

MACHINE LEARNING - BASED FRAMEWORK FOR CONSTRUCTION DELAY MITIGATION

SUBMITTED: April 2020
REVISED: March 2021
PUBLISHED: July 2021
EDITOR: Robert Amor
DOI: [10.36680/j.itcon.2021.017](https://doi.org/10.36680/j.itcon.2021.017)

*Muizz O. Sanni-Anibire, Mr.,
Dammam Community College, King Fahd University of Petroleum and Minerals, Kingdom of Saudi Arabia;
muizzanibire@kfupm.edu.sa*

*Rosli M. Zin, Prof.,
School of Civil Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, Johor, 81310, Malaysia;
roslizin@utm.my*

*Sunday O. Olatunji, Assoc. Prof.,
Department of Computer Science, College of Computer Science and Information Technology,
Imam Abdulrahman Bin Faisal University, Dammam, 31441, Kingdom of Saudi Arabia;
osunday@iau.edu.sa*

SUMMARY: *The construction industry, for many decades, has been underperforming in terms of the success of project delivery. Construction delays have become typical of many construction projects leading to lawsuits, project termination, and ultimately dissatisfied stakeholders. Experts have highlighted the lack of adoption of modern technologies as a cause of underproductivity. Nevertheless, the construction industry has an opportunity to tackle many of its woes through Construction 4.0, driven by enabling digital technologies such as machine learning. Consequently, this paper describes a framework based on the application of machine learning for delay mitigation in construction projects. The key areas identified for machine learning application include "cost estimation", "duration estimation", and "delay risk assessment". The developed framework is based on the CRISP-DM graphical framework. Relevant data were obtained to implement the framework in the three key areas identified, and satisfactory results were obtained. The machine learning methods considered include Multi Linear Regression Analysis, K-Nearest Neighbours, Artificial Neural Networks, Support Vector Machines, and Ensemble methods. Finally, interviews with professional experts were carried out to validate the developed framework in terms of its applicability, appropriateness, practicality, and reliability. The main contribution of this research is in its conceptualization and validation of a framework as a problem-solving strategy to mitigate construction delays. The study emphasized the cross-disciplinary campaign of the modern construction industry and the potential of machine learning in solving construction problems.*

KEYWORDS: *machine learning, construction delays, construction 4.0, delay mitigation, risk assessment, cost estimation, duration estimation*

REFERENCE: *Muizz O. Sanni-Anibire, Rosli M. Zin, Sunday O. Olatunji (2021). Machine learning - based framework for construction delay mitigation. Journal of Information Technology in Construction (ITcon), Vol. 26, pg. 303-318, DOI: 10.36680/j.itcon.2021.017*

COPYRIGHT: © 2021 The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



1. INTRODUCTION

For almost five decades, the construction industry has been overcome by inefficiency and under productivity. Other industries such as the service and manufacturing industries have increased productivity by adopting digital technologies and automation across their value chain. Forbes and Ahmed (2010) report that up to 30% of wastage is due to inefficiency, errors, delays, and poor communication. Unfortunately, the majority of construction work is still based on century-old methods, which do not relieve the construction industry of its problems. The problem of underperformance in construction is even more aggravated in light of large and ambitious construction projects which are characteristic of the industry in the 21st century. Assaf et al., (1995) hinted that a critical issue with large construction projects is that of delays. Aibinu and Jagboro (2002) define delays as situations where a project's duration is extended due to factors related to the client, consultant, and contractor. Delays in construction projects have negative effects on all stakeholders including litigation (Santoso and Soeng, 2016), cost and time overruns, loss of productivity and revenue, and contract termination (Sambasivan and Soon, 2007).

Conventionally, construction research has focused on exploring the causes of construction delays in various economies globally (Sanni-Anibire et al., 2020a). Such exploratory research approaches have been subject to severe criticism. For instance, AlSehaimi et al., (2013) argued for constructive research approaches in the construction industry in general, and delay mitigation in particular. Accordingly, some researchers have proposed frameworks to mitigate the prevalence of delays in the industry. Examples include a conceptual framework based on Case-Based Reasoning (CBR) by Ng et al., (2000); a model based on knowledge management by Abdul-Rahman (2008); the adoption of the Lean Planner System by AlSehaimi (2011); a risk response model by Motaleb (2014); and a management framework by Khair et al., (2018). These studies have viewed delay mitigation as a means of providing preventative measures, recommendations derived from lessons learned, and modified project management frameworks. Despite the contribution these studies have made to the research and professional community, the occurrence of delays is still a common feature of projects. It could be argued that these studies do not treat the issue of delay mitigation from a holistic perspective by considering the cost estimate, schedule, and project risks (Galway, 2004; Meyer, 2015). Moreover, none of these studies has sought to leverage the growth of industrial data which has the potential to transform the construction industry. Success stories on the use of modern digital technologies to resolve performance issues have been reported in other industries (Larrañaga et al., 2018). Interestingly, Woodley (2019) suggests that digitalization has the potential to resolve claims and disputes due to construction delays. Regrettably, the construction industry is notorious for its slow adoption of modern technology; nonetheless, a technological transformation driven by the fourth industrial revolution (IR4.0) is now witnessed in the industry.

