Towards Developing a Predictive Model for Interpersonal Communication Quality in Construction Projects: An Ensemble Artificial Intelligence-based Approach

Ali Rahimian; Keivan Sadeghzadeh*; Saeed Reza Mohandes; Igor Martek; Patrick Manu; Maxwell Fordjour Antwi-Afari; Sajjad Mirvalad; Ibrahim Odeh

*Corresponding author: k.sadeghzadeh@northeastern.edu

Abstract

High-quality communication among stakeholders is salient to securing and maintaining collaboration in construction projects. Indeed, the absence of such communication among the workforce leads to inefficiency, low productivity, and substandard deliverables. Against this backdrop, the body of relevant knowledge is bereft of a study investigating the association between workers' interpersonal skills and interpersonal communication (IC) quality. Thus, this study aims to predict the quality of professionals' IC through a multi-pronged artificial intelligence-based methodological approach. In doing so, the literature is reviewed to capture noticable interpersonal skills (IPSs), followed by utilizing a fuzzy-based algorithm to prioritize them. Then, an extreme Gradient Boosting (XGBoost)-based algorithm is developed to predict the quality of workers' IC. The developed XGBoost is finally applied to three real-life construction projects to check its efficacy. Based on the application of the developed model to the selected case studies, the following conclusions are drawn: (1) the significant skills are "Leadership Style," "Listening," "Team Building," and "Clarifying Expectations"; and (2) the predictions of the developed model equal to what happens to the workers' IC quality in more than 78% of the cases. The developed algorithm can warn interpersonal conflicts before they escalate, enhance job-site productivity, team development, and human resources management, and guide construction managers in developing IPSs training.

Keywords Construction; Communication Quality; Interpersonal Skills; Machine Learning.

1. INTRODUCTION

The role Efficient communication plays in enhancing the performance of construction project is unignorable (Pamidimukkala et al., 2023). The link between communication and performance has been well established. Across countries as diverse as Norway, Indonesia, and Australia, we see that the overall structure of project teams, procedures, and working norms, must be constantly adjusted and fine-tuned in order to raise communication efficiency in the pursuit of productivity (Elghaish et al., 2023). In the same vein, Yolanda et al. (2021) found that barriers to communication stifle employee performance.

There is a need for measuring the Communication Quality (CQ) in the construction industry (Mohr and Sohi, 1995, Kwofie, 2015, Forcada et al., 2017, Hosseini et al., 2017, Safapour et al., 2020), however, few methods such as those utilized by Rahimian et al. (2022) have been proposed for evaluating the CQ, as determined by an individual's IPSs (Nogueira, 2022).

Strong IPSs are a prerequisite for enhancing stakeholder communication and in fostering strong relationships (Ghorbani, 2023). They are also necessary to the cultivation of employee loyalty to an organization (Nogueira, 2022). Previous studies, however, have only placed emphasis on indicators affecting project communication. Yet we know IPSs are integral to managing IC (Kundi et al., 2023). An in-depth understanding of the impact of the IPSs of workers on the CQ can potentially alert managers to conflicts before they escalate. This will in turn improve team development and promote productivity gains and intelligent human resource management.

Previous studies have been limited to methods based on structural equation modeling (Hosseini et al., 2017), the chi-square test (Forcada et al., 2017), and factor analysis (Safapour et al., 2020). In summary, this study aims to address the following Research Question (RQ):

RQ. How could a predictive model for evaluating the CQ among workers, in which the uncertainty of IPSs is overcome, be developed and validated?

To address this question, a machine learning algorithm – XGBoost (XGB) – was developed to evaluate CQ based on the individual's set of IPSs. The tailored predictive model can be used by managers to quantify the CQ with their workers, leading to the identification of conflict-prone communication, enhanced job-site productivity and team development, and improved analysis of the IPSs of job applications during recruitment.

The paper is structured as follows. First, a review of the literature is presented. The research methods section outlines data collection protocols and the analysis techniques. This is followed by a description of the findings, along with a discussion. The implications of findings, research limitations, and future avenues for research are summarized in the final section.

2. CONTEXTUAL BACKGROUND

2.1 Quality of communication

High-quality communication is essential in construction organizations (Mithas et al., 2011), primarily because of its role in improving the effectiveness of project teams (Segerstedt et al., 2010, Bosch-Sijtsema and Henriksson, 2014). Claimed by Mohr and Sohi (1995), researchers apply two indicators when evaluating the CQ. They are communication flow, i.e., the manner and communication frequency and the CQ. Aubert et al. (2013) identify the central features of communications to be timeliness and accuracy, with both of equal importance.

CQ indicators in the context of construction project management have been captured (Ochieng and Price, 2010, Affare, 2012, Senescu et al., 2013, Kwofie, 2015, Forcada et al., 2017, Safapour et al., 2020). Armstrong and Taylor (2003) asserted that the number of distinct geographic regions and CQ have a negative correlation. Moreover, a large proportion of cultural and ethnic groups are associated with low communication effectiveness (Nam et al., 2009). Tone et al. (2009) reported that CQ can be thoroughly affected by bureaucracy level within an organization. Senescu et al. (2013) established that communication complexity significantly impacts an organization's CQ. Westin and Sein (2014) defined accessibility as an essential element of communication while highlighting the breadth of complicated tasks assigned to construction workers. Hosseini et al. (2017) identified five more indicators (i.e., sense of presence, documentability, persuasiveness, accessibility, and relevancy). They determined that team members must have a holistic appreciation of a project while also maintaining high-quality communication. Similarly, Wang et al. (2014) determined that construction project teams must establish strong interrelationships. A summary of CQ indicators is provided in Table 1.

Table 1. Compilation of CQ indicators

Indicator	**Definition/relevance**	**Reference(s)**																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Accuracy	Whether data are correctly transferred without bias, distortion, or withheld information.	✓	✓	✓	✓	✓	✓	✓	✓												
Accessibility	The speed and accuracy with which the exchange of information occurs.							✓		✓											
Bidirectionality	The obtainability of verification through feedback and clarifications.					✓		✓			✓										
Clarity of scope and objectives	There is a positive correlation between clarity of scope, objectives and CQ.				✓							✓									
Completeness	The completeness of data and essential information in the exchange.	✓	✓			✓		✓	✓				✓								
The complexity level of communication	Communications become distorted as complexity increases.													✓							

Documentability	Tacit documentation of information exchanged is an attribute of high-quality communications.							✓													
Frequency	Denotes how frequently parties partake in communication with each other.							✓			✓				✓						
Formality	The extent to which communication flows are planned, structured, planned, and made routine.										✓										
The inexperience of project managers	Inability to apply effective techniques, tools, knowledge, and skills needed meet project requirements.															✓					
Persuasiveness	Degrees to which people can convince others of an idea or the value of an action.							✓									✓				
Reliability	The perceived accuracy and value of the information received.		✓	✓		✓		✓													
Relevancy	Relevancy of communications is a measure of the extent to which information is helpful and applicable to the required task.							✓													
Sense of presence	Team members must feel connected with others to maintain high-quality communication.							✓										✓			
Timeliness	Information is provided when needed, not late nor too early.	✓	✓	✓	✓	✓	✓	✓	✓	✓											
Top-down bureaucracy	It negatively affects the CQ.											✓							✓		
Understandability	The receiving party must correctly comprehend and interpret the information provided.	✓	✓	✓	✓	✓		✓													
Variety of geographic regions	A much smaller range of cultural and ethnic groups are able to establish effective communication.																			✓	✓

Note: 1: Thomas et al., 1998; 2: Kahn et al., 2002; 3: Miller, 2005; 4: Xie et al., 2010;5: Aubert et al., 2013; 6: Forcada et al., 2017; 7: Hosseini et al., 2017; 8: Trach et al., 2021;9: Westin and Sein, 2014; 10: Mohr and Sohi, 1995; 11: Safapour et al., 2020; 12: Xu et al., 2003;13: Senescu et al., 2013; 14: Ellwart et al., 2015;15: Safapour et al., 2019; 16: Den Otter and Emmitt, 2007; 17: Wang et al., 2014; 18: Tone et al., 2009; 19: Armstrong and Taylor, 2003; 20: Nam et al., 2009.

2.2 Interpersonal skills

Robbins and Hunsaker (2011) extensively reviewed the most common IPSs. They can be divided into the sub-categories of motivation, leadership, and communication process (Figure 1). It has been suggested that negotiation is yet another related skill that does not fall into a specific category (Arabi and Khoshneyat, 2022).

