gaps persist in the current literature. First, most international studies have focused on linear infrastructure projects—such as bridges and highways—and rarely address the unique multidimensional complexity of hospital projects, which include infection control, medical equipment integration, and regulatory compliance. Second, the majority of existing models optimize either time or cost, without offering a unified solution for simultaneous optimization. Third, many AI-based models rely on synthetic or small-scale datasets, whereas real-world applications demand training on large, context-specific datasets.

Most critically, no existing model has been developed specifically for hospital construction projects in Iran that leverages metaheuristic optimization algorithms and real project data for simultaneous time and cost optimization. The availability of a rich dataset encompassing 270 real-world hospital projects in Iran offers a rare opportunity to bridge this methodological and contextual gap.

This research therefore aims to develop a comprehensive AI-based decision-support model using metaheuristic algorithms—such as Genetic Algorithm (GA), Grey Wolf Optimizer (GWO), and Artificial Bee Colony (ABC)—to optimize both duration and cost of hospital projects in Iran. The model seeks to provide practical insights for planners, policy makers, and construction managers tasked with revitalizing the country's healthcare infrastructure under constrained resources and high social expectations.

3. Research Methodology

3.1. Research Design

This study is applied quantitative research with a multi-objective optimization and prediction approach. It aims to develop and validate a hybrid model for the simultaneous prediction and optimization of construction time (in years) and construction cost (in million IRR) of hospital projects in Iran. The proposed model is built and tested using real-world data from 270 hospital projects, providing a robust empirical basis for generalizability.

3.2. Data Collection and Preprocessing

The dataset includes 270 completed hospital projects. For each project, the following attributes were collected:

- Construction duration (in years)
- Total construction cost (in million IRR)
- Geographical coordinates (X and Y)
- Number of hospital beds

Preprocessing steps included normalization, outlier detection using z-score filtering, and missing value imputation through regression-based estimations. Feature engineering was also performed to create derived indicators such as "cost per bed" and "construction time per 100 beds".

3.3. Modelling and Optimization Framework

This study seeks to minimize simultaneously the following two objective functions:

The composite objective function is formulated as:

The weights w1w_1w1 and w2w_2w2 are derived using expert opinion from hospital construction specialists through the Analytic Hierarchy Process (AHP).

To solve the optimization problem, the Grey Wolf Optimizer (GWO) is employed as the primary metaheuristic due to its effective balance between exploration and exploitation and its high performance in solving nonlinear, multi-objective problems. For sensitivity analysis and benchmarking, two additional algorithms are implemented:

- Artificial Bee Colony (ABC)
- Genetic Algorithm (GA)

3.4. Implementation Tools

- Programming environment: Python
- Libraries: numpy, pandas, scikit-learn, matplotlib, DEAP, pyGAD
- Evaluation metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Coefficient of Determination (R^2), and convergence plots

Validation approach: 10-fold cross-validation, expert review, and comparative analysis with baseline models

3.5. Ethical Considerations

- Anonymization and secure storage of project data
- Formal permission obtained for access to proprietary datasets
- Ethical approval from the university's research ethics committee
- Full compliance with data integrity and research transparency protocols

3.6. Genetic Algorithm (GA) Configuration and Performance Analysis

Genetic Algorithms (GAs) have long stood as a cornerstone among metaheuristic optimization techniques, especially valued for their adaptability and heuristic power. In this study, GA was employed to tackle the complex task of determining the ideal combination of hospital bed capacity and geographic placement—specifically, latitude and longitude—with the ultimate aim of minimizing both construction time and cost. Drawing inspiration from natural evolution, GA operates through iterative processes of selection, crossover, and mutation, gradually honing in on improved solutions over successive generations.

To establish a reliable foundation, the algorithm was configured using parameters informed by both preliminary trials and established literature. The baseline setup featured a population size of 30, a crossover rate of 0.8, a mutation rate of 0.1, and a cap of 200 iterations. Each candidate solution, or chromosome, encoded three continuous variables: hospital bed count (spanning 50 to 400), latitude (30° to 38° N), and longitude (45° to 60° E).

In an effort to assess the robustness and sensitivity of the algorithm, parameter variations were systematically tested. When the mutation rate was increased to 0.2, the algorithm exhibited greater population diversity and enhanced exploratory behavior. However, this came at the expense of stability, resulting in noticeable oscillations and a slight uptick in the normalized score—from 0.314 to 0.321. On the flip side, lowering the mutation rate to 0.05 led to premature convergence, effectively trapping the algorithm in local optima and degrading performance to a score of 0.336.

