Predicting construction cost index using fuzzy logic and machine learning in Jordan

Heba Al Kailani and Ghaleb J. Sweis  
Civil Engineering Department, Faculty of Engineering and Technology,  
The University of Jordan, Amman, Jordan  
Farouq Sammour  
Department of Construction Science,  
Texas A&M University College Station, College Station, Texas, USA  
Wasan Omar Maaitah  
Data Science Department, Princess Sumaya University for Technology, Amman, Jordan  
Rateb J. Sweis  
Department of Business Administration, The University of Jordan, Amman, Jordan, and  
Mohammad Alkailani  
Civil Engineering Department, Faculty of Engineering and Technology,  
The University of Jordan, Amman, Jordan

Abstract

Purpose – The process of predicting construction costs and forecasting price fluctuations is a significant and challenging undertaking for project managers. This study aims to develop a construction cost index (CCI) for Jordan’s construction industry using fuzzy analytic hierarchy process (FAHP) and predict future CCI values using traditional and machine learning (ML) techniques.

Design/methodology/approach – The most influential cost items were selected by conducting a literature review and confirmatory expert interviews. The cost items’ weights were calculated using FAHP to develop the CCI formula.

Findings – The results showed that the random forest model had the lowest mean absolute percentage error (MAPE) of 1.09%, followed by Extreme Gradient Boosting and K-nearest neighbours with MAPEs of 1.41% and 1.46%, respectively.

Originality/value – The novelty of this study lies within the use of FAHP to address the ambiguity of the impact of various cost items on CCI. The developed CCI equation and ML models are expected to significantly benefit construction managers, investors and policymakers in making informed decisions by enhancing their understanding of cost trends in the construction industry.

Keywords Construction cost index, Fuzzy logic, Machine learning in construction, Construction cost forecasting, Analytical hierarchy process, Construction management

Paper type Research paper

1. Introduction

The global construction industry, valued at over US$10.5tn, remains predominantly non-digitized in the modern era (Businesswire, 2021). The fields of engineering and technology are increasingly intertwined, driven by the rapid advancements in technology, particularly
within the construction sector (Hernandez-de-Menendez et al., 2020). In recent years, artificial intelligence (AI) has gained prominence in construction engineering and management due to its potential to enhance performance and efficiency (Saka et al., 2023).

The construction industry serves as a cornerstone of economies worldwide, encompassing residential, commercial, industrial and educational building projects, along with critical infrastructure such as roads, bridges, ports and airports (Sepasgozar et al., 2022). Beyond construction, the industry plays a pivotal role in maintaining and repairing existing structures (Gu and Wang, 2022). Apart from providing employment opportunities, it significantly contributes to economic development and growth (Ball, 2014), addressing complex societal, environmental and energy-related challenges. Its interconnections with other sectors further amplify its impact on the economy (Jordan Strategy Forum, 2019). In this global competitive environment marked by narrowing profit margins and declining market shares, accurate cost estimation stands as a paramount concern in construction decision-making (Matel et al., 2019). Jordan’s construction sector is a vital economic pillar, contributing 5% to the gross domestic product in 2022, ranking sixth among all sectors. In terms of employment, 5% of Jordanians were employed in the construction industry in 2022 (DoS, 2023).

However, the construction industry is notorious for its inherent uncertainty and complexity, rendering cost estimation and prediction a formidable challenge (Shehadeh et al., 2022a). In recent years, there has been growing interest in leveraging fuzzy logic and machine learning (ML) algorithms to enhance cost estimation and prediction within the construction domain (Elmousalami, 2019).

Fuzzy analytic hierarchy process (FAHP) is a decision-making methodology that combines analytic hierarchy process (AHP) principles with fuzzy logic to handle uncertain and imprecise data (Chan et al., 2019). It extends the AHP approach by accommodating linguistic variables to represent subjective evaluations and uncertainties in decision-making (Kutlu Gündoğdu et al., 2021).

Integral to the project life cycle is the cost and bid estimating process, requiring in-depth investigation, extensive knowledge, substantial experience and continuous development for precise cost estimation (Hatamleh et al., 2018). The construction cost index (CCI) plays a pivotal role in achieving more accurate bids by capturing price fluctuations over the short and long term (Zhang et al., 2018). Owners use this index to forecast project costs, whereas contractors rely on it to provide financial proposals during the tendering process (Xu and Moon, 2011).

The CCI has multifaceted implications, including its ability to guide construction managers in cost forecasting and budgeting, assisting investors in risk assessment, supporting policymakers in formulating growth-focused policies and enhancing the educational sector by aiding the development of relevant teaching materials and courses. In addition, the CCI promotes public awareness of construction cost drivers, improves transparency in the construction industry, facilitates the development of affordable housing and ultimately enhances the quality of infrastructure and public services.

This study introduces an innovative approach to develop a CCI, integrating AHP and fuzzy logic, both recognized as potent decision-making tools. AHP aids in ranking construction cost items based on their relative importance, whereas fuzzy logic manages the ambiguous and imprecise information typically encountered in the construction industry.