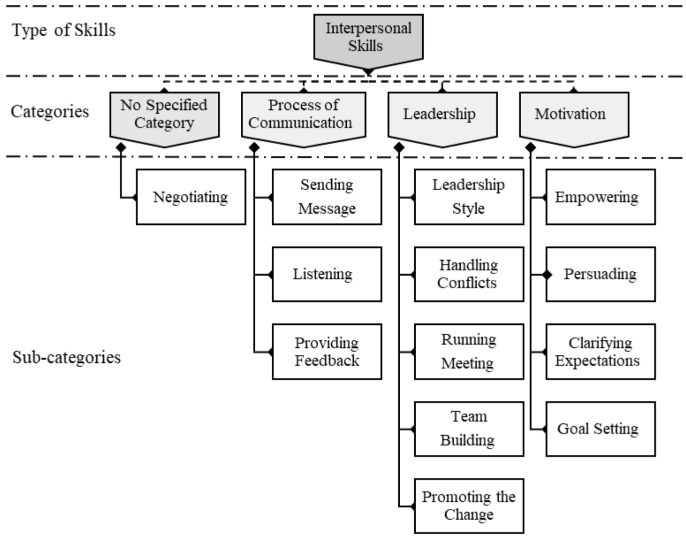

Figure 1. Conceptualization of IPSs

As claimed by Robbins and Hunsaker (2011), practical IPSs involve four skill sets. The first is leadership. Leaders must be able to energize teams to accomplish tasks while supporting the integration of members in achieving that task (Ghorbani, 2023). To that end, leadership style, the ability to handle conflicts, approaches to running meetings, team-building skills, and the ability to promote and foster change all contribute to the success or otherwise of the leader. Secondly, another important role of the manager is to create conditions that foster a sense of motivation to achieve the group's mission (Nogueira, 2022). This involves the leaders' ability to set goals, clarify expectations, and persuade others. Thirdly, the preceding depends on the leaders' ability and process used in communicating with others. Finally, although the mission presented by the leader may appear fixed, its realization involves many parties, each of whom may have specific views on the value of the mission and how it will be realized. The leader must navigate a mutually acceptable compromise in bringing people to a common purpose.

2.3. Point of departure

Although numerous papers on CQ evaluation have been published, there is a relative dearth of information on the IC of workers. This type of communication significantly influences team development, productivity, and facilitates intelligent resource management. This study critically examines the IC of construction workers and focuses on their IPSs, as the core of IC, to make predictions accordingly. To address the aforementioned limitation, a hybrid AI-based model was suggested in this study. The predictive nature of the suggested approach could allow managers to identify and address interpersonal conflicts before escalation.

3. RESEARCH METHODOLOGY AND MODEL DEVELOPMENT

This section elaborates on the data gathering protocols, the data analysis techniques, and the developed XGB algorithm. Overall, the method comprises three steps: data collection and preparation, XGB development, and XGB performance in a real context (Figure 2). Step 1 commences upon data collection, considers the IPSs and preparation, followed by the ranking of these skills using Fuzzy Analytic Hierarchy Process (F-AHP). The essential skills are then fed into the XGB algorithm to train and tailor it (step 2). The suitability of these techniques is well explained in the respective sub-sections. In the third and last step, the developed algorithm is implemented in three construction projects to assess its capability to predict the quality of IC between construction professionals within a specified time period.

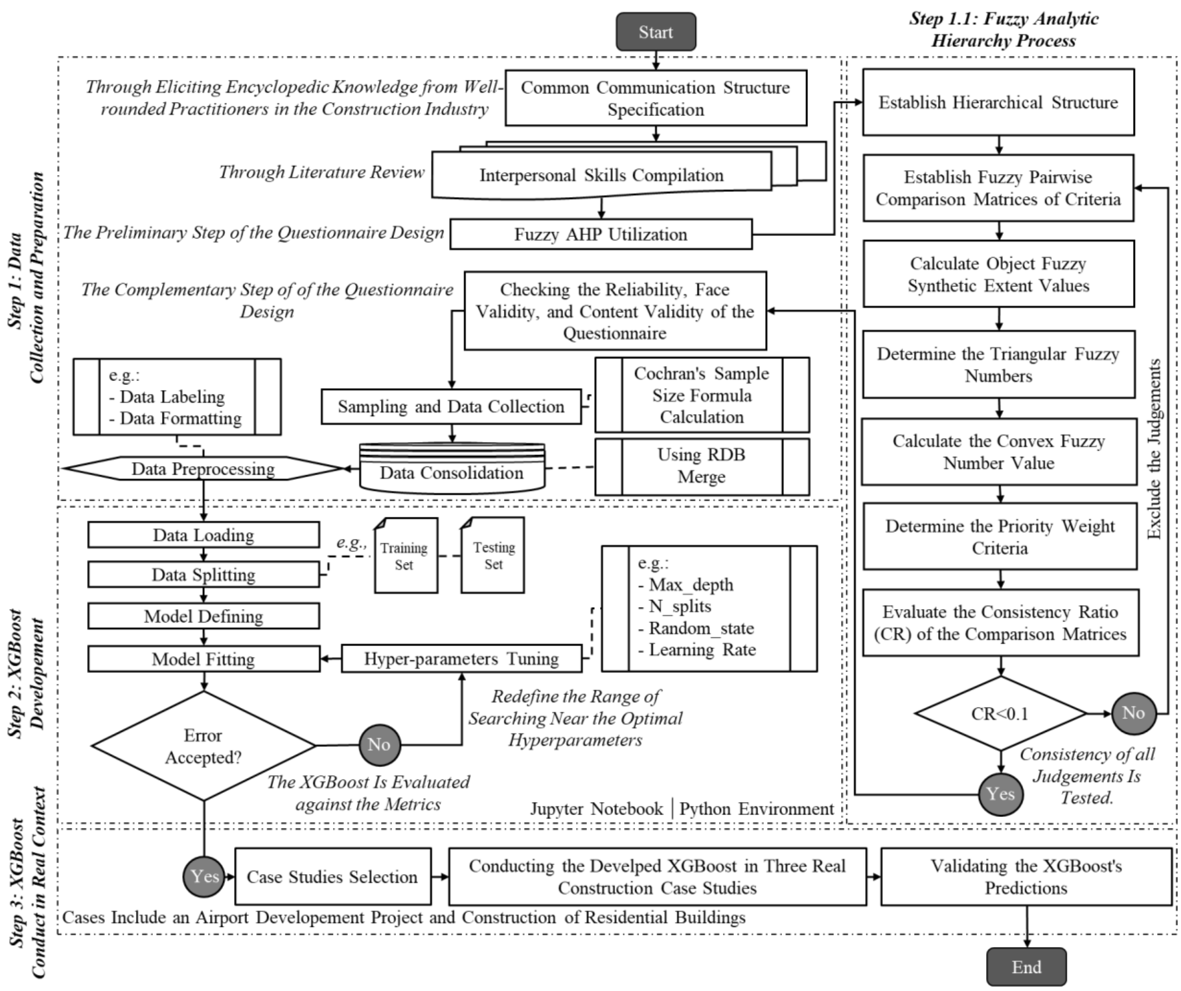

Figure 2. Flowchart of the established three-step method

3.1 Step 1: Data collection and preparation

3.1.1. Communication structure typical of the construction industry

At the outset, a typical construction project communication network was modeled, given the experience and views of experts with broad knowledge about a variety of construction projects. The information about the experts was clarified in the sampling sub-section. These relationships defined the connection between construction project workers, for example, a site-head, and other practitioners in a project. In practice, the relationships imply the connections between individuals, which revealed the respondents of the questionnaire whose IPSs should be assessed.

3.1.2. Interpersonal skills compilation

As mentioned in sub-section 2-2, the IPSs set identified by Robbins and Hunsaker (2011) was used in this study. It is important to appropriately collect the data required in relation to these skills and to follow the ultimate goal of the study, which is to develop a predictive algorithm. However, the development of algorithms of this kind demands specific data types. Therefore, before presenting the other sub-sections, a brief explanation of the rationale for the questionnaire used to gather the required data is examined.

3.1.3 Rationale behind the data types that were utilized

Predictive models allow decision-makers to identify routes to the available information (Zhang et al., 2021) and while demonstrating the odds that something will (or will not) occur. Broadly speaking, they consist of three principal types: 1) Reinforcement learning, 2) supervised learning, and 3) unsupervised learning (Kapoor et al., 2022). In addition, if the outputs are discrete, the problem involves classification, whereas if the outputs are continuous, the problem involves regression. The model proposed in this report is a supervised classification problem. This category involves a predictive model design with data containing the results that will be predicted (targets). The primary and cross-sectional dataset were gathered using the designed questionnaire.