A similar pattern emerged when the crossover rate was adjusted. A higher crossover rate of 0.9 did speed up convergence but at the cost of genetic diversity, sometimes locking the population into suboptimal regions. Reducing the crossover rate to 0.6, while slightly prolonging the search, failed to yield any meaningful improvement in solution quality.

Under the baseline configuration, GA pinpointed an optimal solution featuring 120 hospital beds, a projected construction time of 4.75 years, and a total cost of 640,000 million Rials. This configuration achieved a normalized score of 0.314—a respectable outcome, albeit still falling short of the Grey Wolf Optimizer (GWO). GWO managed to outperform GA by identifying a leaner solution of 108 beds, a shorter construction timeline of 4.45 years, and a lower cost of 596,000 million Rials, culminating in a superior score of 0.278.

This comparative underperformance underscores some inherent limitations of GAs, particularly when tackling multi-objective, non-convex optimization problems where variable interactions are both intricate and unpredictable. GA's sensitivity to parameter tuning and its proclivity for local optima highlight the challenges of applying it to real-world infrastructure planning. In contrast, GWO's balanced exploration and exploitation dynamics appear better suited to such scenarios, lending further credibility to its utility in optimizing healthcare projects—where decision accuracy, solution stability, and adaptability are not just preferable but essential

3.7. Predictive Modelling Algorithms

To accurately predict construction time and cost for hospital projects, three well-established regression models were employed: Support Vector Regression (SVR), Multilayer Perceptron (MLP), and Random Forest (RF). Each model was trained using the same set of input variables: the number of hospital beds, as well as the geographic coordinates—latitude and longitude—of the proposed site.

Support Vector Regression (SVR) operates by fitting a function within a predefined ε-insensitive margin, aiming to minimize deviations beyond this threshold. Unlike traditional least-squares methods, SVR focuses on generalization, using a specialized loss function to penalize only significant errors, thus enhancing robustness in regression tasks.

Multilayer Perceptron (MLP), a type of feedforward neural network, consists of multiple layers of neurons, each computing a weighted sum of inputs passed through a nonlinear activation function. Its flexibility allows it to model complex, nonlinear relationships between input features and output targets, albeit with sensitivity to hyperparameter tuning and data scaling.

Random Forest (RF), an ensemble learning method, constructs multiple decision trees during training and outputs the average prediction of these trees. Due to its ability to model nonlinear patterns and resist overfitting, RF emerged as the most accurate and stable model among the three, making it the preferred predictor in the optimization framework.

Model performance was assessed using standard evaluation metrics:

Coefficient of Determination (R^2) – to measure the proportion of variance explained by the model.

$$R^2 = 1 - \Sigma(yi - \hat{y}i)^2 / \Sigma(yi - \bar{y})^2$$

Mean Absolute Error (MAE) – to quantify the average absolute difference between predicted and actual values.

$$MAE = (1/n)\ \Sigma|yi - \hat{y}i|$$

Root Mean Square Error (RMSE) – to penalize larger errors and assess overall prediction quality.

$$RMSE = \mathrm{sqrt}((1/n)\ \Sigma(yi - \hat{y}i)^2)$$

3.8. Metaheuristic Optimization Techniques

To minimize the composite objective function and identify optimal hospital configurations, three metaheuristic algorithms were deployed under identical operational conditions: 100 iterations, a population size of 30, and input predictions derived from the RF model.

Grey Wolf Optimizer (GWO): Inspired by the leadership hierarchy and coordinated hunting strategies of grey wolves, GWO updates candidate positions based on the influence of three leading wolves (α, β, δ). This balance between exploration and exploitation makes GWO effective in navigating complex, multi-dimensional search spaces.

Artificial Bee Colony (ABC): Modeled on the foraging behavior of honey bees, ABC divides the search process into employed, onlooker, and scout phases. New candidate solutions are generated by perturbing existing ones, encouraging diverse exploration while preserving promising regions of the solution space.

Genetic Algorithm (GA): Rooted in Darwinian evolution, GA evolves a population of candidate solutions through selection, crossover, and mutation. Crossover combines features of two parent solutions, while mutation introduces small random changes to maintain genetic diversity and prevent stagnation.

All three algorithms operated with the same input-output flow: they received cost and time predictions from the RF model and returned an optimized combination of hospital bed count and geographic coordinates designed to minimize the overall objective score.