Importantly, this research focuses on the Jordanian context. Jordan’s construction industry, as a significant economic sector, poses unique challenges and opportunities. By conducting this study in Jordan, we aimed to address specific issues and contribute to the body of knowledge in a context that has received relatively less attention in the literature.
This study bridges the gap in cost estimation and prediction in the construction industry by integrating advanced methodologies within the specific context of Jordan, addressing a critical need for more accurate and informed decision-making in this vital sector.

The research aims are as follows:

- Identify the major cost drivers for estimating CCI in the Jordanian construction sector.
- Evaluate the significance of factors contributing to cost and their impact on CCI through the application of fuzzy AHP.
- Determine the ML model that achieves the highest level of accuracy.
- Predict future CCI values using ML models.
- Provide insights into improving predictive models in future research and development.

2. Literature review

Project success in the construction industry hinges on several factors, with cost being a crucial determinant. Radjuković and Sjekavica (2017) emphasize that effective project management involves balancing time, cost and quality. In particular, the cost of construction is widely considered the most critical determinant of project success. Therefore, it is essential to measure any deviation from the estimated cost to assess its impact on project performance and profitability, as highlighted by Vaardini (2016).

In the context of construction projects, the accuracy of cost estimation plays a pivotal role in ensuring project success (Garg and Misra, 2021). This significance arises from the pervasive problem of cost overruns, which are especially problematic given the industry’s current focus on tight budgets, as noted by Zhu et al. (2010). Large-scale construction projects, in particular, are susceptible to rework, which can result in substantial cost overruns and significant schedule delays, as reported by Safapour and Kermanshachi (2019). These findings underscore the critical importance of accurately identifying the major cost drivers for estimating CCI in the Jordanian construction sector.

The process of accurately estimating building construction costs during the design phase has posed challenges for Jordanian designers and their clients. The precision of expense projections for building construction significantly impacts the success or failure of a given project (Dandan et al., 2020).

Fuzzy logic and AHP are two significant techniques in the field of decision-making under uncertainty (Bouamrane et al., 2020). Fuzzy logic is a mathematical approach that enables reasoning with uncertain or ambiguous data, by assigning degrees of membership to concepts or variables rather than binary values (Rodriguez-Valderrama et al., 2022). FAHP, on the other hand, is a method that extends the traditional AHP model to handle imprecise or vague information by using fuzzy sets and membership functions (Liu et al., 2020). This approach allows decision makers to more effectively handle subjective criteria and preferences more in the decision-making process (Shi and Lai, 2023). Combining these two techniques, decision-makers can use fuzzy logic to represent and manipulate uncertain or vague data, and then use fuzzy AHP to prioritize and rank multiple decision criteria based on their relative importance (Saivaew and Butdee, 2020). The result is a more robust and comprehensive decision-making process that is better suited to handling complex real-world problems (Figueiredo et al., 2021).

The effective execution of projects in construction engineering and management remains crucial, prompting ongoing efforts among researchers and practitioners to enhance construction methodologies. Fuzzy logic stands as a significant component within various applications of construction engineering and management (Fayek, 2020). Nguyen and Nguyen (2020) proposed
a fuzzy decision-making method to confront the complexity of rapidly estimating costs in construction investment endeavours, using a construction pricing index. This challenge stems from the fluctuations in economic and social factors that influence the construction industry’s framework.

Over the years, there has been significant scholarly interest in the integration of AI tools within the construction sector to enhance process efficiency and facilitate informed decision-making (Akinosho et al., 2020; Dang-Trinh et al., 2023). This is particularly relevant in the context of resource planning, risk management and logistical complexities, which are known to frequently contribute to issues such as design flaws, delays in project delivery, budgetary overruns and contractual disputes (Sammour et al., 2023; Sánchez-Garrido et al., 2023).

In McKinsey (2020) report titled “The Next Normal in Construction”, it is identified that AI can add value throughout the project lifecycle, including design, financing, construction, operations and asset management, as well as transforming business models. The execution of AI in construction has the potential to overcome obstacles such as shortages of workers, safety concerns and cost and schedule overruns, ultimately benefiting the industry as a whole (Rao, 2022).

Construction cost predictions to reduce time risk assessment are indispensable steps for the process of decision-making of managers. ML techniques need adequate data set size to model and forecast the cost of projects (Tayefeh Hashemi et al., 2020). Shoar et al. (2022) presents an innovative application for ML in the construction industry, emphasizing how to transform cost prediction techniques so that they include engineering services in addition to customary building costs. Besides advancing the field of building cost estimates, the Shoar et al. (2022) study emphasizes the wider ramifications of using ML to improve project management in the sector.

Forecasts of construction costs play a pivotal role in mitigating time-related risks and informing managerial decision-making processes. Successful application of ML techniques in cost prediction necessitates sufficiently large datasets (Tayefeh Hashemi et al., 2020). Shoar et al. (2022) introduced a novel utilization of ML within the construction industry, highlighting the expansion of cost prediction methodologies to encompass engineering services alongside conventional building expenses. Beyond enhancing the realm of building cost estimations, the study by Shoar et al. (2022) underscored the broader implications of using ML to enhance project management within the sector.