3.1.4 Designing the questionnaire: the preliminary step

In eliciting information from participants, quantitative data collection is desirable since it allows for a generalization of findings (Chun Tie et al., 2019). The preliminary step includes the identification of salient IPSs. According to experts from academia and industry, the skills must first be ranked to select the most important ones. This ranking was done using the F-AHP.

3.1.4.1 F-AHP utilization

As previously indicated, this study adopted the model developed by Robbins and Hunsaker (2011), in which 13 skills were synthesized in total. Assessing all these skills is would be ideal. Nevertheless, there is a time constraint. The more skills that are assessed using the questionnaire, the more time is needed to address them, with the respondents' patience being a limited commodity. Therefore, it was decided to rank the different skills according to their relative importance (weights).

AHP, developed by Saaty (1980), was utilized as a suitable approach for two reasons. It ranks a decision maker's judgment by assessing the vitality of a decision maker's intuitive decisions and the consistency of the comparison alternatives in the decision-making process.

However, simple AHP alone cannot manage the uncertainties associated with criteria decision-making problems based on quantitative data. Fuzzy set theory is required for greater robustness and accuracy in judgments (Khan et al., 2021). This combination of the two approaches has the benefit of being able to manage decision-making problems based on both quantitative and qualitative data. The criteria priority weight and preferred rating were measured using Fuzzy triangular numbers. The concept of the F-AHP approach developed by (Khan et al., 2021) was utilized to calculate the priority weight of IPSs.

Let $V = \{v_1, v_2, \dots, v_n\}$ be a set of objects for the main categories and $U = \{u_1, u_2, \dots, u_n\}$ represent the goal set of each category. (Khan et al., 2021)stated that the extent analysis for each object and goal (gi) should be performed respectively. Therefore, the (m) extent analysis of each object is performed using Eq. (1):

$$X_{gi}^1, X_{gi}^2, \dots, X_{gi}^m, i = 1,2, \dots, n \quad (1)$$

where X_{gi}^i (i = 1, 2, …, n) are Fuzzy triangular numbers (TFNs). The in-depth extent analysis instructions by (Khan et al., 2021) are considered next.

Step 1. The i_{th} object Fuzzy synthetic extent value should be calculated using Eq. (2):

$$S_i = \sum_{j=1}^{m} X_{gi}^{j} \otimes \left[\sum\nolimits_{i=1}^{n} \sum\nolimits_{j=1}^{m} X_{gi}^{j}\right]^{-1} \quad (2)$$

where $\sum_{j=1}^{m} X_{gi}^{j}$ is calculated using the m extent analysis Fuzzy addition operation such as:

$$\sum_{j=1}^{m} X_{gi}^{j} = \left(\sum_{j=1}^{m} X_{gi}^{l}, \sum_{j=1}^{m} X_{gi}^{jm}, \sum_{j=1}^{m} X_{gi}^{r}\right) \quad (3)$$

Similarly, $\left[\sum_{i=1}^{n} \sum_{j=1}^{m} X_{gi}^{j}\right]^{-1}$ is calculated using an extensive Fuzzy addition operation on the X_{gi}^{j} $(j = 1, 2, \ldots, m)$ value:

$$\sum_{i=1}^{n} \sum\nolimits_{j=1}^{m} X_{gi}^{j} = \left(\sum\nolimits_{i=1}^{n} X_{i}^{l}, \sum\nolimits_{i=1}^{n} X_{i}^{m}, \sum\nolimits_{i=1}^{n} X_{i}^{r}\right) \quad (4)$$

Eq. 5 is utilized to calculate the inverse of the vector:

$$\left[\sum_{i=1}^{n} \sum\nolimits_{j=1}^{m} X_{gi}^{j}\right] = \left(\frac{1}{\sum_{i=1}^{n} X_{i}^{r}}, \frac{1}{\sum_{i=1}^{n} X_{i}^{m}}, \frac{1}{\sum_{i=1}^{n} X_{i}^{l}}\right) \quad (5)$$

Step 2. Both X_1 and X_2 are Fuzzy triangular numbers, and Eq. 6 defines the degree to which $X_2 = (X_2^l, X_2^m, X_2^r) \geq X_1 = (X_1^l, X_1^m, X_1^r)$:

$$Y(X_2 \geq X_1) = sup\big[min\big(\mu_{x1}(v)\big)\big], \big(\mu_{x2}(v)\big)\big] \quad (6)$$

Eq. 6 can be simplified as follows:

$$Y(X_2 \geq X_1) = hgx(X_1 \cap X_2) = \mu_{x2}(d) = \begin{Bmatrix} 1 & if\ x_2^m \geq x_1^m \\ 0 & x_1^l \geq x_2^r \\ \dfrac{x_1^l x_2^r}{(x_2^m - x_2^r) - (x_1^m - x_1^l)} & Otherwise \end{Bmatrix} \quad (7)$$

where d is the highest intersection point between μ_{x1} and μ_{x2}.

Step 3. The value of the convex Fuzzy number $X_i (i = 1, 2, \ldots, K)$ could be measured using the following equation:

$$Y(X \geq X_1, X_2, \ldots, X_k) = min\, Y(X \geq X_i) \quad (8)$$

Suppose that $d'(X_i) = min\ Y(X_i \geq X_k)$ for $K = 1, 2, \dots, n; K \neq i$.

The weight vector of each element could be calculated using Eq. 10:

$$W' = \left(d'\ (X_1), \quad d'\ (X_2), \quad d'\ (X_e), \quad d'\ (X_n)\right)^X \tag{10}$$

where $X_i (i = 1, 2, \dots, n)$ consists of n distinct elements.

Step 4. The priority weight criteria could be determined using Eq. 11 by normalizing the weight vector and converting the results into a non-fuzzy number:

$$W = \left(d(X_1), d(X_2), d(X_3), \dots, d(X_n)\right)^X \tag{11}$$

The value of W shows a non-Fuzzy number.

Step 5. Checking the consistency ratio:

It is necessary to evaluate the consistency of the pairwise comparison matrices (Mohandes et al., 2022). In this study, the graded mean integration ratio was used to check the consistency of the results obtained as follows:

$$Q_{crisp} = \frac{(4m + l + r)}{6} \tag{12}$$

After the preceding calculations are performed, the following equations are used to determine the Consistency Index values (CI) and Consistency Ratio (CR), respectively:

$$CI = \frac{\lambda_{max} - n}{n - 1} \tag{13}$$

$$CR = \frac{CI}{RI} \tag{14}$$

where λ_{max} is the maximum eigenvalue, n is the size of the pairwise comparison, and RI could be determined using the predefined tables. The comparison matrices are consistent if the value of CR is greater than 0.1; otherwise, the consistency ratio should be re-evaluated until an acceptable deal is achieved.

3.1.5 Designing the questionnaire: the complementary step

3.1.5.1. Checking the reliability, face validity, and content validity of the questionnaire

Next, a close-ended survey questionnaire containing a Likert scale ranging from 1 (Novice) to 5 (Expert) for the first two sections and from 1 (Very Low) to 5 (Very High) for the third section was designed and developed based on interviews with experts and a literature review (see Appendix A).

The reliability or internal consistency of the questionnaire was verified using Cronbach's alpha (α) (Blunch, 2012), calculated based on Eq. (15).

$$A = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k} \sigma_i^2}{\sigma_X^2}\right) \tag{15}$$

where k is the number of items, σ_i^2 is the variance of the i_{th} item, and σ_X^2 is the variance of the total score formed by summing all the items. Notably, A ranges from 0 (unreliable) to 1 (reliable). If the items that make up the score have such a perfect correlation, then α = 1. However, the reverse is true when they are independent, in which α = 0. Therefore, it is evident that the reliability of the generated scale is positively correlated with the score. In the present study, α was 83%, much higher than the threshold value of 70%.

Moreover, the questionnaire's face and content validity were confirmed both qualitatively and quantitatively by 20 experts using the approach suggested by Hosseini et al. (2018). The quantitative value of face validity indicated that all the items in the questionnaire had an impact score of more than 1.5. The content validity of the items (selected IPSs for use in the questionnaire) in the quantitative analysis was also measured and validated based on the Content Validity Ratio (CVR) (Spoto et al., 2023), calculated using Eq. (16):

$$CVR = \frac{n_e - \left(\frac{N}{2}\right)}{\frac{N}{2}} \tag{16}$$

where n_e is the number of experts indicating "essential," and N is the total number of experts. The final evaluation to retain the items based on the CVR depends on the number of experts. Since the number of experts was 20, the minimum CVR could not be less than 0.42—all the items met this condition, and the CVR value was not lower than 0.53 for any item.