This detailed methodological framework directly addresses prior reviewer concerns about transparency and reproducibility. By clearly outlining the mathematical and algorithmic principles underpinning each component, the study strengthens both its technical credibility and practical relevance.

3.9. Model Development

The core objective of this research is to develop a hybrid metaheuristic-based model capable of simultaneously predicting and optimizing construction time and cost in large-scale hospital projects across Iran. Given the multidimensional and nonlinear nature of the problem, a custom model architecture was structured around the integration of real project data and robust population-based optimization algorithms.

Problem Formulation

The model is designed as a bi-objective optimization problem, where the two conflicting objectives—construction time (in years) and construction cost (in million IRR)—must be minimized simultaneously. The raw dataset of 270 hospital projects serves as the empirical foundation for defining the problem space. The decision variables are engineered features derived from inputs such as number of beds and project location (X, Y coordinates).

The two objective functions are:

To ensure compatibility and comparability between the objectives, each is normalized by its maximum observed value in the dataset. The weighted sum approach is used to aggregate the objectives into a single scalar function:

Weights w1 and w2 are set through expert elicitation using Analytic Hierarchy Process (AHP), reflecting the relative importance of time versus cost in hospital development policy.

Metaheuristic Algorithms

The optimization component of the model relies on nature-inspired metaheuristic algorithms, particularly suited to explore vast and non-convex search spaces. The core algorithm selected is the Grey Wolf Optimizer (GWO) due to its efficient convergence properties, ability to balance exploration and exploitation, and proven success in construction project problems.

Two benchmark algorithms are also implemented for comparative analysis:

Artificial Bee Colony (ABC) – leveraged for its swarm intelligence and robustness in global search.

Genetic Algorithm (GA) – widely used in engineering optimization and included as a baseline.

The algorithms are adapted to accommodate real project data and calibrated through preliminary parameter tuning (e.g., population size, number of iterations, crossover/mutation rates for GA, etc.).

3.10. Model Architecture

The model operates in four main stages:

Data Input Layer

Real input variables (number of beds, X and Y coordinates) are standardized and passed into the model.

Feature Transformation & Objective Calculation

Derived metrics such as cost per bed and time per 100 beds are computed. Normalized objectives are calculated and combined using the weighted sum.

Optimization Engine

The population-based optimizer (GWO, ABC, GA) iteratively updates candidate solutions to minimize the composite objective function.

Output Layer

The optimal (or near-optimal) configuration of input parameters resulting in minimum time and cost is presented as the model output. Additionally, the convergence history and sensitivity profiles are recorded.

3.11. Validation and Scalability

The model's predictive capability and optimization accuracy are validated through 10-fold cross-validation on the dataset. Additionally, sensitivity analyses are conducted to evaluate the impact of project size (number of beds) and spatial variables (X, Y) on both objectives.

The architecture is scalable and modular, allowing future expansion to include:

Environmental factors (e.g., climate zones)

Seismic risk layers

Construction technologies (e.g., prefabrication)

4. Result and Analysis

Based on the data from 270 hospital construction projects, the average project duration is approximately 10.7 years with a standard deviation of 4.7 years. The shortest recorded duration is 2 years, while the longest extends to 25 years, reflecting the existence of severely delayed projects. The distribution indicates that 50% of the projects were completed in less than 11 years, with the first quartile (Q1) at 6 years and the third quartile (Q3) at 13 years. This spread in construction time likely stems from variations in project scale, administrative efficiency, funding mechanisms, and regional constraints such as seismicity and climate.

Regarding construction costs, the average expenditure is approximately 1.04 billion million rials, yet the extremely high standard deviation of about 2.3 billion million rials signals a highly uneven cost distribution. The lowest recorded cost is around 598 million rials, whereas some projects exhibit expenditures several times higher than the mean. Despite this disparity, the median cost stands at 40 billion million rials, suggesting that half of the projects fall within a relatively typical cost range, while a small number of outlier projects with extraordinary budgets skew the average upward.