The current trend of the construction industry is Construction 4.0 – aimed at digitalization and automation in the industry for improved productivity. Machine learning—a subset of artificial intelligence—is considered one of the top ten technologies driving IR4.0 (PricewaterhouseCoopers, 2017). It is worth mentioning that machine learning has been applied to various domains of construction and civil engineering (Adeli, 2020); however, only a few studies are directly related to delay mitigation in construction. Promisingly, there is a current interest in applying machine learning to delay risk prediction in construction (Gondia et al. 2019, El-Kholy 2019), though such studies already predominate in other industries (Yaghini et al. 2013, Kim et al. 2016, Choi et al. 2016, Takeichi et al. 2017). While the prediction of delay risk is valuable towards the central objective of mitigating delays, it is crucial to mention that there exists a gamut of factors that influence construction delay. Aziz and Abdel-Hakam (2016) identified two hundred and ninety-three (293) construction delay factors, while Enshassi et al., (2009) presented 110 delay factors. Therefore, there could be a wide range of possibilities in identifying relevant areas of machine learning application in delay mitigation.

In light of the foregoing, the goal of this study is to develop a framework based on machine learning to mitigate delays in construction projects. According to the Project Management Institute, project risk mitigation should holistically cover three areas of cost estimate, schedule, and identified project risks (Galway, 2004; Meyer, 2015). Thus, the researchers proposed three concepts: (1) delay mitigation by accurate estimation of cost (Abdul Rahman et al. 2006, Olawale and Sun 2010, Love et al. 2000, Smart Market Report 2011, El-Kholy 2019); (2) delay mitigation by accurate estimation of duration (Ng 2007, Abdul Rahman et al. 2006, Love et al. 2000, El-Kholy 2019); and (3) delay mitigation through delay risk management (Ng,2007, Abdul Rahman et al. 2006, Abedi et al. 2011, Olawale and Sun 2010, Gondia et al. 2020, Motaleb 2020). The researchers thus theorized that these three areas could serve as control points in the planning stage of construction projects to mitigate the occurrence of

delays, and ultimately, ensure the success of the project. Abdul Rahman et al. (2006) state “it is important to predict and identify the problems in the early stages of construction and diagnose the cause to find and implement the most appropriate and economical solutions”.

Thus, to demonstrate the framework, relevant data was obtained and machine learning models were developed in the three areas of cost, duration, and delay risk. Subsequently, the results, as well as a graphical representation of the framework, were presented to industry experts for validation. The conclusion from the study was the proposal of a framework based on machine learning for delay mitigation in construction. The framework was developed as an adapted format of the most prevalent analytics model: The Cross-Industry Standard Process for Data Mining (CRISP-DM) (Wirth and Hipp, 2000). The ensuing sections discuss productivity in the construction industry, Construction 4.0 and machine learning, a summary of other frameworks developed by researchers for delay mitigation, the methodology of the study, findings, discussions, and conclusions.

2. LITERATURE REVIEW

The following sections highlight the initiative of various governments to increase productivity in their respective construction industries. Industry 4.0 as a new paradigm in the construction industry, as well as the role of machine learning in this paradigm, is also discussed. Finally, a descriptive summary of previous efforts to develop frameworks to mitigate construction delay is presented.

2.1 Productivity in the construction industry

Construction is crucial to a country’s industrialization, and thus, it occupies the center stage in a nation's socio-economic development. The worth of the construction industry is estimated to be about \$8.7 trillion, accounting for 12.2% of the world's economic output and providing employment for about 200 million people worldwide (Zou and Sunindijo, 2015). Hence, problems in the construction industry have far-reaching consequences to a nation's socio-economic structure. It is therefore reasonable that stakeholders are interested in the construction industry's productivity, which is lagging behind that of other industries. In the UK, a study by Sir Michael Latham (1994) titled "Constructing the Team" provided a 30-point executive summary of strategies that could be used to enhance the productivity of the construction industry (Latham, 1994). Another study titled "Rethinking Construction" was made by Sir John Egan (Egan, 1998). A more recent study by Mark Farmer (2016) commonly known as the "Farmer Review" or by its subtitle "Modernise or Die" identified key deficiencies of the British Construction Industry. Farmer (2016) recommended the adoption of modern techniques including greater use of robotics, machine learning, and automated planning decisions by use of digital design. Other countries have undertaken similar studies and developed action points as well as key performance indicators for the development of the construction industry (Sawhney et al., 2020). The continuous proliferation of such studies may suggest that the industry has not made significant progress in its battle for enhanced productivity.

Notably, few countries like China, Singapore, and Japan have made giant strides in the past few decades to adopt technologies such as prefabrication (Xu et al. 2020, Gao et al. 2020). However, on a global scale, construction productivity is largely affected by the industry's slow adoption of modern technology. Farmer (2016)'s report states, "Construction has not even made the transition to 'Industry 3.0' status which is predicated on large scale use of electronics and IT to automate production". Unlike the manufacturing, automotive, and aerospace industries, the construction industry has failed to embrace the opportunities afforded by technology and advances in data management. Modern digital technologies offer unique opportunities in construction project management through automated means of capturing, storing, and processing large quantities of data for effective decision-making (Forbes and Ahmed, 2010). Positively, the industry has embraced a new era of faster, more automated, and smarter construction processes (Keith, 2018).