The questionnaire was then operationalized, as is discussed next. In the first section, respondents were required to perform a self-assessment and rate their IPSs level—the exact meaning of each level is described in appendix C. The second section asked the respondents to assess another employee with whom they communicate interpersonally, given the predetermined connections. Finally, in the third section, they were asked to rate the CQ with the employee in question. Using this approach, the data necessary for developing the machine learning model were gathered.

3.1.6 Sampling and data collection

A list of certified construction companies working in Iran was downloaded from the licensed contractors' data bank. Generally, the sample sizes were large. However, this difficulty is inherent when a host of samples is gathered. Cochran's sample size formula (Eq. 7) reduces the sample size and logically allows the representative statistical population to be determined. The precision was ±5%, the confidence level was 95%, q was 0.5, and Z was 1.96 (Z^2 = 3.8461) for the first recommended sample which consisted of 240 construction firms in Tehran (the capital of Iran). Seventy-five companies were ultimately used to conduct the survey. Blank survey questionnaires were distributed by mail or in person to 230 senior employees. The details, including the respondents' occupation, major concentration, educational qualification, working party, work experience, and gender, are tabulated in Table 2. The data were collected over five months, and, in total, 185 correctly completed questionnaires (166 by men and 19 by women) were retrieved, with an 80% response rate. The Population size is calculated using Eq. 17:

$$n = \frac{\frac{Z^2pq}{e^2}}{1 + \frac{1}{N}\left(\frac{Z^2pq}{e^2} - 1\right)} \tag{17}$$

where N is the population size, e is the desired level of precision, z is the selected critical value of the desired confidence level, p is the estimated proportion of an attribute present in the population, and q = 1 – p. Given that the value of p is unknown, the value of 0.5 usually is used.

Table 2. Demographic profile of questionnaire respondents

Characteristic	Category	Percentage (%)
Occupational position	Technical Office	12 %
	Sub-Contractors	15 %
	Contractor Project Manager	9 %
	Supervision Consultancy Team	7 %
	Consultant Project Manager	7 %
	Industrial Engineer	11 %
	Site Manager	24 %
	Design Team Members	11 %
	Client Project Manager	4 %
Major concentration	Civil Engineering	52 %
	Construction Engineering and Management	20 %
	Architectural Engineering	11 %

	Industrial Engineering	8 %
	Structural Engineering	9 %
Educational background	B.Sc.	72 %
	M.Sc.	15 %
	Ph.D.	13 %
Work experience	5 to 10	19 %
	11 to 15	33 %
	16 to 20	22 %
	21 to 30	26 %
Gender	Female	8 %
	Male	92 %
	Prefer not to Answer	0 %
Working party	Contractor	71 %
	Consultant	25 %
	Client	4 %

3.1.7 Data pre-processing: data encoding and formatting

Generally, data fall into three categories: structured, semi-structured, and unstructured. Structured data, contrary to the other two, are represented in matrix form with rows and columns. The collected data of this study were structured and contained ordinal variables with a finite number of classes or categories. However, this is not the appropriate category of data for machine learning algorithms. Instead, they need data in numerical form. As such, each category was assigned an integer value between 1 to 5. Thus, "Expert" was encoded as 5, "Journeyman" as 4, etc. In addition, ordinal encoding was used to encode target labels (last column) representing the CQ. Ordinal encoding assigns a sequence of numerical values between 0 and the number of classes minus one ($n_{class} - 1$), per the order of data. As such, "Very High" was encoded as 4, "High" as 3, and so forth (see Table 3). Ordinal encoding was performed using the Scikit-learn library in PYTHON.

It is noteworthy that the dataset consists of 5616 numerical values, 624 rows and nine columns (eight features and one target). Appendix B includes a table containing some of the data used in this paper.

Table 3. The selection approach for encoding numerical values in the present study

Proficiency Level	Equivalent Value	Communication Quality	Ordinal Encoded Value
Expert	5	Very high	4
Journeyman	4	High	3
Apprentice	3	Medium	2
Initiate	2	Low	1

Novice	1	Very low	0

The survey data were then cleaned. This task dealt with missing values (NaN entries). PYTHON data manipulation library was used for this purpose (Wu, 2013). Furthermore, classes of CQ were formatted using one-hot encoding to develop the intended model that predicts the probability of an instance that belongs to each of the five classes.

3.2 Step 2: XGB model establishment

In this study, the XGB algorithm was chosen to forecast the quality of IC for two principal reasons. First, XGB is one of the most well-known boosting tree algorithms for gradient boosting machines (GBM). It is frequently employed in the field since it performs well in problem-solving tasks, and the minimal requirement for feature engineering is high (Möller et al., 2016, Tamayo et al., 2016, Dong et al., 2020). Second, compared to deep learning algorithms, XGB is better suited to small datasets running on a CPU (Dong et al., 2020). Considering the database size in this study (624 instances), the XGB algorithm may be more appropriate than deep learning approaches. XGB is an open-source library that provides a boosting algorithm in Python and other languages such as C++ and Java.

3.2.1 Classification model

The classification model in Figure 3 depicts how the dataset was interrogated in the model. The dataset includes testing and training data. Both sets were consolidated into one data file. Python was used to combine the data and XGB and Anaconda, which was an in-built package. The dataset was opened using Python based on various parameters pertinent to XGB used to run the dataset.

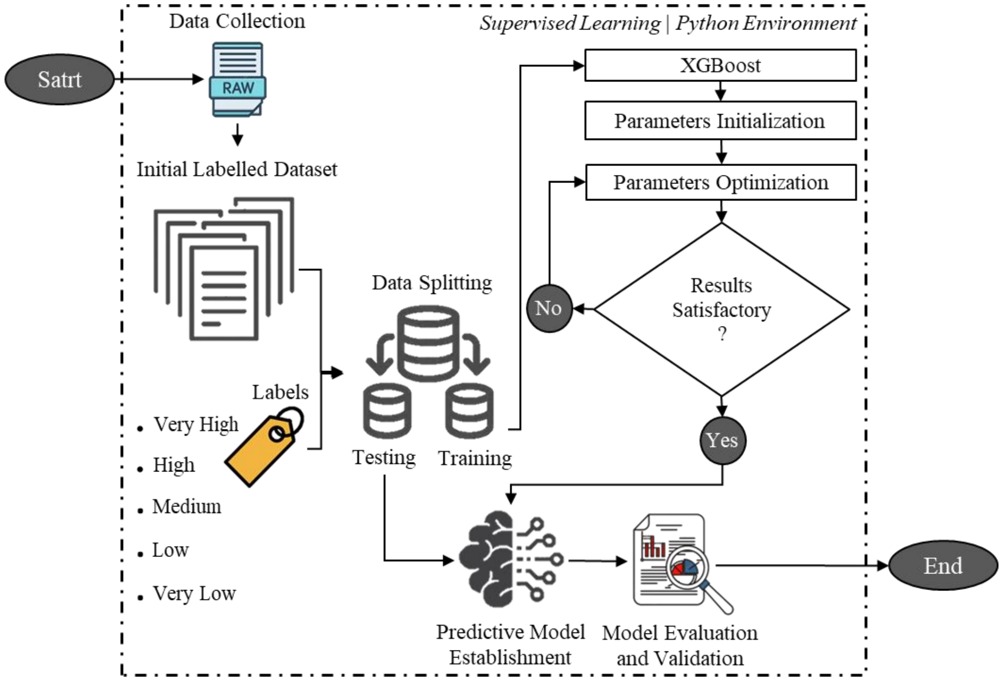

Figure 3. Classification model

As previously indicated, the data were labeled, and the predictive model was used to learn the relationships between the training inputs and targets; hence, the task involved supervised learning. The problem is also s multi-class classification given that the IPSs were rated on a five-point scale and were single-label since each data point should be classified into one category. Thus, respondents were only allowed to choose the extent to which they were proficient in specific IPSs; therefore, the problem is a single-label, multi-class classification. As new data was received, it was mapped onto the already-established functions, which in turn, improved the algorithm, and is the core of machine learning (Talukder et al., 2023).