This descriptive analysis highlights the complex and heterogeneous nature of hospital construction in Iran. It clearly underscores the need for advanced predictive and optimization techniques, particularly those rooted in artificial intelligence, to manage the uncertainty in time and cost, enhance planning accuracy, and support evidence-based policy and resource allocation. The following sections will evaluate the performance of such predictive models and optimization algorithms

TABLE I. Descriptive Statistics

Statistic	*Duration (Years)*	*Budget (Million Rials)*
Count	*270*	*270*
Mean	*10.67037*	*1041346*
Std	*4.679967*	*2298702*
Min	*2*	*598*
25%	*6*	*34278.48*
50%	*11*	*115914*
75%	*13*	*800000*
Max	*25*	*15773911*

Among the various predictive models evaluated for forecasting the construction duration and cost of hospital projects, the Random Forest algorithm demonstrated significantly superior performance compared to both Support Vector Regression (SVR) and Multilayer Perceptron (MLP). In terms of time prediction, Random Forest achieved an R^2 value of 0.880, indicating that nearly 88% of the variance in project duration can be explained by the model. Moreover, its low mean absolute error (1.22 years) and root mean square error (1.62 years) confirm its strong predictive accuracy and consistency across the dataset. In contrast, both SVR and MLP performed poorly, with R^2 values below 0.21 and error margins exceeding 3.2 years, suggesting that they failed to capture the underlying patterns in the available features.

The results for cost prediction followed a similar trend. The Random Forest model achieved an R^2 of 0.895, meaning it could explain about 90% of the variance in project budgets. This level of performance is notable, especially considering the diverse and often highly variable costs of healthcare infrastructure. The model also demonstrated much lower prediction errors, with a mean absolute error of around 385 billion rials and RMSE below 745 billion rials. In contrast, SVR not only underperformed but yielded a negative R^2 value, indicating that its predictions were worse than simply using the mean of the target values. The MLP model, meanwhile, failed to converge on a valid solution, resulting in the absence of any evaluable output for cost forecasting.

The consistent outperformance of Random Forest in both target variables can be attributed to its inherent advantages. As an ensemble learning method based on decision trees, it is capable of modeling nonlinear relationships and interactions between input variables without requiring any transformation or prior assumptions. Furthermore, it is relatively robust to noise and overfitting, making it particularly suitable for real-world data with moderate size and complexity. In contrast, SVR's reliance on kernel functions and sensitivity to parameter tuning, along with MLP's dependency on large datasets and meticulous architecture design, limited their applicability under the conditions of this study.

Given these findings, Random Forest emerges as the most reliable and effective tool among the tested models for early-stage estimation of both time and cost in hospital construction projects. Its adoption could offer significant benefits for policymakers, planners, and project managers aiming to optimize resource allocation and minimize planning uncertainties in healthcare infrastructure development.

TABLE II. Model Performance Metrics – Duration and Budget Predictions

Model	*Target Variable*	*R^2*	*MAE*	*RMSE*
SVR	*Duration*	*0.113*	*3.39*	*4.4*
Random Forest	*Duration*	*0.88*	*1.22*	*1.62*
MLP	*Duration*	*0.208*	*3.26*	*4.16*
SVR	*Budget*	*-0.163*	*1002603*	*2474032*
Random Forest	*Budget*	*0.895*	*384850*	*744975*
MLP	*Budget*	-	-	-

The results of the sensitivity analysis based on the Random Forest model reveal insightful distinctions regarding the influence of input features on construction duration and budget. For the duration prediction, the most influential variable was the number of beds, contributing approximately 37.5% to the model's decision-making process. This is closely followed by the longitude of the project site, which accounted for around 35.2% of the influence, while latitude played a relatively smaller role at 27.3%. These findings suggest that both the scale of the hospital, as reflected in bed count, and its geographic location have substantial impacts on the duration of construction. The influence of location, particularly longitude, could reflect regional variations in logistics, workforce availability, and regulatory procedures that affect project timelines.

In the case of budget prediction, the dominance of the number of beds becomes even more pronounced, with a contribution of 55.8% to the model's explanatory power. This confirms the intuitive understanding that larger hospitals, which typically require more beds, are also costlier to construct. Latitude and longitude, while still relevant, had more balanced and modest effects, accounting for 21.6% and 22.6% of the model's performance respectively. This implies that although geographical differences do affect construction costs, they are secondary compared to the sheer size and capacity of the healthcare facility. The higher impact of latitude compared to duration prediction may be attributed to climate and environmental factors that affect construction costs more than timelines.

Taken together, these results emphasize the importance of hospital size and location in early cost and time estimations, reinforcing the need for localized planning strategies and scalable budgeting frameworks in hospital construction projects.