2.2 Industry 4.0 and machine learning

The production world, to date, has witnessed four industrial revolutions. The first industrial revolution relied on water and steam, the second on electricity, the third on electronic systems and information technology. The fourth industrial revolution is a convergence of the digital and physical worlds. It merges the third industrial revolution's IT, such as computer integrated manufacturing, machine learning, the internet, and many other technologies, with operational technologies to create the disruptive technologies that are the backbone of the IR4.0 (Larrañaga et al., 2018). Industry 4.0 was defined by Kagermann et al. (2013) as an initiative to secure the future of the German



manufacturing industry. Although, a German concept, it has been adopted globally by governments and organizations. Global examples include Industrie du Futur in France, Industria Conectada in Spain, Made in China 2025, Made in India, ASEAN 4.0 with manufacturing leaders such as Singapore and Malaysia, and Society 5.0 in Japan (Larrañaga et al., 2018).

In light of the fourth industrial revolution, the construction industry is presented with the opportunity to propel to more efficient production, business models, and value chains. Such a transformation could be achieved through existing and emerging technologies that form part of the IR4.0 paradigm (Oestereich and Teuteberg, 2016). The construction industry's response to Industry 4.0 is Construction 4.0. The idea of Construction 4.0 is based on a confluence of trends and technologies (both digital and physical) that promise to reshape the way built-environment assets are designed and constructed. Liao et al. (2017) presented a systematic literature review of 224 papers published until June 2016 to determine the technologies and key features of IR4.0. Notably, machine learning was identified as one of the enabling digital technologies to drive IR4.0. Machine learning is one of the key techniques able to generate data-driven predictive models that can be used for decision-making. Cost management, scheduling, and construction management are applications that can be developed using digital technologies as an inherent concept of Construction 4.0 (Sawhney et al., 2020).

2.3 Frameworks for construction delay mitigation

Construction delay is one of the most significant problems in the construction industry, and it is to this end that numerous studies have been made to explore the main causes of construction delays in various countries (Sanni-Anibire et al., 2020a). AlSehaimi and Koskela (2008) opine that the failure of existing delay studies is attributed to their descriptive and exploratory nature, and thus suggested the need for alternative research approaches. The following is a descriptive summary of research work that has been achieved towards that aim.

Love et al. (2000) proposed a systems dynamics model for delay mitigation due to prolonged overtime work on project costs and quality. The study suggested that utility theory can be applied to determine the most appropriate solutions to mitigate project delays. The study validated its procedure with 14 projects in Hong Kong. Ng et al. (2000) investigated the application of Case-Based Reasoning to develop a conceptual framework for delay mitigation. The framework consists of various components, including "delay identification", "delay analysis", "crashing activity scrutiny", "estimate", "schedule re-estimation", "data input" and "output". Abdul-Rahman (2008) proposed a delay mitigation model based on the adoption of knowledge management. The model was based on knowledge obtained through lessons learned in construction projects and noted that the accuracy and volume of information obtained are potential limitations of the model. AlSehaimi (2011) theorized that traditional project management tools and practices are inadequate in modern construction. Consequently, a proposal to adopt the Last Planner System as a viable solution was made. A risk response model was developed by Motaleb (2014). The study described preventative measures and mitigation measures in developing the model, and further validated the same through interviews with professionals from selected case studies. Similarly, Chai et al. (2015) presented a structural equation model based on preventive measures, predictive measures, organizational measures, and corrective measures for delay mitigation. Khair et al. (2018) proposed a project management framework based on the 'stage-gate' approach and validated the same with input from focused group discussions with experts. Recently, Gunduz and Al-Naimi (2021) proposed a delay mitigation framework through the identification of 41 delay mitigation factors categorized into financial and enabler objectives. The main contribution of the study was the integration of the balanced scorecard approach and quality function deployment in developing the framework.

Despite the acknowledged contribution of these studies, construction delays remain unmitigated in various construction sectors globally (Sanni-Anibire et al., 2020a). Simplistic frameworks which do not consider the complex interplay among multiple delay factors have been perceived as ineffective (Woodley, 2019). Effective strategies to delay mitigation should cover three main aspects including the project's timeframe, estimated costs, and performance in terms of identified risks (Galway, 2004; Meyer, 2015). These three domains are also recognized by other researchers (Abdul Rahman et al. 2006, El-Kholy 2019; Gondia et al. 2020). Furthermore, adopting digital technologies has the potential to mitigate most delay factors through improved situational awareness and information insights derived from data (Woodley, 2019). The current study hopes to fulfil the research need for a digital and holistic delay mitigation framework.