3.2.2 Boosting

Boosting is a machine-learning algorithm that reduces dataset bias and variance. It allows weak learners to become more robust. Algorithms achieve this outcome in a process called "boosting" (Dhaliwal et al., 2018). In XGB, trees are optimized with gradient boosting (Friedman, 2001, Linardatos et al., 2020). Let the output of a tree be:

$$f(x) = w_q(x_i) \tag{18}$$

where x is the input vector and w_q is the score of the corresponding leaf q. The output of an ensemble of K trees includes:

$$y_i = \sum_{k=1}^{K} f_k(x_i) \tag{19}$$

The XGB algorithm seeks to minimize the following objective function J at step t:

$$J(t) = \sum_{i=1}^{n} L\left(y_i, \hat{y}_i^{t-1} + f_t(x_i)\right) + \sum_{i=1}^{t} \Omega(f_i) \tag{20}$$

where the first term encompasses the train loss function L (e.g., mean squared error) between real class y and output $\hat{y}$ for the n samples, and the second term is the regularization term, which controls the complexity of the model and aids in avoiding overfitting. In XGB, the complexity can be defined as:

$$\Omega(f) = \Upsilon T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{21}$$

where T refers to the number of leaves, γ is the pseudo-regularization hyperparameter, depending on each dataset, and λ to the L2 norm for leaf weights.

Applying gradients for second-order approximation of the loss function and finding the optimal weights w, the optimal value of the objective function is:

$$J(t) = -\frac{1}{2}\sum_{j=1}^{T} \frac{\left(\sum_{i \in I} g_i\right)^2}{\sum_{i \in I} h_i + \lambda} + \gamma T \tag{22}$$

where $g_i = \partial_{\hat{y}^{t-1}} L(y, \hat{y}^{t-1})$and $h_i = \partial^2_{\hat{y}^{t-1}} L(y, \hat{y}^{t-1})$are the gradient statistics of the loss function, and I is the set of leaves.

3.2.3. XGB parameters and evaluation

XGB has many hyper-parameters, which can be used to perform specific tasks. The tuned hyper-parameters are listed in Table 4. A corresponding model is generated given the pre-processed training set after setting the (initial) values for the hyper-parameters to train the XGB algorithm. During this, the K-fold cross-validation method is used to refine the training performance by randomly splitting the training set into k distinct subsets, called folds. Next, it trains and evaluates the established XGB model k times, picking one of the folds for evaluation every time and training on the other (k − 1) folds. The readers are referred to the following reference (Dhaliwal et al., 2018) for a more thorough review of the hyper-parameters used in this study.

Table 4. Set values of the tree booster hyperparameters for the XGB model.

Hyperparameters	Values
Max_depth	3
Learning_rate	0.3
N_estimators	100
N_splits	10
Objective	Multi: Softprop
Random_state	7
Booster	Gbtree
Reg_alpha	10
Num_class	5

The correctness of classification can be evaluated by computing the number of correctly recognized class examples (True Positives), the number of correctly identified examples that do not belong to the class (True Negatives), and examples that were incorrectly assigned to the class (False positives) or that were not recognized as class examples (False negatives). According to the confusion matrix values, the most often used measures for a multi-class setting are Accuracy (Zhu et al., 2010), Recall (Sokolova and Lapalme, 2009), and Precision (Sokolova and Lapalme, 2009), as was the case for XGB. Thus, classification accuracy is the proportion of correct predictions. Classification accuracy is generally converted into a percentage, where 100% is a perfect classifier, and an accuracy of 0% is a perfectly wrong classifier. The Recall is the proportion of the true class predictions that are correctly predicted over the number of true predictions. This is also known as the True Positive Rate or Sensitivity. A Recall of 1 is a perfect Recall, while 0 is a "bad" Recall. Precision is the proportion of a class predicted to be in a given class and is actually in that class. These metrics were calculated using the following equations:

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \tag{23}$$

$$Precision_M = \frac{\sum_{i=1}^{l} \frac{TP_i}{TP_i + FP_i}}{l} \tag{24}$$

$$Recall_M = \frac{\sum_{i=1}^{l} \frac{TP_i}{TP_i + FN_i}}{l} \quad (25)$$

Where for an individual class C_i: TP_i are True Positive for C_i, and FP_i – False Positive, FN_i– False Negative, and TN_i – are True Negative counts. l and M also represent the number of classes and macro-averaging (the average of the same measures calculated for $C_1, \dots, C_i$) showing the quality of overall classification, respectively.

3.3 Step 3: XGB conducted in real context

The trained XGB was evaluated using three real illustrative cases to evaluate its predictions (Figure 2). One case (illustrative example I) was an airport development project, and the other two (illustrative example I and II) were both the construction of residential buildings. They were selected for two primary reasons. Firstly, the contractors and consultants had a good reputation and extensive history in the field and were successful in various construction projects. According to Iran's classification of contractors, their firms were also classified in the top-rank organizations. The second reason is because of their employees' interest in this research, their commitment, and the cooperation of clients in providing access to the research-related data. Appendix D provides the detailed introductory information in each instance.

In each case, an expert with the most connections related to project personnel was tasked with validating the predictions of XGB based on what occurred between construction workers over time. To better capture the XGB's performance, evaluated workers accounted for a selection of team members from various departments and levels. The experts indicated whether a relationship exhibited interpersonal problems or mistreatment behaviors, contrary to the accepted norm of the workplace. This step of the study was carried out in 2020 over eight months, from predicting 183 qualities to finalizing the experts' evaluation.

4. RESULTS

This section explains the results obtained from each step of the method. The first two sections outline the minor findings, which pave the way for preparing, training, and implementing the predictive model.

4.1 Interpersonal skills ranking

All the IPSs are significant to a degree; however, considering all of them in the questionnaire was unreasonable. This could have caused the responding process to be unnecessarily time-consuming, resulting in a sub-optimal completion of the questionnaire by the experts. Having identified the relative importance of each skill using F-AHP, the authors narrowed down the skills pool to four—based on the judgment and timeframe, the number of skills that require the shortest time to assess while maintaining the maximum accuracy of the responses was considered to be 4. They include leadership style, listening, team building, and clarifying expectations (Figure 4). The selected skills were considered in the questionnaire, and respondents were tasked to individually determine the extent to which they possessed these skills. In addition, the definition or explanation of each skill was presented to the respondents to better understand the assessed items. Notably, the prioritization consistency ratio was less than 10%, indicating that the given matrix was sufficiently consistent.

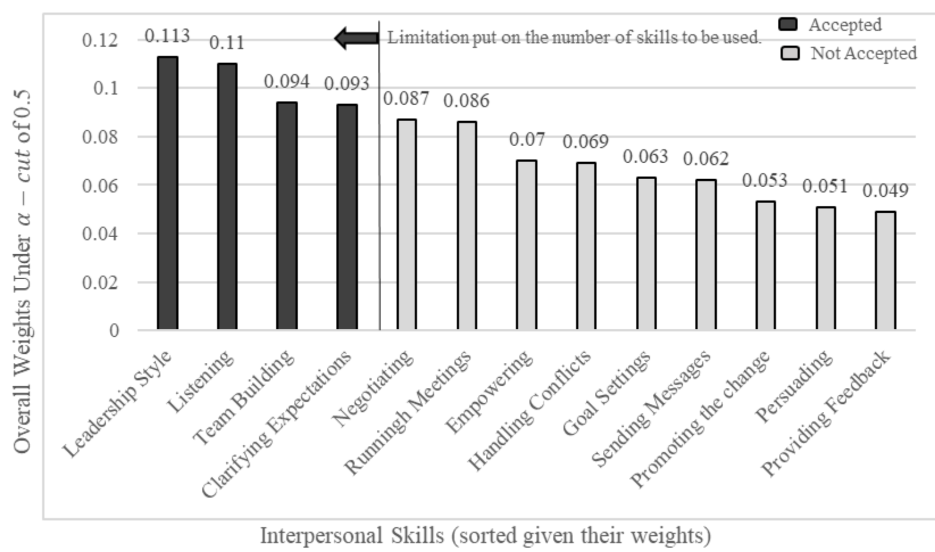

Figure 4. F-AHP output demonstrates the significance of IPSs based on experts' judgment.

4.2 Communication structure typical of construction projects

Error! Reference source not found. shows the ubiquitous Interrelated relationships among the c onstruction stakeholders, consisting of 10 major roles and 23 connections. All the experts were unanimous on the ubiquitous of this formal communication network. As tabulated, the site manager and supervision team had the most links with other members. The tabulated connections played an essential role in the questionnaire. As indicated, in the first two sections of the questionnaire, respondents were requested to perform a self-assessment regarding the given four IPSs and to assess other employees with whom they communicated regarding the predetermined connections shown in Table 5.