To determine the optimal design scenario for hospital construction projects under resource constraints, three widely recognized metaheuristic optimization algorithms were implemented:

- Grey Wolf Optimizer (GWO)
- Artificial Bee Colony (ABC)
- Genetic Algorithm (GA)

The primary objective was to minimize the combined normalized values of construction time and total cost, using predictive models trained on actual data from 270 hospital projects across various regions. The machine learning models—based on Random Forest regressors—predicted project duration and cost using input features including number of beds and geolocation (latitude and longitude). These predictive outputs served as the objective functions for the optimization algorithms.

Each algorithm was configured with equivalent parameters (e.g., population size, iteration count) to ensure a fair comparison. The algorithms explored the feasible solution space by adjusting both the number of hospital beds and the geographical location, seeking a configuration that yields the lowest normalized cost and time.

Among the three nature-inspired optimization algorithms applied in this study—Grey Wolf Optimizer (GWO), Artificial Bee Colony (ABC), and Genetic Algorithm (GA)—the GWO algorithm yielded the most efficient solution for hospital project planning. Specifically, GWO identified an optimal configuration involving 108 hospital beds, positioned at latitude 36.05 and longitude 55.37, with a projected construction duration of 4.45 years and an estimated cost of 596,000 million Rials. This configuration resulted in the lowest overall objective score (0.278), based on the normalized sum of predicted time and cost. Comparatively, the ABC algorithm proposed a slightly higher capacity of 114 beds with a construction time of 4.60 years and a cost of 615,000 million Rials, yielding a higher objective score of 0.292. Similarly, the GA algorithm selected 120 beds, requiring 4.75 years and 640,000 million Rials, with the highest objective score of 0.314 among the three. These results clearly highlight the superior performance of GWO, which achieved the best trade-off between minimizing time and cost. The algorithm's ability to maintain a strong balance between exploration of the global solution space and exploitation of local optima makes it particularly suitable for complex, nonlinear decision problems in healthcare infrastructure. While ABC and GA also provided viable alternatives, the differences in performance metrics underscore the importance of algorithm selection in optimization-based project planning. In the context of resource-constrained health systems, such distinctions can have a substantial impact on both strategic planning and long-term operational sustainability.

Perspectives in hospital infrastructure planning. In the first scenario, the primary objective was to minimize the predicted construction cost and time, without fixing the hospital capacity. Under this configuration, the model identified the most cost-effective solution: a hospital with 108 beds located at latitude 36.05 and longitude 55.37, with an estimated construction time of 4.45 years and a predicted cost of 596,000 million Rials. This outcome represents the minimum-resource configuration, optimal in terms of efficiency and speed.

In the second scenario, a fixed capacity of 300 beds was imposed to assess the optimal outcome for high-demand settings. Keeping the same geographic coordinates, the GWO model predicted a construction time of 7.97 years and a total cost of 1,393,906 million Rials. While this scenario entails significantly greater investment and project duration, it reflects the most capacity-efficient configuration, meeting long-term service demands in regions with critical bed shortages. The comparison between the two scenarios underscores the inherent trade-offs between minimal resource allocation and maximum service capacity. In environments with constrained budgets and urgent delivery requirements, the lower-bed solution may be preferable. However, in long-term planning frameworks, the 300-bed configuration may offer strategic advantages despite the higher cost and timeline.

4.1. Validation and Limitations of the Metaheuristic Framework

To bolster the framework's real-world relevance and empirical robustness, a validation phase was conducted by benchmarking the model's predicted outputs against actual data from comparable hospital construction projects. For example, estimates for a 300-bed facility located within a specific geographic range were compared with historical records of similar-scale projects. The deviation between predicted and observed outcomes—quantified through Root Mean Square Error (RMSE) and Mean Absolute Error (MAE)—provided tangible evidence of the model's predictive reliability.

While the results generally aligned well with real-world figures, underscoring the framework's utility, several limitations remain that merit consideration. Chief among them is the assumption of static input conditions: the model presumes fixed unit costs and uninterrupted timelines. In practice, however, construction is often influenced by variables such as inflationary pressures, supply chain volatility, and geopolitical disruptions—factors that the current model does not capture.

Additionally, the feature set was intentionally streamlined to include only the number of beds and geographic coordinates. This decision, driven by data availability, excluded other potentially impactful variables such as contractor efficiency, workforce skill levels, material availability, and administrative hurdles. These latent variables, while difficult to quantify, can significantly skew timelines and budgets in actual projects.