3. METHODOLOGY

The methodology employed in this research can be summarily described as follows:

3.1 Identify areas of ML application for construction delay mitigation

A thorough review of the extant literature was made to identify the most common delay risk factors in the construction industry. This exercise led to the identification of 36 delay risk factors extracted from relevant studies published in the last 15 years (Abd El-Razek 2008, Sambasivan and Soon 2007, Aibinu and Odeyinka 2006, Doloi et al. 2012, Faridi and El-Sayegh 2006, Fugar and Agyakwah-Baah 2010, Gündüz et al. 2013, Lo et al. 2006, Sweis et al. 2008, Toor and Ogunlana 2008, Enshassi et al. 2009). Furthermore, the research defined three key areas for ML application based on relevant literature, including: “accurate estimation of cost”; “accurate estimation of duration”; and “delay risk assessment” as previously discussed.

3.2 Develop conceptual delay mitigation framework

Conceptual frameworks are representations of ideas as understood by the researcher. Miles and Huberman (1984) define a conceptual framework as "the current version of the researcher's map of the territory being investigated". Likewise, Weaver-Hart (1988) views a conceptual framework as "a structure for organizing and supporting ideas; a mechanism for systematically arranging abstractions; sometimes revolutionary or original, and usually rigid". Thus, the areas identified for ML application in delay mitigation were converged to develop the conceptual framework. The Cross-Industry Standard Process for Data Mining (CRISP-DM) was adapted as the vehicle to convey the graphical format of the framework. The CRISP-DM is a standard process model for data mining which was developed in a project partly sponsored by the European Commission. The value of the model is in its effectiveness, reliability, and adaptability to various situations regardless of the industry sector and technology used (Wirth and Hipp, 2000). Though other data mining frameworks have been developed over the years (Shafique and Qaiser, 2014), CRISP-DM is widely adopted across various industries. For instance, the CRISP-DM has been used to develop machine learning frameworks for quality management (Schäfer et al., 2018), and sports management (Bunker and Thabtah, 2019; Schelling and Robertson, 2020).

3.3 Implement delay mitigation framework

An implementation of the delay mitigation framework was carried out based on data obtained from past case studies. Data (see table 1 for statistical summary and description) on the impact and likelihood of 36 identified delay risk factors were obtained from 48 industry professionals in skyscraper projects. These included professionals across the life cycle of the project, where the contractors represent 30% of the population, while the consultants represent 35%, and similarly the clients' representatives/facility managers. Likewise, at the time of the survey, 23% of the respondents were designated as project managers while 21% were holding facility manager roles, 15% director roles and 11% executive director roles. In terms of professional experience, 48% which represents the majority of the respondents had greater than 15 years of experience. UAE and Saudi Arabia, with 42% and 33%, represents the majority in terms of the location of the respondents. The obtained values for impact and likelihood were used to compute risk classes for each response and then presented in a suitable data structure as shown in table 4 (Sanni-Anibire et al., 2020b). Relevant cost and duration data (see tables 2 and 3 for statistical summary and description) were obtained for skyscrapers from the Mega Project Case Study Centre of China (http://www.mpcsc.org/case_search.htm). Thirty-five projects with information on the project duration and other relevant data were selected, and there were 5 missing values for the project cost as shown in table 2. Hence, 35 projects were adopted for the duration, while 30 projects were adopted for the cost.

Subsequently, machine learning algorithms were implemented on the dataset and evaluated based on standard performance metrics including the Root Mean Squared Error (RMSE), Correlation Coefficient (R^2), Mean Absolute Percentage Error (MAPE), the Classification Accuracy and Misclassification Error. The Waikato Environment for Knowledge Analysis (WEKA 3.8.3)-one of the most popular tools used in the machine learning community, has been used in this study (Larrañaga et al., 2018). In this study, four ML algorithms have been considered including Multi Linear Regression Analysis (MLRA), Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN), and Support Vector Machines (SVM). Detailed theoretical and mathematical descriptions of these algorithms are presented in relevant references (Cortes and Vapnik 1995, Lek and Guégan 1999, Olatunji et al. 2013, Wauters and Vanhoucke 2017, Sethi et al. 2017, Olatunji 2017).



TABLE. 1: Descriptive statistics of the delay risk factors (DRFs) from questionnaire responses.