Table 5. Interrelated relationships among the construction stakeholders

Stakeholders	Client Project Manager	Site Engineer	Design Team	Site Manager	HSE	Consultant Project Manager	Supervision Team	Contractor	Sub-Contractor	Technical Office
Client Project Manager	-	-	-	-	-	*	*	*	-	-
Site Engineer	-	-	-	-	-	-	*	-	*	*
Design Team	-	-	-	-	-	*	*	-	-	*
Site Manager	-	*	-	-	*	*	*	*	*	*
HSE	-	*	-	*	-	-	-	-	*	-
Consultant Project Manager	*	-	*	*	-	-	*	*	-	-
Supervision Team	*	*	*	*	-	*	-	*	-	*
Contractor	*	-	-	*	-	*	*	-	-	-
Sub-Contractor	-	*	-	*	*	-	-	-	-	-
Technical Office	-	*	*	*	-	-	*	-	-	-

Note: * denotes the relationships between the stakeholders

4.3 Accuracy, Recall, and Precision

All the preliminary steps were performed to train the machine learning algorithm and to save the metrics by which XGB was evaluated, including Accuracy, Recall, and Precision calculations. The accuracy calculated using XGB was 82% (0.82659574), higher than the conventional threshold value of 75%. The Recall was 79% (0.79089015). The Precision of the model was 78% (0. 78555023), indicating that was correct just over 80% of the time. The best results for XGB were achieved for Accuracy, followed by Recall and Precision. Accuracy, Precision, and Recall were also tested as K fractions, the number of splits in tree formations. As the number of splits increased, there was a corresponding decrease in the values since splits allow the model to better learn about the datasets.

4.4. Validating the results obtained from the method implementation

In the final step, the trained algorithm was applied in a real context to evaluate its predictive capabilities. In total, it predicted 183 communication qualities. As Table 5 reveals that the site manager and supervision consultancy team shared the most connections with other project members. They were asked to comment on any exhibited interpersonal problems or conflicts of interests.

Among the predictions, fifty-five were rated as "Very High" or "High" quality. Among them, 48 had no signs of mistreatment or interpersonal problems such as unwillingness to cooperate, tension, or IC issues. Seven experienced interpersonal conflicts or other types of interpersonal issues at some level. Eighty-nine qualities were predicted as "Medium." A total of 67 relationships did not exhibit conflicts of any kind, but 22 exhibited this issue. Among the rest, 39 relationships were predicted as "Low" or "Very Low," 12 had no conflicts, but 27 had struggled with interpersonal problems of high intensity, such as rudeness or yelling. If those predicted as "Medium" were not expected to demonstrate interpersonal problems, the developed XGB was correct in predicting 142 IC qualities, meaning that it was accurate in nearly 78% of the cases.

5. DISCUSSION

5.1. Benchmarking against previous studies

The outputs of F-AHP reveal that leadership style, listening, team building, and clarifying expectations are the most important skills. Leadership style is crucial to construction experts because it directly and significantly affects productivity, efficiency, and performance of not only themselves but also their team members (Matin et al., 2010, Nogueira, 2022). Listening is considered to be one of the four most important communication skills, which aligns with the finding by (Matin et al., 2010, Abbas et al., 2019). Effective communication cannot occur without proper team-building (Hassan et al., 2022), which is consistent with the results of (Pollack and Matous, 2019). This is why this skill ranked third in importance compared to the others. Clarifying expectations skill is among the most important owing to its critical role in addressing ambiguity (Matin et al., 2010, Nogueira, 2022) and enhancing site safety (Mohammadi et al., 2018, Aldossary and Bubshait, 2022).

This study presents several useful insights. First, the network of relationships Table 5 presented in this report highlighted the prevailing communication style used in construction projects (Gamil and Abd Rahman, 2021). It showed that there are significant differences between cultures that prioritize individualism (the United States and English-speaking countries) compared to collectivistic cultures (Asian countries), as well as between low-context (LC) and high-context (HC) communication cultures. Therefore, the results obtained for the network align with that of

(Balakrishnan, 2022) in that LC and HC communications are ubiquitous in individualistic and collectivistic cultures. However, the findings contradict the findings of Hosseini et al. (2018), who argued that national culture has no bearing on the effectiveness of construction project teams or how team members collaborate.

The CQ indicators is shown in Table. 1 dominates the literature. Noticeably absent are IPSs, despite their recognized importance. This study augments this deficiency by proposing the developed XGB, which predicts and quantifies the relationship between IPSs indicators and the CQ, while also considering the effects of concurrent or co-existed indicators that may positively or negatively impact each other. Such considerations were achieved by using a developed machine-learning algorithm.

This study can also lay the foundation for use in Social Network Analysis (SNA) when investigating networks of communication as weighted networks. This is significant as construction projects are now regarded as network-based organizations (Castillo et al., 2023). After forecasting the quality of IC via XGB, the links between workers will carry different weights. Such weightings distinguish links from points of resistance, intensity, or capacity. The weight of links can make a significant difference when using dependent network metrics, such as the most direct path between workers. Although other studies have also analyzed communication networks, they used weights derived from other indicators (Pryke et al., 2018, Trach and Bushuyev, 2020) or relied on the total number of links sent by one participant to another (Jafari et al., 2020). To date, none of these studies have considered weights based on IPSs. It is this insight that adds to the significance of this study.

5.2. Practical implications

In terms of practical implications, (Mignone et al., 2016, Mani et al., 2022) showed that most leaders intervene immediately once disagreements and differences of opinions result in interpersonal conflict. Leaders, however, are expected to predict such issues and intervene as early as possible to avoid a loss of productivity. The developed XGB, with its predictive nature, provides a solution. Firstly, it evaluates the CQ, given different levels of CQ, which is consistent with the work by Pryke et al. (2018). The investigator of this study examined the CQ using three levels (low, medium, and high). Secondly, it was based on IPSs and made predictions accordingly. This lays the foundation for leaders to intervene, hopefully, before IC conflicts arise. This substantiates the claim (Ayodele et al., 2020) that early identification of interpersonal conflicts is one of two key ways of cultivating an efficient and amicable workplace. Moreover, the developed XGB is novel in terms of shifting the discourse around CQ from an approach based on conflict resolution to a predictive one that helps management to anticipate conflicts in projects in advance, which is consistent with the argument about predictive modeling proposed by Omar et al. (2019).

Job-site productivity is affected by many factors, including the IPSs of workers (Gamil and Abd Rahman, 2023). CQP shows how workers can relate to their peers effectively and productively. Understanding these abilities of workers is essential given that lower-skilled workers will reduce efficiency and productivity owing to the increase in the frequency of required communication, training, and cooperation (Guide, 2001, Karamoozian et al., 2019).

6. CONCLUSION

While the impact of IPSs on CQ is generally appreciated, no research has attempted to develop a machine-learning-based approach to predict the CQ given such skills. To fill this gap, this study aims at predicting the CQ of construction workforces based on their inherent IPSs, using a novel hybridization of a fuzzy-based algorithm and machine-learning-based technique. According to the application of the developed AI-based framework to the selected case studies, the following contributions are noted:

(1) From a theoretical perspective, the conceptual linkages between IPSs and CQ are uncovered.
(2) From a practical perspective, an accurate prediction of workers' CQ are achieved. It is observed that the selected skills with the highest importance, together with their weights, are Leadership Style (0.114), Listening (0.109), Team Building (0.096), and Clarifying Expectations (0.091).

Project managers may utilize the proposed method to assess the quality of IC between project participants. This research contributes to the field by postulating an innovative way to draw more precisely from the predictive models. Practitioners may press this method into service to obtain further insight into their projects' quality of IC. The method may also warn managers of looming interpersonal conflicts. Furthermore, social network analysts in the project management domain can now model the IC networks as a weighted graph and calculate those weights-of-links dependent metrics. Such a prediction advances workers' understanding of the quality of their communication and forewarns them about potential interpersonal conflicts before they escalate. Future research may as well be directed towards testing and using the rest of the predictive models and making real-time predictions possible using cyber-physical systems.

References

ABBAS, A., CHOI, M., SEO, J., CHA, S. H. & LI, H. 2019. Effectiveness of immersive virtual reality-based communication for construction projects. *KSCE journal of civil engineering,* 23**,** 4972-4983.

AFFARE, M. A. W. 2012. *An assessment of project communication management on construction projects in Ghana.*

ALDOSSARY, M. & BUBSHAIT, A. Factors Affecting Construction Safety Performance in Saudi Arabia. Proceedings of the Canadian Society of Civil Engineering Annual Conference 2021: CSCE21 Construction Track Volume 2, 2022. Springer, 371-381.

ARABI, M. & KHOSHNEYAT, G. R. 2022. A Qualitative and Quantitative Study of the Relationship between Interpersonal Communication Skills and Successful Teaching. *International Journal of Multicultural and Multireligious Understanding,* 9**,** 91-97.