The accurate forecasting of the future trends of the CCI has been regarded as a critical component, given the significance of budget planning and contract bidding in construction cost management. However, it should be emphasized that construction costs are subject to significant variance due to regional differences in market conditions and environmental factors (Choi et al., 2021). In this context, ML models offer a powerful tool for predicting CCI trends. These models can analyse historical data and identify patterns, enabling them to make predictions about future trends (Zhang et al., 2021).

Cao and Ashuri (2020) introduced an innovative AI algorithm for cost index forecasting, demonstrating superior predictive accuracy compared to existing time series models, especially for volatile cost indexes. Similarly, Alshboul et al. (2022b) focused on financial and social development in US highway construction projects, addressing delays leading to liquidated damages (LDs). Their study developed an ensemble machine learning technique (EMLT) combining various algorithms to predict LDs, identifying key attributes and assessing correlations among influential features for accurate forecast models. Different ML-based models were created and evaluated using performance metrics like root mean square error (RMSE), MAE, mean absolute percentage error (MAPE) and $R^2$, with the EMLT model showing the highest accuracy of 0.997 compared to other ML models. The research underscores the efficacy of the EMLT model as an administrative decision-making tool for
forecasting LDs, highlighting ML's potential in advancing computational aspects within construction projects.

The comprehensive analysis conducted by Baduge et al. (2022) examined the application of AI and ML techniques throughout the building and construction sectors, encompassing the entire building lifecycle. The review encompassed diverse areas such as material design, construction management, progress monitoring, building management and sustainability. Furthermore, Baduge et al. (2022) highlighted the prevalent utilization of ML algorithms within the construction industry, particularly in predicting the characteristics of construction materials and facilitating the development of cost-effective and sustainable materials. This is attributed to the increased processing capacity and the abundance of data generated, which enable more robust outcomes (Uddin et al., 2022).

Moreover, Elmousalami (2020) study comprehensively investigates cost drivers using qualitative and quantitative procedures while exploring computational intelligence (CI) techniques for crafting reliable cost prediction models. It extensively reviewed prevalent AI techniques applicable to cost modelling, including fuzzy logic, artificial neural networks, regression models, case-based reasoning, decision trees and others. The research highlighted the significance of these techniques in developing parametric cost models during the conceptual phase of projects. Among the 20 AI techniques evaluated, Extreme Gradient Boosting (XGBoost) emerges as the most accurate method, achieving 9.091% MAPE and 0.929 adjusted R-squared. The study emphasized the non-linear adaptability, handling of missing values and outliers, model interpretation and considerations of uncertainty across these AI models.

3. Methods

The study introduces a novel method by combining FAHP with ML techniques to forecast the CCI within Jordan's construction industry. This innovative approach stands out for its capability to navigate the intricacies and uncertainties inherent in predicting construction costs within a region known for its distinctive market dynamics. The choice of FAHP stems from its adeptness in handling imprecise and uncertain data commonly found in construction cost estimations in Jordan. FAHP plays a pivotal role in fostering a deeper comprehension of the varying significance of different cost influencers, taking into account subjective judgements and data variability. This method thus enables a more refined and comprehensive analysis of the factors impacting construction costs. To fulfil the objectives of this study, it is necessary to develop a formula for calculating the CCI in Jordan. This formula will use Fuzzy AHP to determine the weight of each cost driver. The research approach in this study is mix design, using both quantitative and qualitative data to achieve its objectives. The study uses a mixed-methods research strategy, combining data collection techniques, quantitative analysis (fuzzy AHP, regression, autoregressive integrated moving average [ARIMA] and ML algorithms) and qualitative analysis (interviews) to address the research objectives. The methodology flowchart is illustrated in Figure 1.

3.1 Data collection

The data collection combined a comprehensive review of academic and industry literature (Taylor, 2016) with the collection of historical price data from credible sources, ensuring a robust and comprehensive cost catalogue. The choice of monthly data collection over a span of 12 years provides a detailed and comprehensive view of the market trends, essential for accurate forecasting.

To collect quantitative data for this study, historical prices for the identified cost items were collected from multiple sources, including the Department of Statistics, The Jordanian
Figure 1. Methodology flowchart

Source: Authors’ own creation
Engineers Association, the Jordanian Construction Contractor Association and the General Tender Department. The sources of data were selected based on their credibility and authority in the area of construction cost analysis. Data were collected at monthly intervals spanning from January 2010 through December 2022, yielding a total of 156 discrete data points.

The qualitative data collection process via interviews was meticulously planned, following Maxwell’s (2018) systematic approach. The decision to limit the interviews to five experts was grounded in Creswell (2019) principles, prioritizing depth and quality over quantity. This strategy aimed to extract detailed and concentrated insights from each expert, fostering more precise and meaningful data analysis. After identifying potential interviewees, invitations to participate in the study were extended via email or phone. Table 1 presents the list of study participants, outlining their professional backgrounds, years of experience and the duration of the interviews. A flexible set of open-ended interview questions was crafted to elicit responses that could shed light on the most significant cost components in construction projects within Jordan. These questions provided interviewees with the opportunity to articulate their experiences and perspectives in considerable detail.