S/N	Causes of delay	Likelihood		Impact	
		Mean	Standard Deviation	Mean	Standard Deviation
Causes related to material					
1	Shortage in construction materials/unforeseen material damages	2.96	1.18	3.57	1.06
2	Slow delivery of materials	3.23	0.99	3.77	0.94
3	Waiting for approval of shop drawings and material samples	3.28	1.16	3.72	1.06
Causes related to manpower					
4	Shortage in manpower (skilled, semi-skilled, unskilled labor)	3.28	1.19	3.89	0.98
5	Poor labor productivity	3.5	0.96	3.89	0.92
6	Labor disputes and strikes	2.47	1.25	3.35	1.39
Causes related to equipment					
7	Poor equipment productivity (breakdown/maintenance problem)	2.79	1.02	3.57	1.08
8	Shortage in equipment	2.84	1.15	3.63	0.99
Causes related to contractual relations					
9	Inappropriate construction/contractual management/ construction methods	3.51	1.1	4.19	0.79
10	Slowness in decision making	3.45	1.12	3.98	1.04
11	Delay in mobilization	3.34	1.15	3.59	1.02
12	Excessive bureaucracy/interference by the owner	3.62	0.95	3.8	1.07
13	Delay in approval of completed work	3.38	1.05	3.82	0.89
14	Delay in sub-contractor's work	3.71	0.79	4.05	0.65
Causes related to government					
15	Slow permits from municipality/government	3.79	1.13	4.07	0.98
16	Government regulations	3.31	1.13	3.67	1.08
Causes related to financing					
17	Contractor's financial difficulties	3.92	1.01	4.30	0.84
18	Client's cash flow problems/Delays in contractor's payment	4.1	0.88	4.44	0.69
19	Price escalation/fluctuations	3.29	0.98	3.56	0.89
Causes related to environmental factors					
20	Weather condition	2.71	1.13	2.96	1.03
21	Civil disturbances/Hostile political conditions	2.15	1.13	2.87	1.13
Causes related to changes					
22	Design errors/incomplete made by designers (Architects and structural drawing)	3.48	1.24	3.98	0.95
23	Design variations/change orders/increase in scope of work	3.85	0.89	4.09	0.69
24	Errors committed due to lack of experience	3.42	1.16	3.96	0.94
25	Unexpected foundation conditions encountered in the field	2.79	1.15	3.5	1.05
26	Changes in materials types and specifications during construction	3.1	1.02	3.67	0.82
27	Inaccurate site/soil investigation	2.96	1.16	3.52	1.07
28	Frequent change of sub-contractor	3.1	0.95	3.61	0.93
Causes related to scheduling and controlling techniques					
29	Poor site organization and coordination between various parties	3.75	1.19	4.24	0.89
30	Poor planning of resources and duration estimation/scheduling	3.61	1.18	3.93	1.01
31	Inadequate supervision, inspection and testing procedures	3.4	1.16	3.73	0.99
32	Accidents during construction/lack of safety measures	3.06	1.09	3.67	1.15
33	Poor communication/documentation and detailed procedures	3.13	0.97	3.59	0.87
34	Unrealistic time schedule imposed in contract	3.53	0.95	3.69	1.16
35	Poor qualification of the contractor or consultant	3.78	0.87	4.18	0.72
36	Architects'/structural engineers' late issuance of instruction	3.21	0.91	3.62	0.81

TABLE 2: Descriptive statistics of cost and duration dataset.

Features	Mean	Standard deviation	Maximum	Minimum	Missing values
GDP (bill USD)	302.91	108.22	446.31	80.77	0
# of elevators	50.66	31.31	130	6	6
Building area (m ²)	289364.06	152174.12	602401	91600	1
Floor area (m ²)	30569.68	45035.52	197000	4126	6
Height to tip (m ²)	386.18	113.64	636	237.5	0
# of floors above GF	76.97	23.22	128	37	0
Height of occupied floors (m)	339.30	115.88	610	213.9	1
# of total floors	80.91	23.49	133	39	0
# of basement floors	6.31	5.204	30	2	6
# of parking spaces	1058.32	619.86	2702	128	7
Cost (bill Yuan)	8.34	8.53	30	0.38	5
Duration (days)	1783	682.16	4555	730	0

TABLE 3: Description of the non-numeric features of cost and duration dataset.

Features	Description	Conversion to dummy variables	Missing values
Facility type	O/Office, BOH/Business, office, hotel, ROH/Residential, office, hotel, BO/Business, office, BOR/Business, office, residential	O = 1; BOH = 2; ROH = 3; BO = 4; BOR = 5	0
Structural form	T-T/Tube in Tube, D/Diagrid, C-T/Core-Tube, T/Tubular	T-T = 1; D = 2; C-T = 3; T = 4	9
Structural material	RC/ Reinforced concrete, RCS/Reinforced concrete and steel, S/Steel C/Composite	RC = 1; RCS = 2; S = 3; C = 4	0
Commencement period	Summer, autumn, winter and spring	Summer = 1; Autumn = 2; Winter = 3; Spring = 4	0

TABLE 4: Compiled data set structure for delay risk.

Project Case	DRF 1	DRF 2	DRF 3	-	-	-	DRF 4	DRF 5	DRF 6	Risk Class
Case 1	Moderate	Low	Very Low	-	-	-	Low	Low	Very Low	High
Case 2	Low	Moderate	High	-	-	-	Very Low	Moderate	High	Moderate
Case 3	Very Low	High	Moderate	-	-	-	High	High	Moderate	Very High
-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-
Case 47	High	Moderate	Very Low	-	-	-	High	Very Low	Low	Moderate
Case 48	Moderate	Low	Very Low	-	-	-	Low	Very Low	High	Very High

3.4 Validation of the framework with construction professionals

The validation approach utilized in this study was based on structured interviews of professional experts on the applicability, appropriateness, practicability, and reliability of the framework. Applicability means “the framework is suitable to the context of the project”, appropriateness means “the framework is clear, and easy to follow”, “practicability means “the framework is technically sound and correct”, and reliability means “the results of the framework is dependable” (Bassioni et al., 2005). A total of 10 highly qualified professionals (see table 5) were contacted for the validation interview. This was based on the recommendation of 5-10 interviewees as an adequate number of participants for the framework validation process (Bryman and Bell, 2003; Gao and Low, 2014). The professionals were initially briefed on the aim of the interview, and then a validation document describing the framework in figure 1 was discussed with them. This also included the performance of the ML models as shown in figures 2 and 3, and table 8. Subsequently, the professionals were requested to give their feedback on the validity of such a framework to their professional practice. A scoring sheet with a Likert scale of validation from (1) to (5) as presented in table 6 was provided to the professionals to validate the framework based on relevant criteria.