ARMSTRONG, M. & TAYLOR, S. 2003. A Handbook of Human Resource Management Practices, Kogan Page Limited. London.

AUBERT, B., HOOPER, V. & SCHNEPEL, A. 2013. Revisiting the role of communication quality in ERP project success. *American Journal of Business*.

AYODELE, O. A., CHANG-RICHARDS, A. & GONZÁLEZ, V. 2020. Factors affecting workforce turnover in the construction sector: A systematic review. *Journal of construction engineering and management,* 146**,** 03119010.

BALAKRISHNAN, K. 2022. Influence of Cultural Dimensions on Intercultural Communication Styles: Ethnicity in a Moderating Role. *Journal of Communication, Language and Culture,* 2**,** 46-62.

BLUNCH, N. 2012. *Introduction to structural equation modeling using IBM SPSS statistics and AMOS*, Sage.

BOSCH-SIJTSEMA, P. M. & HENRIKSSON, L.-H. 2014. Managing projects with distributed and embedded knowledge through interactions. *International Journal of Project Management,* 32**,** 1432-1444.

CASTILLO, T., HERRERA, R. F. & ALARCÓN, L. F. 2023. The Quality of Small Social Networks and Their Performance in Architecture Design Offices. *Journal of Construction Engineering and Management,* 149**,** 04022162.

CHUN TIE, Y., BIRKS, M. & FRANCIS, K. 2019. Grounded theory research: A design framework for novice researchers. *SAGE open medicine,* 7**,** 2050312118822927.

DHALIWAL, S. S., NAHID, A.-A. & ABBAS, R. 2018. Effective intrusion detection system using XGBoost. *Information,* 9**,** 149.

DONG, W., HUANG, Y., LEHANE, B. & MA, G. 2020. XGBoost algorithm-based prediction of concrete electrical resistivity for structural health monitoring. *Automation in Construction,* 114**,** 103155.

ELGHAISH, F., HOSSEINI, M. R., KOCATURK, T., ARASHPOUR, M. & LEDARI, M. B. 2023. Digitalised circular construction supply chain: An integrated BIM-Blockchain solution. *Automation in Construction,* 148**,** 104746.

FORCADA, N., SERRAT, C., RODRÍGUEZ, S. & BORTOLINI, R. 2017. Communication key performance indicators for selecting construction project bidders. *Journal of Management in Engineering,* 33**,** 04017033.

FRIEDMAN, J. H. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics***,** 1189-1232.

GAMIL, Y. & ABD RAHMAN, I. 2021. Studying the relationship between causes and effects of poor communication in construction projects using PLS-SEM approach. *Journal of Facilities Management,* ahead-of-print.

GAMIL, Y. & ABD RAHMAN, I. 2023. Studying the relationship between causes and effects of poor communication in construction projects using PLS-SEM approach. *Journal of Facilities Management*, 21**,** 102-148.

GHORBANI, A. 2023. A Review of Successful Construction Project Managers' Competencies and Leadership Profile. *Journal of Rehabilitation in Civil Engineering,* 11**,** 76-95.

GUIDE, A. Project management body of knowledge (pmbok® guide). Project Management Institute, 2001. 7-8.

HASSAN, B., WAZIRI, A. Y., USMAN, H. & IBRAHIM, Y. 2022. The Influence of Construction Project Team Effectiveness in Higher Institutions' Building Projects: A Case from Nigeria. *International Journal of Real Estate Studies,* 16**,** 37-50.

HOFFMAN, R. R. 1998. How can expertise be defined? Implications of research from cognitive psychology. *Exploring expertise.* Springer.

HOSSEINI, M. R., MARTEK, I., CHILESHE, N., ZAVADSKAS, E. K. & ARASHPOUR, M. 2018. Assessing the influence of virtuality on the effectiveness of engineering project networks:"Big Five Theory" perspective. *Journal of construction engineering and management,* 144**,** 04018059.

HOSSEINI, M. R., ZAVADSKAS, E., XIA, B., CHILESHE, N. & MILLS, A. 2017. Communications in hybrid arrangements: case of Australian construction project teams. *Engineering Economics,* 28**,** 290-300.

JAFARI, P., MOHAMED, E., LEE, S. & ABOURIZK, S. 2020. Social network analysis of change management processes for communication assessment. *Automation in Construction,* 118**,** 103292.

KAPOOR, A., GULLI, A., PAL, S. & CHOLLET, F. 2022. *Deep Learning with TensorFlow and Keras: Build and deploy supervised, unsupervised, deep, and reinforcement learning models*, Packt Publishing Ltd.

KARAMOOZIAN, A., WU, D., CHEN, C. P. & LUO, C. 2019. An approach for risk prioritization in construction projects using analytic network process and decision making trial and evaluation laboratory. *IEEE Access,* 7**,** 159842-159854.

KHAN, A. A., SHAMEEM, M., NADEEM, M. & AKBAR, M. A. 2021. Agile trends in Chinese global software development industry: Fuzzy AHP based conceptual mapping. *Applied Soft Computing,* 102**,** 107090.

KUNDI, Y. M., BADAR, K., SARFRAZ, M. & ASHRAF, N. 2023. Interpersonal conflict as a barrier to task performance: the mediating role of workplace deviance and the moderating role of emotional intelligence. *International Journal of Conflict Management,* 34**,** 104-124.

KWOFIE, E. T. 2015. *Contribution of unique features of mass housing projects to project team communication performance.*

LINARDATOS, P., PAPASTEFANOPOULOS, V. & KOTSIANTIS, S. 2020. Explainable ai: A review of machine learning interpretability methods. *Entropy*, 23**,** 18.

MANI, S., AHMADI EFTEKHARI, N., HOSSEINI, M. R. & BAKHSHI, J. 2022. Sociotechnical dimensions of BIM-induced changes in stakeholder management of public and private building projects. *Construction Innovation*.

MATIN, H. Z., JANDAGHI, G., KARIMI, F. H. & HAMIDIZADEH, A. 2010. Relationship between interpersonal communication skills and organizational commitment (Case Study: Jahad Keshavarzi and University of Qom, Iran). *European Journal of Social Sciences*, 13**,** 387-398.

MIGNONE, G., HOSSEINI, M. R., CHILESHE, N. & ARASHPOUR, M. 2016. Enhancing collaboration in BIM-based construction networks through organisational discontinuity theory: a case study of the new Royal Adelaide Hospital. *Architectural Engineering and Design Management,* 12**,** 333-352.

MITHAS, S., RAMASUBBU, N. & SAMBAMURTHY, V. 2011. How information management capability influences firm performance. *MIS quarterly***,** 237-256.