The data collection process involved recording the interviews and taking notes. Following data collection, a qualitative data analysis technique was used to select the most essential cost items in construction projects in Jordanian construction based on the responses obtained. The analysis involved a ranking scheme and a pair-wise comparison matrix to identify frequently mentioned and highly ranked cost items.

3.2 Fuzzy analytic hierarchy process analysis

The FAHP analysis was chosen due to its ability to translate subjective expert judgements into quantifiable measures, a method extensively supported in literature for its efficacy and reliability (Kubler et al., 2016; Noor et al., 2017; Vinogradova-Zinkevič et al., 2021). The use of FAHP in the context of Jordan’s construction industry is a novel application, providing a more localized and relevant analysis of cost drivers.

3.2.1 Analytic hierarchy process. A hierarchical model of the decision problem was constructed (Figure 2), with the goal positioned at the highest level, criteria in the middle and alternatives at the lowest level.

<table>
<thead>
<tr>
<th>Interviewee</th>
<th>Position</th>
<th>Years of experience</th>
<th>Education level</th>
<th>Relevant experience in Jordanian construction</th>
<th>Duration of interview</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Project Manager</td>
<td>15</td>
<td>Master’s degree</td>
<td>Extensive experience in urban development</td>
<td>20 min</td>
</tr>
<tr>
<td>2</td>
<td>Site Supervisor</td>
<td>11</td>
<td>Bachelor’s degree</td>
<td>Supervised large-scale infrastructure projects</td>
<td>15 min</td>
</tr>
<tr>
<td>3</td>
<td>Construction Worker</td>
<td>5</td>
<td>Vocational training</td>
<td>Hands-on experience with local labor practices</td>
<td>10 min</td>
</tr>
<tr>
<td>4</td>
<td>Architect</td>
<td>24</td>
<td>Master’s degree</td>
<td>Designed award-winning buildings in Jordan</td>
<td>15 min</td>
</tr>
<tr>
<td>5</td>
<td>Design Engineer</td>
<td>6</td>
<td>Bachelor’s degree</td>
<td>Involved in national infrastructure planning</td>
<td>10 min</td>
</tr>
</tbody>
</table>

Source: Authors’ own creation

Table 1. List of interviewees
The next step involved determining the relative importance of the criteria by conducting pairwise comparisons using Saaty’s (2008) 1–9 scale, and using the fundamental AHP matrix to obtain the criteria weights. Then, the alternatives can be evaluated against each criterion using the same pairwise comparison method, and the alternative weights can be determined by implementing the criteria weights to the alternative evaluations.

A consistency metric was used to evaluate the level of consistency attained. The consistency index was then calculated by quantifying the variance between the maximum eigenvalue and the number of criteria, divided by the number of criteria minus one. Equation (1) outlines the specific formula used for computing the consistency index:

$$CI = \frac{\lambda_{max} - n}{n - 1}$$  \hspace{1cm} (1)

where $CI$ is the consistency index, $\lambda_{max}$ is the maximum eigenvalue and $n$ is the number of criteria or alternatives.

The consistency ratio is then calculated by dividing the consistency index by the random index (Sahoo et al., 2016), as shown in equation (2). The random index is a measure of the expected consistency for a given number of criteria or alternatives:

$$CR = \frac{CI}{RI}$$  \hspace{1cm} (2)

The judgements are regarded as consistent and can be used for decision-making if the consistency ratio is less than or equal to 0.1.

3.2.2 Fuzzy logic. The methodology used in this study involved the derivation of fuzzy and normalized weights for a specified set of criteria. To achieve this, the approach
implemented the fuzzification technique, using a triangle membership function, in conjunction with the fuzzy geometric mean value (Thaker and Nagori, 2018).

To implement fuzzification in the developed pairwise comparison matrix, the triangle membership function was used in the first instance. This function comprises three distinct parameters, namely, a lower bound (a), a mode (b) and an upper bound (c). Notably, a triangular fuzzy number, which is used to visually represent the fuzzy set, is represented by a triangle with its peak at the mode and its base defined by the lower and upper bounds (Krejčí, 2018).

The triangular fuzzy number scale represents values or variables via triangular fuzzy numbers, as illustrated in Table 2. This scale associates each value or variable with a triangular fuzzy number that reflects the degree of membership or preference of the corresponding value or variable within the relevant fuzzy set.

Finally, the fuzzy geometric mean value was ultimately used to identify the fuzzy weights of the criteria, whereby the $n$th root of the product of each criterion’s degree of membership with itself was taken (Wang, 2015). Following this, the fuzzy weights were subjected to normalization to ensure that their sum equated to 1. To convert the fuzzy weights to weights, the de-fuzzification method was applied, using the centroid method to calculate the weights of each criterion. The main objective of the fuzzy set was computed to determine the weight of each criterion via the aforementioned method (Helmy et al., 2021).

3.3 Developing and calculating the construction cost index

The key cost items previously identified will be used in index calculation, with a base period selected to serve as the point of comparison for all subsequent periods. The base period, selected for its relatively stable prices, is January 2010.

The weight of each cost item, as determined by the FAHP, was used in computing the coefficients for the CCI formula. The equations used for this computation are as follows in Table 3.