TABLE 5: Details of professionals consulted for the validation process.

S/N	Position	Education and professional certifications	Expertise	Years of experience	Location
1	Head of Project Management Office	MSc, PMP, RMP, CCP, LEED AP, Registered Engineer	Project risk management	20	United Arab Emirates
2	Managing Partner	LLM MSc FRICS	Contracts, delay, claims, and disputes	15	United Arab Emirates
3	Project Manager	PhD, FCI Arb	Delay, claims, and disputes	20	United Arab Emirates
4	Project Supervisor	PhD, PMP, LEED AP, Registered Engineer	Project management	10	Saudi Arabia
5	Senior Project Manager	PhD, PMP, CCP, Registered Engineer	Project management	15	Saudi Arabia
6	Senior Project Manager	PhD, PMP	Project management	20	Saudi Arabia
7	Project Manager	BSc, PMP	Project management	10	Saudi Arabia
8	Deputy Project Manager	BSc, PMP	Project management	10	Saudi Arabia
9	Consultant	LLM	Contract management and data analysis	5	United Arab Emirates
10	Associate Consultant	BSc	Data science	3	United Arab Emirates

MSc: Master of Science; BSc: Bachelor of Science; PhD: Doctor of Philosophy; PMP: Project Management Professional; RMP: Risk Management Professional; CCP: Cost Certified Professional; LEED AP: Leadership in Energy and Environmental Design Accredited Professional; LLM: Master of Laws; FRICS: Fellow, Royal Institution of Chartered Surveyors; FCI Arb: Fellow, Chartered Institute of Arbitrators

TABLE 6: Validation scoring sheet (modified from Bassioni et al., 2005)

Validation criteria	1	2	3	4	5
Applicability	Not applicable	←	←	←	→ Exceptionally applicable
Appropriateness	Not appropriate	←	←	←	→ Exceptionally appropriate
Practicality	Not practical	←	←	←	→ Exceptionally practical
Reliability	Not reliable	←	←	←	→ Exceptionally reliable

4. RESULTS AND FINDINGS

The delay mitigation framework proposed in this study has been developed as an adapted format of the CRISP-DM framework. The framework as presented in figure 1 is briefly described as follows:

4.1 Business understanding

This entails defining the problem from a business perspective (i.e. cost, duration, and delay risk) and obtaining relevant data. As stated earlier, the problem of delay could be mitigated through three key areas including accurate estimation of cost, accurate estimation of duration, and delay risk assessment. Thus, three models would be developed with relevant data in these domains. The dataset developed for delay mitigation has been described in stage 3 of the methodology section.

4.2 Data understanding

To understand data, statistical analysis and visualization of the dataset containing information about the problem are required. Table 1 provides descriptive statistics of the delay risk factors (DRFs) obtained from professional experts, while tables 2 and 3 provide descriptions of the cost and duration data obtained from the Mega Project Case Study Centre of China.

4.3 Data preparation

The necessary activities to construct the final dataset from the initial raw data are carried out in this phase. The study first examined the various filtering options that expose the understanding of the data structure to the machine-learning algorithm. Consequently, the *replacemissingvalues* filter was adopted when deploying the ANN and KNN algorithm for cost estimation, while the standardization filter was adopted for MLRA and SVM algorithms. In developing the model for duration estimation, the entire dataset (input features and output class) was normalized, while the untransformed data was adopted for predicting delay risk as presented in table 4. Another aspect of data pre-processing is feature subset selection. This is necessary, as data used for machine learning sometimes suffer from a phenomenon known as the “curse of dimensionality”. Thus, correlation-based feature selection was adopted for cost and duration estimation; while wrapper-based feature selection (SVM as a base classifier) was adopted for delay risk prediction. The results revealed that the most relevant feature in estimating the cost was the "floor area", while the most relevant feature in estimating the duration was the "# of total floors". The most relevant feature in delay risk prediction was "slowness in decision making".