MOHAMMADI, A., TAVAKOLAN, M. & KHOSRAVI, Y. 2018. Factors influencing safety performance on construction projects: A review. *Safety science,* 109**,** 382-397.
MOHANDES, S. R., KINEBER, A. F., ABDELKHALEK, S., KADDOURA, K., ELSAYED, M., HOSSEINI, M. R. & ZAYED, T. 2022. Evaluation of the critical factors causing sewer overflows through modeling of structural equations and system dynamics. *Journal of Cleaner Production,* 375**,** 134035.
MOHR, J. J. & SOHI, R. S. 1995. Communication flows in distribution channels: Impact on assessments of communication quality and satisfaction. *Journal of retailing,* 71**,** 393-415.
MÖLLER, A., RUHLMANN-KLEIDER, V., LELOUP, C., NEVEU, J., PALANQUE-DELABROUILLE, N., RICH, J., CARLBERG, R., LIDMAN, C. & PRITCHET, C. 2016. Photometric classification of type Ia supernovae in the SuperNova Legacy Survey with supervised learning. *Journal of Cosmology and Astroparticle Physics,* 2016**,** 008.
NAM, C. S., LYONS, J. B., HWANG, H.-S. & KIM, S. 2009. The process of team communication in multi-cultural contexts: An empirical study using Bales' interaction process analysis (IPA). *International Journal of Industrial Ergonomics,* 39**,** 771-782.
NOGUEIRA, M. 2022. Interpersonal Communication in the workplace in the perspective of People with Disability co-workers. *International journal of Engineering, Business and Management,* 6.
OCHIENG, E. G. & PRICE, A. D. 2010. Managing cross-cultural communication in multicultural construction project teams: The case of Kenya and UK. *International Journal of Project Management,* 28**,** 449-460.
OMAR, N. S., HATEM, W. A. & NAJY, H. I. 2019. Predictive modeling for developing maintenance management in construction projects. *Civil Engineering Journal,* 5**,** 892-900.
PAMIDIMUKKALA, A., KERMANSHACHI, S. & KAMALI RAD, S. 2023. Ranking and Weighting Effective Project-Based Communication Indicators for Primary and Secondary Stakeholders in Construction Projects. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction,* 15**,** 05022006.
POLLACK, J. & MATOUS, P. 2019. Testing the impact of targeted team building on project team communication using social network analysis. *International Journal of Project Management,* 37**,** 473-484.
PRYKE, S., BADI, S., ALMADHOOB, H., SOUNDARARAJ, B. & ADDYMAN, S. 2018. Self-organizing networks in complex infrastructure projects. *Project management journal,* 49**,** 18-41.
RAHIMIAN, A., HOSSEINI, M. R., MARTEK, I., TAROUN, A., ALVANCHI, A. & ODEH, I. 2022. Predicting communication quality in construction projects: A fully-connected deep neural network approach. *Automation in Construction,* 139**,** 104268.
RITCHIE, M. J., PARKER, L. E. & KIRCHNER, J. E. 2020. From novice to expert: a qualitative study of implementation facilitation skills. *Implementation science communications,* 1**,** 1-12.
ROBBINS, S. & HUNSAKER, P. L. 2011. *Training in interpersonal skills: Tips for managing people at work*, Pearson Higher Ed.
SAATY, T. L. 1980. The analytical hierarchy process, planning, priority. *Resource allocation. RWS publications, USA.*
SAFAPOUR, E., KERMANSHACHI, S. & KAMALIRAD, S. 2020. Analysis of effective project-based communication components within primary stakeholders in construction industry. *Built Environment Project and Asset Management*.
SEGERSTEDT, A., OLOFSSON, T., XIE, C., WU, D., LUO, J. & HU, X. 2010. A case study of multi-team communications in construction design under supply chain partnering. *Supply Chain Management: An International Journal*.
SENESCU, R. R., ARANDA-MENA, G. & HAYMAKER, J. R. 2013. Relationships between project complexity and communication. *Journal of Management in Engineering,* 29**,** 183-197.
SOKOLOVA, M. & LAPALME, G. 2009. A systematic analysis of performance measures for classification tasks. *Information processing & management,* 45**,** 427-437.

SPOTO, A., NUCCI, M., PRUNETTI, E. & VICOVARO, M. 2023. Improving content validity evaluation of assessment instruments through formal content validity analysis. *Psychological Methods*.
TALUKDER, M. A., HASAN, K. F., ISLAM, M. M., UDDIN, M. A., AKHTER, A., YOUSUF, M. A., ALHARBI, F. & MONI, M. A. 2023. A dependable hybrid machine learning model for network intrusion detection. *Journal of Information Security and Applications,* 72**,** 103405.
TAMAYO, D., SILBURT, A., VALENCIA, D., MENOU, K., ALI-DIB, M., PETROVICH, C., HUANG, C. X., REIN, H., VAN LAERHOVEN, C. & PARADISE, A. 2016. A machine learns to predict the stability of tightly packed planetary systems. *The Astrophysical Journal Letters,* 832**,** L22.
TONE, K., SKITMORE, M. & WONG, J. K. W. 2009. An investigation of the impact of cross-cultural communication on the management of construction projects in Samoa. *Construction Management and Economics,* 27**,** 343-361.
TRACH, R. & BUSHUYEV, S. 2020. Analysis communication network of construction project participants. *Przegląd Naukowy Inżynieria i Kształtowanie Środowiska,* 29.
WANG, X., LOVE, P. E., KIM, M. J. & WANG, W. 2014. Mutual awareness in collaborative design: An Augmented Reality integrated telepresence system. *Computers in Industry,* 65**,** 314-324.
WESTIN, S. & SEIN, M. K. 2014. Improving data quality in construction engineering projects: An action design research approach. *Journal of Management in Engineering,* 30**,** 05014003.
WU, S. 2013. A review on coarse warranty data and analysis. *Reliability Engineering & System Safety,* 114**,** 1-11.
YOLANDA, A., NURISMILIDA, K. W. & WULANSARY, D. 2021. The Effect of Miscommunication on the Quality of Employee Performance.
ZHANG, J. Z., SRIVASTAVA, P. R., SHARMA, D. & EACHEMPATI, P. 2021. Big data analytics and machine learning: A retrospective overview and bibliometric analysis. *Expert Systems with Applications,* 184**,** 115561.
ZHU, W., ZENG, N. & WANG, N. 2010. Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. *NESUG proceedings: health care and life sciences, Baltimore, Maryland,* 19**,** 67.

Appendix A: The designed questionniare used for gathering the required data

Self-assessment				*Other Colleagues Assessment*				*To What Extent Do You Feel that Your Communication with Each Other Is:*
Leadership Style	*Listening*	*Team Building*	*Clarifying Expectations*	*Leadership Style*	*Listening*	*Team Building*	*Clarifying Expectations*	
Expert	*Expert*	*Expert*	*Expert*	*Expert*	*Expert*	*Expert*	*Expert*	*Very High*
Journeyman	*Journeyman*	*Journeyman*	*Journeyman*	*Journeyman*	*Journeyman*	*Journeyman*	*Journeyman*	*High*
Apprentice	*Apprentice*	*Apprentice*	*Apprentice*	*Apprentice*	*Apprentice*	*Apprentice*	*Apprentice*	*Medium*
Initiate	*Initiate*	*Initiate*	*Initiate*	*Initiate*	*Initiate*	*Initiate*	*Initiate*	*Low*
Novice	*Novice*	*Novice*	*Novice*	*Novice*	*Novice*	*Novice*	*Novice*	*Very Low*

Figure 5. Questionnaire used for gathering the data needed to train the predictive algorithm

Appendix B: A bunch of data collected through the designed questionnaire

Table 6. A bunch of data collected through the designed questionnaire

Self-assessment				Other Colleagues Assessment				Communication Quality
Leadership Style	Listening	Team Building	Clarifying Expectations	Leadership Style	Listening	Team Building	Clarifying Expectations	
4	3	4	3	3	4	4	4	3
5	5	5	5	4	4	5	4	3
3	4	4	4	4	3	2	2	1
3	5	4	3	2	3	2	3	1
4	5	5	5	3	4	4	4	3
4	5	4	5	3	3	3	3	3
4	4	5	5	5	3	3	4	4
3	4	3	4	3	4	3	3	3
3	2	3	2	3	2	2	2	1
4	4	3	3	2	2	2	3	1
4	4	5	5	3	4	3	3	3
4	4	3	3	4	3	3	4	3
4	3	4	4	3	3	3	3	2
4	5	5	4	4	4	4	4	4
4	5	3	3	3	4	3	3	4
3	4	3	4	4	4	3	3	2
4	5	4	4	4	2	4	3	3
4	3	4	4	4	3	3	4	3
4	4	4	5	3	2	2	1	0
4	5	5	4	2	2	2	3	2

Appendix C: The proficiency scale used in the survey

Table 7. The proficiency scale used in the survey (*Hoffman, 1998*, *Ritchie et al., 2020*)

Proficiency level	Definition
Expert	The distinguished or brilliant journeyman, highly regarded by colleagues, whose judgments are remarkably accurate and reliable, whose performance indicates consummate skill, and who can deal optimally with certain kinds of rare or "tough" cases. Also, an expert is one who has marked skills or knowledge derived from comprehensive experience with subdomains.

Journeyman	Literally, a person who can perform a day's labor with no supervision, while following orders. A well-experienced and reliable worker, or one who has a level of competence. For all high levels of motivation, it is possible to remain at this proficiency level for life.
Apprentice	Literally, one who is learning. Traditionally, the apprentice is immersed in the domain by assisting someone at a higher level. The length of an apprenticeship hinges upon the domain, ranging from one to 12 years.
Initiate	Literally, a novice who has begun introductory instruction.
Novice	Literally, a probationary person. There has been some minimal exposure to the domain.

Appendix D: Details of each construction case study

Table 8. Details of each case study

Projects	Number of project participants	Number of communication-quality predicted	Observation hours per day (hrs/day)	Size	Type of contract	Sector	Description
Illustrative example I	32	53	Approximately 5-10 hrs/day	Large-sized licensed	Design-bid-build	Building	Airport development project
Illustrative example II	41	66	Approximately 3-12 hrs/day	Large-sized licensed	Design-bid-build	Building	Construction of residential buildings
Illustrative example III	23	64	Approximately 3-15 hrs/day	Large-sized licensed	Design-bid-build	Building	Construction of residential buildings