The formula for the CCI represents a weighted average of the prices of the main cost items, with the respective weights reflecting their significance in construction. The equation (3) to be used in this study is presented as follows:

$$CCI = S_{wi} \times SU_i + CO_{wi} \times COU_i + CE_{wi} \times CEU_i + A_{wi} \times AU_i + D_{wi} \times DU_i$$ (3)

where:

$SU_i = $ Steel unit price for period $i$;

<table>
<thead>
<tr>
<th>Definition of linguistic terms</th>
<th>Triangular fuzzy no. scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equal (EQ)</td>
<td>(1, 1, 1)</td>
</tr>
<tr>
<td>Moderate (MD)</td>
<td>(2, 3, 4)</td>
</tr>
<tr>
<td>Strong (ST)</td>
<td>(4, 5, 6)</td>
</tr>
<tr>
<td>Very strong (VS)</td>
<td>(6, 7, 8)</td>
</tr>
<tr>
<td>Extremely strong (ES)</td>
<td>(9, 9, 9)</td>
</tr>
<tr>
<td>Intermediate values</td>
<td>(1, 2, 3)</td>
</tr>
<tr>
<td></td>
<td>(3, 4, 5)</td>
</tr>
<tr>
<td></td>
<td>(5, 6, 7)</td>
</tr>
<tr>
<td></td>
<td>(7, 8, 9)</td>
</tr>
</tbody>
</table>

*Source: Authors’ own creation*
3.4 Forecasting models

The study uses a mix of traditional techniques (linear regression, ARIMA) and advanced ML techniques (KNN, RF, XGBoost). Linear regression and ARIMA are well-established methods in time series forecasting, providing a baseline for comparison. The inclusion of ML techniques, chosen for their scalability, accuracy and versatility, introduces a novel element to the study, enabling the handling of more complex, non-linear patterns in the data.

3.4.1 Traditional techniques. The train data and test data were divided in a ratio of 80:20 to achieve accurate predictions with a limited number of data points. Linear regression and ARIMA were applied as univariate models to forecast the CCI in this study.

A common method for modeling the relationship between scalar variables in time series forecasting is linear regression (Dudek, 2016). This method assumes that there is a linear relationship between the dependent variable (CCI) and one or more independent variables (time), which allows for the prediction of future values based on historical data (Altay, 2005).

The ARIMA model is a frequently used technique for forecasting future values based on historical data in time series analysis (Ho and Xie, 1998; Siami-Namini et al., 2019). The first step in the ARIMA modeling process was to identify the model’s parameters, which consist of the order of differencing \(d\), the order of autoregression \(p\) and the order of moving average \(q\) (Newbold, 1983).

3.4.2 Machine learning techniques. The selection criteria incorporated considerations such as scalability, accuracy, speed, versatility and regularization capabilities for mitigating overfitting. Based on these criteria, four distinct ML algorithms, K-nearest neighbours (KNN), random forest (RF) and XGBoost were developed and implemented on Google Collab.

<table>
<thead>
<tr>
<th>CCI component</th>
<th>Formula</th>
<th>Variables description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steel coefficient</td>
<td>(S_w = \frac{W_s \times 100}{SU_b})</td>
<td>(S_w) Steel coefficient (W_s) Steel normalized weight percentage (SU_b) Steel unit price at the base year in JOD/ton.</td>
</tr>
<tr>
<td>Concrete coefficient</td>
<td>(CO_w = \frac{W_{CO} \times 100}{COU_b})</td>
<td>(CO_w) Concrete coefficient (W_{CO}) Concrete normalized weight percentage (COU_b) Concrete unit price at the base year in JOD/m³</td>
</tr>
<tr>
<td>Cement coefficient</td>
<td>(CE_w = \frac{W_{CE} \times 100}{CEU_b})</td>
<td>(CE_w) Cement coefficient (W_{CE}) Cement normalized weight percentage (CEU_b) Cement unit price at the base year in JOD/ton</td>
</tr>
<tr>
<td>Aggregate coefficient</td>
<td>(A_w = \frac{W_A \times 100}{AU_b})</td>
<td>(A_w) Aggregate coefficient (W_A) Aggregate normalized weight percentage (AU_b) Aggregate unit price at the base year in JOD/m³</td>
</tr>
<tr>
<td>Diesel coefficient</td>
<td>(D_w = \frac{W_D \times 100}{DU_b})</td>
<td>(D_w) Diesel coefficient (W_D) Diesel normalized weight percentage (DU_b) Diesel unit price at the base year in JOD/Liter</td>
</tr>
</tbody>
</table>

Source: Authors’ own creation
software for forecasting the CCI in Jordan. The optimal performance of each algorithm was attained via hyperparameter tuning, ensuring the attainment of optimal forecasting outcomes for the CCI.

3.4.3 Models evaluation. The selection of MAPE, RMSE and $R^2$ as performance metrics is based on their widespread acceptance and reliability in model evaluation literature (Alshboul et al., 2022a, Chai and Draxler, 2014; Hyndman and Koehler, 2006), ensuring a comprehensive assessment of the models’ predictive capabilities.