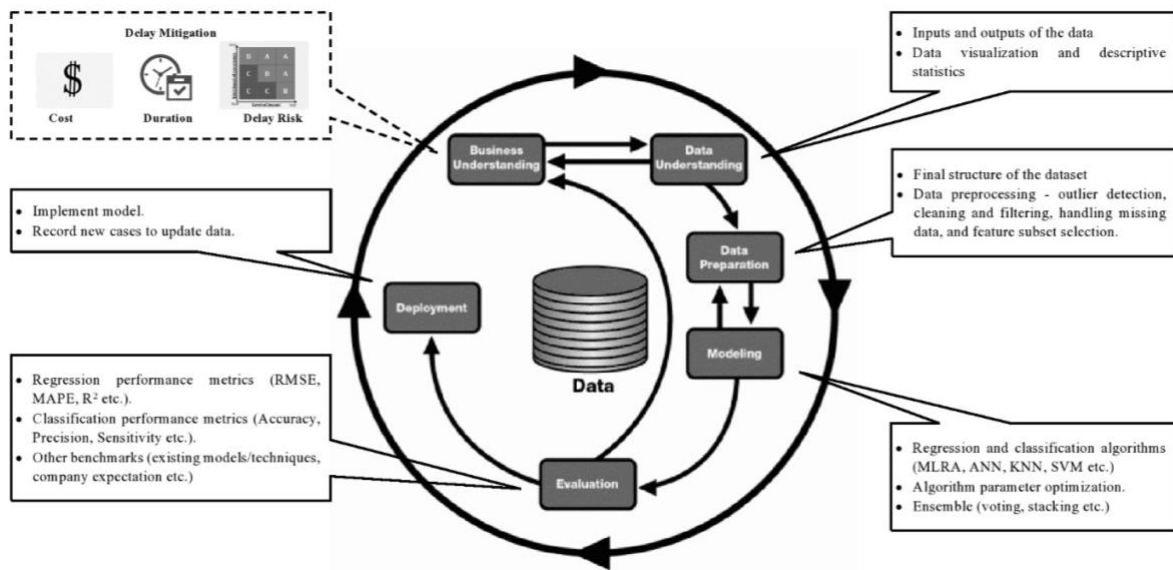


FIG. 2: Delay mitigation framework based on CRISP-DM.

4.4 Modeling

This stage involved the application of various machine-learning algorithms, as well as ensemble methods to the dataset. The dataset was split into a train-test ratio of 66% to 34%. The final model developed in predicting the duration was based on an ensemble method with ANN as the combining classifier of three base classifiers as described in table 7. In estimating the cost, an ensemble model was also developed with KNN as the combining classifier of three base classifiers as described in table 7. Likewise, the model for predicting delay risk was developed based on ANN as presented in table 7.

4.5 Evaluation

Evaluating the models developed is based on established performance metrics depending on the type of problem i.e. classification or regression. Developed models may also be evaluated based on targets set by the company or based on existing techniques. In this study, relevant performance metrics from the literature including the Root Mean Squared Error (RMSE), Correlation Coefficient (R^2), Mean Absolute Percentage Error (MAPE) were adopted for cost and duration (regression problem), while the Classification Accuracy and Misclassification Error were adopted for delay risk (classification problem). In essence, the performance of the model for the duration was characterized by an R^2 of 0.69, MAPE of 0.18, and RMSE of 301.76, while the model for predicting cost by an R^2 of 0.81, MAPE of 80.95%, and RMSE of 6.09. The results also show that the model was able to correctly classify the risk of delay, except in one case where a risk of "Very High" was erroneously classified as "Moderate". The

model thus achieved a 93.75% accuracy. Cross plots have also been made to illustrate the performance of the various models. Figures 2 and 3 show a satisfactory level of performance for the duration and cost models respectively. The data point of the actual and predicted observations was closely correlated and sometimes matching, while in some instances significant gaps were observed between the actual and predicted values. Improved predictive performance of the models could be achieved through the establishment of larger datasets.

TABLE 7: Description of machine learning models for cost, duration, and delay risk.

Application Area	Machine Learning Model		Description	
	Ensemble	Base Classifier	Selected features	Optimization hyperparameters
Duration	ANN combiner (0.3 learning rate; one hidden layer with four neurons)	ANN (<i>Multilayer Perceptron</i>)	# of total floors; # of floors above ground floors; # of parking spaces; cost	0.3 learning rate; one hidden layer with four neurons
		KNN (<i>IBk</i>)	# of total floors; # of floors above ground floors; # of parking spaces; cost; building area; height to tip; floor area; # of elevators	Nearest Neighbor: LineraNN Distance function: Manhattan Distance K: 1
		SVM (<i>SMOReg</i>)	# of total floors; # of floors above ground floors; # of parking spaces; cost	Kernel: Polykernel Cost function, C: 1 Epsilon: 1E -12 Epsilon parameter: 1E -3
Cost	KNN Nearest Neighbor: LinearNN Distance function: Euclidean Distance K: 1	MLRA (<i>LinearRegression</i>)	Floor area; structural material; height of occupied floors; height to tip	Not Applicable
		SVM (<i>SMOReg</i>)	Floor area; structural material; height of occupied floors; height to tip; # of parking spaces; building area; # of total floors; # of floors above ground floors	Kernel: Polykernel Cost function, C: 106 Epsilon: 1E -12 Epsilon parameter: 0.12
		SVM (<i>SMOReg</i>)	Floor area; structural material; height of occupied floors; height to tip	Kernel: RBFkernel Cost function, C: 3 Epsilon: 0.1 Epsilon parameter: 0.001
Delay risk	ANN (<i>Multilayer Perceptron</i>)		“slowness in decision making”; “delay in sub-contractors work”; “architects’/structural engineers’ late issuance of instruction”; and “waiting for approval of shop drawings and material samples”.	Learning rate 0.3, Momentum 0.2, Network topology: one hidden layer with 3 neurons

TABLE 8: Performance of developed machine learning models

Model	Regression performance			Classification performance	
	R ²	RMSE	MAPE	Classification Accuracy	Misclassification Error
Duration model	0.69	301.76	0.18	-	-
Cost model	0.81	6.09	80.95	-	-
Delay risk model	-	-	-	93.75	6.25