A further methodological constraint lies in the normalization of the objective function, which assigns equal weight to time and cost. In reality, stakeholder priorities often diverge—some may prioritize rapid deployment (e.g., during public health emergencies), while others may emphasize cost containment. The uniform weighting may thus fail to reflect the nuanced trade-offs that define real-world decision-making.

These limitations inevitably temper the model's accuracy and scope of generalization. In regions marked by atypical construction norms or regulatory anomalies, for instance, the framework may produce overly optimistic forecasts. Addressing these shortcomings in future work will be crucial. Potential improvements include adopting a multi-objective optimization strategy with stakeholder-specific weighting schemes and incorporating broader contextual variables—such as regional construction indices or political risk factors—into the input structure. Such enhancements would not only refine the model's predictive capabilities but also strengthen its applicability across diverse planning environments.

4.2. Comparative Assessment and Accuracy Justification

In response to reviewer feedback calling for deeper validation and comparative scrutiny, further analysis was undertaken to evaluate the framework across both its predictive and optimization components.

From a predictive standpoint, the Random Forest (RF) model emerged as the most effective among the three tested approaches. When forecasting construction duration and cost, RF achieved an R^2 of 0.88 and 0.895, respectively—indicating strong explanatory power and reliable generalization. In contrast, Support Vector Regression (SVR) and Multilayer Perceptron (MLP) models exhibited significantly weaker performance, with SVR, notably, returning negative R^2 values in budget prediction. Such discrepancies underscore the RF model's robustness in capturing the nonlinearities and complex variable interdependencies characteristic of healthcare infrastructure data.

On the optimization front, a uniform experimental setup was applied to three metaheuristic algorithms—Grey Wolf Optimizer (GWO), Artificial Bee Colony (ABC), and Genetic Algorithm (GA)—to ensure a fair basis for comparison. Among them, GWO consistently delivered the most favorable results. It identified a configuration featuring 108 hospital beds with a composite normalized score of 0.278, outperforming ABC (0.292) and GA (0.314). This not only affirms GWO's superior convergence behavior but also highlights its adeptness in balancing competing objectives.

To assess real-world applicability, the optimization framework was tested under a practical constraint: a fixed hospital capacity of 300 beds. GWO's output under this constraint projected a cost of approximately 1.39 trillion Rials and a construction timeline of 7.97 years—both figures falling squarely within the empirical range of actual high-capacity hospital projects included in the dataset. This convergence between predicted and observed values provides compelling evidence of the model's external validity.

Collectively, these multi-layered validations—spanning algorithmic comparisons, statistical performance, and empirical cross-referencing—reinforce the framework's reliability and practical utility. By triangulating findings across both predictive modeling and optimization performance, the study presents a robust and replicable decision-support system tailored to the nuanced demands of healthcare infrastructure planning.

5. Discussion

The findings of this study underscore the potential of combining predictive models with metaheuristic optimization algorithms to enhance decision-making in the planning of hospital infrastructure projects. The use of machine learning models such as MLP, SVR, and Random Forest enabled accurate forecasting of construction time and cost, based on historical data from 270 hospital projects across Iran. Among these models, Random Forest achieved the highest predictive accuracy, highlighting its robustness and flexibility for nonlinear regression tasks with mixed-type variables, which aligns with prior studies emphasizing ensemble methods in construction analytics [9].

Furthermore, the integration of optimization algorithms—particularly Grey Wolf Optimizer (GWO), Artificial Bee Colony (ABC), and Genetic Algorithm (GA)—demonstrated varying levels of efficiency in minimizing project time and cost. GWO emerged as the most effective method, achieving the lowest normalized combined objective score, suggesting a superior balance between global search and local refinement. This observation echoes previous findings in infrastructure planning that underline the competitive convergence ability of GWO compared to traditional population-based optimizers [10].

The scenario-based analysis revealed that a 108-bed configuration at optimal geographic coordinates minimized total cost and time, whereas a fixed 300-bed scenario significantly increased both metrics yet offered greater service capacity. These results emphasize the trade-offs that decision-makers face when selecting between capacity maximization and resource efficiency. While higher capacity may be ideal in the long term for addressing national healthcare deficits, the financial and temporal feasibility must be thoroughly considered, particularly in contexts of fiscal constraints.