4. Data analysis, results and discussion

4.1 Fuzzy analytic hierarchy process analysis

The FAHP analysis introduced a novel approach to understanding the relative importance of construction cost components in Jordan. The pairwise comparison matrix presented in Table 4 was gathered from the interviews. The values represented the average relative importance of each material compared to every other material, as assessed by the experts. This provided an estimate of the relative significance of each material based on the collective judgements of the experts. These values were expressed as ratios, where a value of 1 indicated equal importance between two materials. Higher values indicated greater importance, whereas lower values indicated lesser importance.

The pairwise comparison matrices generated were then tested for consistency using the consistency ratio. The consistency ratio was found to be 0.085 which is less than 0.1, indicating that the judgements provided by the experts were consistent enough to be used for decision-making.

Following the generation of the pairwise matrix, a fuzzification process was applied using the triangular membership function. This function, distinguished by its triangular shape, allows for the assignment of diverse values to different levels of membership. Moreover, it facilitates a smooth and gradual transition between varying degrees of membership, thereby enhancing the overall understanding of the decision-making process.

Afterward, the fuzzy geometric mean value was computed to aggregate the fuzzy numbers and calculate the degree of importance of each criterion. This was followed by determining the fuzzy weights of each criterion to represent their degree of significance relative to the others. The fuzzy weights were calculated by dividing the fuzzy geometric mean of each criterion by the sum of all the fuzzy geometric means.

Finally, the normal weights are calculated by de-fuzzifying the fuzzy weights using the centre of gravity method. The centre of gravity method calculates the mean of the triangular membership function for each criterion, weighted by the corresponding fuzzy weight. The resulting values are then normalized to obtain the final normal weights shown in Figure 3.

These results suggest that concrete, steel, diesel and cement should receive greater attention in cost management strategies for the construction of concrete structures in Jordan. In addition, these findings may have broader implications for the construction industry, as

<table>
<thead>
<tr>
<th>Cost item</th>
<th>Concrete</th>
<th>Cement</th>
<th>Steel</th>
<th>Aggregate</th>
<th>Diesel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concrete</td>
<td>1</td>
<td>8</td>
<td>3</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Cement</td>
<td>1/8</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Steel</td>
<td>1/3</td>
<td>1/6</td>
<td>1</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Aggregate</td>
<td>1/7</td>
<td>1/7</td>
<td>1</td>
<td>1/5</td>
<td>1</td>
</tr>
<tr>
<td>Diesel</td>
<td>1/5</td>
<td>2</td>
<td>1/6</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

Source: Authors’ own creation

Table 4. Pairwise comparison matrix
they highlight the relative importance of different cost items and can inform decision-making processes related to cost optimization and resource allocation.

The FAHP analysis revealed the relative importance of various construction cost components in Jordan. When compared to existing literature, such as studies conducted in different geographical contexts, the findings align with the general trend where concrete and steel are often major cost drivers (Ahmed and Arocho, 2021; Balasbaneh and Ramli, 2020; Ji et al., 2021; Val and Stewart, 2003). However, the significant weight assigned to diesel in our study is unique to Jordan’s context, likely due to the local economic and logistical factors influencing construction costs (Alshboul et al., 2021).

4.2 Construction cost index formula and time series plot

Subsequently, the weights that have been normalized through the implementation of the FAHP are used in the computation of the CCI coefficients for distinct cost elements. Equation (4) presents the CCI formula that can be used for any given period of consideration:

\[
CCI = 0.0701 \times SU_i + 0.799 \times COU_i + 0.0706 \times CEU_i + 0.506 \times AU_i + 23.0 \times DU_i
\]  

(4)

where \(SU_i\) is the steel unit price for period \(i\), \(COU_i\) is the concrete unit price for period \(i\), \(CEU_i\) is the cement unit price for period \(i\), \(AU_i\) is the aggregate unit price for period \(i\) and \(DU_i\) is the diesel unit price for period \(i\).

The time series data illustrated in Figure 4 starts from January 2010 and continues until December 2022.

The pattern of increasing construction costs over time, as observed in our time series analysis (Figure 4), aligns with global trends in the construction industry. However, the
distinctive seasonal pattern, with peaks in summer and lows in winter, highlights the unique market dynamics in Jordan, potentially driven by climatic and economic cycles specific to the region (Shehadeh et al., 2022b).

4.3 Models’ development

4.3.1 Linear regression. The CCI forecasting model was fine-tuned by optimizing its parameters through the utilization of the linear regression method, aiming to achieve the highest performance as measured by MAPE. The resultant optimized linear regression equation [equation (5)] serves as a valuable tool for estimating the CCI:

\[
y = 0.0019x + 25
\]  

(5)

It is noteworthy that in Excel, dates are represented as numeric values where the integer part denotes the number of days since 1 January 1900. Furthermore, the coefficient of the independent variable, 0.0019, indicates the extent to which a unit change in the independent variable affects the value of the dependent variable. There is a positive correlation between the CCI and time, as it is expected to grow over time due to natural inflation and interest rates. Inflation and interest rates affected the prices of raw materials, labour and other construction-related expenses, which in turn impacted the CCI.

4.3.2 Autoregressive integrated moving average. Various ARIMA models were constructed to minimize the MAPE, and the most effective model was found to have \( p = 3, \ d = 0 \) and \( q = 3 \) with a square root transformation. The MAPE value calculated for this model was 6.42%, indicating that the forecasted values deviated from the actual values by an average of 6.42% of the actual values.

The ARIMA model parameters were designed to address autocorrelation and serial correlation in the time series data. The value of \( p = 3 \) suggests that the model incorporates three lagged values of the dependent variable, whereas \( d = 0 \) demonstrates that the data are steady and do not need to be differentiable. The value of \( q = 3 \) indicates that the model accounts for three lagged values of the error term to address serial correlation in the errors.

In addition, the model applied a square root transformation to stabilize the variance of the data. Overall, this model is designed to accurately capture the time series patterns and minimize the deviation between the forecasted and actual values.
4.3.3 Machine learning models. Firstly, the non-parametric KNN algorithm was used in this study to perform regression and predict the values of CCI. Secondly, leveraging the interrelationships among different factors influencing CCI, the XGBoost algorithm was used to forecast its values. Thirdly, an ensemble ML approach known as RF was applied.

It is important to note that all three ML models underwent an auto-tuning process to optimize their hyperparameters and achieve the highest predictive accuracy, as determined by the MAPE test.

4.4 Models’ performance evaluation

To provide a comprehensive summary and facilitate a comparative analysis of the predictive capabilities among the five different forecasting models, Table 5 presents the calculated values of the relative accuracy measures for each model.

The findings suggest that the RF, XGBoost and KNN models performed the best in predicting the CCI, as they achieved the lowest MAPE and RMSE values and the highest $R^2$ value. The RF model achieved the lowest MAPE and RMSE values, indicating a high level of accuracy and precision. The linear and ARIMA models, on the other hand, performed the poorest, achieving higher MAPE and RMSE values and lower $R^2$ values. Overall, the results suggest that the ML models, XGBoost, KNN and RF outperformed the traditional statistical models, linear and ARIMA, in predicting the CCI.

The ML models’ performance, particularly the RF model’s high accuracy, offers a novel perspective on CCI prediction in Jordan. While ML techniques have been widely used in construction cost forecasting globally, their application in Jordan’s market, characterized by

<table>
<thead>
<tr>
<th>Model</th>
<th>MAPE</th>
<th>RMSE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>6.29%</td>
<td>7.295</td>
<td>0.892</td>
</tr>
<tr>
<td>ARIMA</td>
<td>6.42%</td>
<td>7.612</td>
<td>0.904</td>
</tr>
<tr>
<td>KNN</td>
<td>1.46%</td>
<td>1.973</td>
<td>0.907</td>
</tr>
<tr>
<td>XGBoost</td>
<td>1.41%</td>
<td>1.804</td>
<td>0.922</td>
</tr>
<tr>
<td>RF</td>
<td>1.09%</td>
<td>1.534</td>
<td>0.944</td>
</tr>
</tbody>
</table>

Table 5. Models’ performance

*Source: Authors’ own creation*
its specific challenges, provides unique insights. For instance, the RF model’s ability to capture non-linear trends in the data reflects the complex interplay of economic, material and labour factors in the Jordanian construction industry.

4.5 Forecasts
Future CCI for the period of January 2023 to December 2023 was predicted using the top three performing models. The outputs of the models are displayed in Figure 5. The RF, KNN and XGBoost algorithms predicted several future events with comparable trends and seasonality. The level of anticipated demand, however, varies between the three algorithms. All the algorithms predicted a rise in demand in November 2023, with CCI values for XGBoost, KNN and RF, respectively, of 127.13, 124.15 and 117.42.

The forecasted rise in CCI in November 2023 (Figure 5) aligns with global economic trends of increasing construction costs but also reflects specific regional factors such as economic policies and material supply chains. The comparative analysis of different ML models showcases their varied predictive capabilities, with each model capturing different aspects of the market dynamics in Jordan.

According to the RF algorithm, the CCI of Jordan is estimated to reach 117.79 in May 2023, indicating a 5.04% rise compared to the same month of 2022. The greatest demand is anticipated to occur in November 2023, with a value of 127.13, whereas the lowest level is expected to be recorded in March 2023, with a value of 106.03. The average CCI during the year 2022 was found to be 113.12, whereas the index is forecasted to increase by 2.41% and reach 115.85 during the next year (2023), as shown in Figure 6.

Similarly, the XGBoost algorithm suggests that Jordan’s CCI would reach 113.86 in May 2023, indicating a 1.54% increase from May 2022. The highest and lowest levels of demand are anticipated to occur in November 2023 and April 2023, with values of 124.15 and 101.91, respectively. The average CCI during the year 2022 was found to be 113.12, whereas the index is predicted to increase by 0.81% and reach 112.22 during 2023, as shown in Figure 7.