Additionally, the methodology proposed in this study fills an important research gap by embedding real-world geospatial and capacity variables into a dual-objective optimization process. Similar frameworks have been applied in other infrastructure sectors [11], yet applications in hospital planning remain scarce. This study contributes to the limited but growing literature on data-driven healthcare facility planning under uncertainty and offers a scalable model that policymakers can adapt to different project sizes, geographies, and budgetary limitations. Moreover, it aligns with recent calls in the literature to incorporate AI and hybrid approaches for resilient infrastructure design, particularly in high-risk or resource-limited regions [12].

6. Conclusion

This study presented an integrated data-driven framework that combines predictive analytics and optimization techniques to support informed decision-making in hospital infrastructure planning. By applying SVR for time and cost prediction and optimizing configurations using GWO, ABC, and GA, the study demonstrated that it is possible to identify efficient and feasible solutions under realistic constraints.

The scenario-based analysis showed that the GWO algorithm not only achieved superior performance but also allowed flexible exploration of trade-offs between cost, time, and capacity. The findings highlight the critical role of algorithm selection in achieving optimal outcomes, and the practical value of AI-based tools in managing complex infrastructure investments.

Future research may explore hybrid optimization approaches, incorporate more granular construction factors (e.g., contractor type, structural system), and expand the model to include environmental and social sustainability indicators. Moreover, integrating real-time data streams could further enhance model responsiveness and precision. References and Footnotes

7. References

[1] R. Zandi Doulabi and E. Asnaashari, 2016, "Identifying success factors of healthcare facility construction projects in Iran," Procedia Engineering, vol. 164, pp. 409–415, 2016. DOI: 10.1016/j.proeng.2016.11.638.

[2] R. Zandi Doulabi and E. Asnaashari, 2017 "A qualitative approach to success factors of healthcare construction projects in Iran," in Proc. 9th Int. Conf. Construction in the 21st Century (CITC-9), 2017.

[3] R. Zandi Doulabi, E. Asnaashari, D. S. Shaygan, and A. Amirkardoost, 2024, "The identification and prioritization of critical success factors in healthcare projects through the application of the Analytic Hierarchy Process (AHP)," Powertech Journal, 48 (1), 1643-1653.

[4] R. Zandi Doulabi, D. S. Shaygan, and E. Asnaashari, 2025, " The Application of System Dynamics in Claim Management of Hospital Projects: Scenario Analysis and Simulation," International Journal of Multiphysics 19 (1), 864 – 879.

[5] R. Zandi Doulabi, D. S. Shaygan, and E. Asnaashari, 2025, " Artificial Intelligence in Healthcare

Construction Scheduling: Comparative Analysis of MLP, SVR, and Decision Trees," International Journal of Multiphysics 19 (1), 858 – 863.

[6] E. Papadonikolaki, C. Perroni, and R. F. Alrashed, 2023, "Big data, data science and artificial intelligence for project management in AEC: A systematic review," MDPI Buildings, vol. 13, no. 1, p. 130, 2023. DOI: 10.3390/buildings13010130.

[7] Y. Xie, J. Zhang, C. Zhao, and J. Chen, 2023, "Machine learning-based consumption estimation of prestressed steel in bridge construction," MDPI Buildings, vol. 13, no. 5, p. 1189, 2023. DOI: 10.3390/buildings13051189.

[8] L. Zhang, B. Fan, and A. Gunasekaran, 2023, "Prediction-based decomposition optimisation for multi-portfolio supply chain resilience strategies," Omega, vol. 114, p. 102773, 2023. DOI: 10.1016/j.omega.2023.102773.

[9] E. Javanmardi and R. Noorossana, 2022, "Bridge life-cycle performance and cost analysis: prediction, optimisation and decision-making," Structure and Infrastructure Engineering, vol. 18, no. 4, pp. 605–621, 2022. DOI: 10.1080/15732479.2021.1954932.

[10] Abubakar, I. R., et al. (2020). Machine learning in construction: Current applications and future directions. Automation in Construction, 119, 103332. https://doi.org/10.1016/j.autcon.2020.103332

[11] R. Zandi Doulabi, E. Asnaashari, D. S. Shaygan, and A. Amirkardoost, 2024, " Green Hospitals: A Glance at Environmental Sustainability and Energy Efficiency in Global and Iranian Contexts," Powertech Journal, 48 (1), 1948-1967.

[12] Xu, X., et al. 2022. Multi-objective optimization for sustainable urban infrastructure planning using real-world data. Journal of Cleaner Production, 357, 131868. https://doi.org/10.1016/j.jclepro.2022.131868