Moreover, the KNN algorithm suggests that Jordan’s CCI would decrease by 3.35% and reach 108.37 in May 2023 compared to May 2022. The greatest level of construction cost is expected in November 2022, with a value of 117.42, whereas the lowest amount of demand is expected in April 2023, with a value of 96.90. The average CCI during the year 2022 was found to be 113.12, whereas the index is predicted to decrease by 5.55% and reach 106.84 during 2023 compared to 2022, as shown in Figure 8.

Source: Authors’ own creation

Figure 6. Random forest model forecast
5. Conclusion
This study aimed to devise a CCI equation for Jordan’s construction sector by using fuzzy AHP and conducting a comprehensive literature review. Furthermore, the study aimed to build and assess models that incorporate traditional and ML algorithms for predicting forthcoming CCI values. In addition, the study endeavoured to put forward recommendations for future research to enhance the precision of forecasting models.

A systematic method was used to identify and select the most significant cost items that affect the CCI. This method aimed to establish a comprehensive cost catalogue. The study conducted a meticulous review of academic literature, conference proceedings and industry reports. The selection process was carefully carried out to ensure the relevance of the chosen cost items to the Jordanian construction market.

Data was gathered from credible sources to provide a historical perspective on the selected cost items. Subsequently, to determine which cost items have the highest impact on the CCI, the study conducted interviews with industry experts who were selected based on their experience and knowledge in the area of construction. The selection criteria included evaluating the impact of the cost items on overall construction project costs, as well as their

**Figure 7.**
XGBoost model forecast

**Source:** Authors’ own creation

**Figure 8.**
KNN model forecast

**Source:** Authors’ own creation
availability and practicality across different regions. After careful investigation, five items were identified as the most influential, namely concrete, cement, steel, aggregate and diesel.

Accordingly, the expert’s collective judgements were used to develop a pairwise comparison matrix to estimate the relative importance of each material. The values were expressed as ratios. Subsequently, the pairwise matrix underwent fuzzification using the triangular membership function, which allowed assigning of varying values to distinct degrees of membership. Then fuzzy weights for each criterion were calculated. The normal weights were then obtained by de-fuzzifying the fuzzy weights using the centre of gravity method.

Concrete was found to be the most significant cost item in the construction projects in Jordan, followed by steel, diesel, cement and aggregate. Consequently, the calculated weights from the fuzzy AHP method were applied to construct the CCI equation for calculating the CCI time series.

The CCI values were forecasted using various models including traditional ARIMA and linear models, and ML algorithms such as RF, XGBoost and KNN. This study used three effective measures, namely, MAPE, RMSE and $R^2$, to evaluate the accuracy of forecasted data. These metrics were computed on testing data sets to assess the developed model’s predictive accuracy.

The results indicate that the RF, XGBoost and KNN models exhibited the most favourable performance in forecasting the CCI, demonstrating lower MAPE and RMSE values and higher $R^2$ values. Specifically, the RF model yielded a MAPE of 1.09%, an RMSE of 1.53% and an $R^2$ of 0.94%, which implies a higher level of precision and accuracy. The findings suggest that the ML algorithms, including RF, XGBoost and KNN, surpassed the traditional statistical models of linear and ARIMA in predicting the CCI.

In addition, the three most effective models were used to predict future CCI values from January to December 2023. The XGBoost, KNN and RF algorithms provided diverse scenarios for the upcoming period with comparable patterns and seasonal trends. However, there were variations in the predicted levels of demand among the three models. Notably, all models projected a spike in demand for November 2023, with the CCI values predicted as 127.13, 124.15 and 117.42 for XGBoost, KNN and RF, respectively.

The research conducted has significant implications for the field of construction management and the development of forecasting models for construction cost trends. The utilization of the FAHP to address the inherent ambiguity in determining the impact of various cost items on CCI represents a novel contribution to the field. These models provide construction managers, investors and policymakers with a reliable means of forecasting construction cost trends, enabling them to make informed decisions and optimize project outcomes. Overall, this study highlights the potential benefits of integrating advanced techniques of FAHP and ML in the construction industry, paving the way for further advancements in the field.

The study encountered several limitations, primarily stemming from the restricted accessibility of data of superior quality, a challenge exacerbated by governmental regulations governing data access. Furthermore, ML models are susceptible to the issues of overfitting and underfitting, thus posing a threat to the dependability and soundness of the predictions. Another limitation arises from the restricted scope of analysis, as the study exclusively concentrated on particular construction cost factors within the context of Jordan.

Future research could build on this foundation by exploring additional variables that could impact construction costs and developing more robust forecasting models. In addition, opportunities in the domain of construction cost forecasting include the exploration of more refined data frequencies, such as weekly or daily data, to improve the accuracy and responsiveness of predictive models. Integrating the forecasting models with construction
management software has the potential to optimize project schedules, asset management and materials procurement, resulting in more cost-effective and timely project completion. This integration encourages data-driven decision-making and strengthens competitiveness in the industry.

Future studies are needed to externally validate this researcher’s findings applicability and relevance beyond the Jordanian context. Investigating how the methodologies and models developed in this study can be adapted or extended to other regions or countries would not only enhance its global appeal but also contribute to the field of CCI prediction on a larger scale.

References


Corresponding author
Heba Al Kailani can be contacted at: hebakailani98@gmail.com

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com