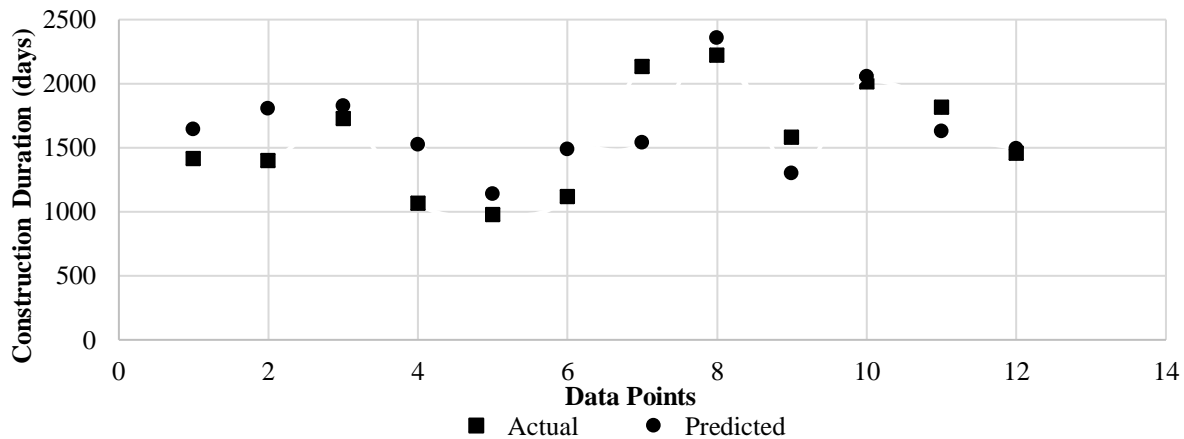


FIG. 2: Cross-plots of the actual vs. predicted values for construction duration.

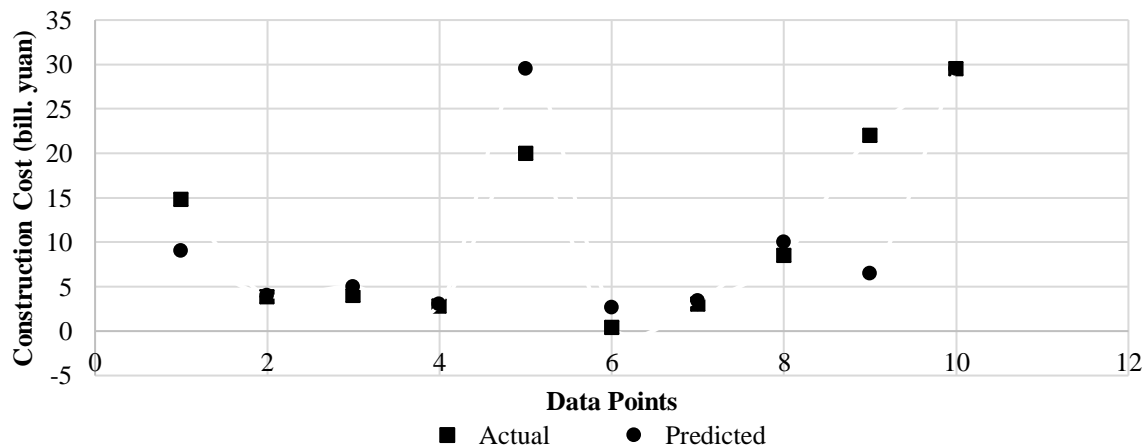


FIG. 3: Cross-plots of the actual vs. predicted values for construction cost.

4.6 Deployment

The implementation of the machine-learning model could be achieved through the development of standalone digital tools or as an addition to existing project management tools. To validate the potential for deployment by project managers, interviews with ten highly experienced professionals (described in table 5) involved in the planning and delivery of construction projects were carried out. The results of their validation of the framework are presented in table 9. Their feedback on the validity of the proposed framework can be summarised as follows:

The professionals believed in the applicability and appropriateness of the framework in mitigating construction delays. They emphasized its value in the pre-planning stage of construction projects. It can be observed from table 9 that the mean validation score for “applicability”, “appropriateness”, and “practicality” was 4.6, 4.4, and 4.1 respectively. On the other hand, the professionals were not highly confident of the “reliability” of the framework, where a mean validation score of 3.3 was achieved. Further discussions with the professionals revealed that the credibility of the dataset used to develop such models is crucial to the dependability of the decisions made due to it. The professionals opined that proper documentation of historical data from past projects needs to be confirmed for the reliability of the framework. Generally, the professionals welcomed the idea of the framework should it be

developed further as tools to be adopted in project planning. They also generally suggested that it is valuable as a baseline for comparison with existing techniques in the industry.

TABLE 9: Mean validation scores from expert interviews.

Validation criteria	Professional Interviewees										Mean
	1	2	3	4	5	6	7	8	9	10	
Applicability	4	5	4	5	5	4	5	4	5	5	4.6
Appropriateness	4	4	5	4	5	5	5	4	4	4	4.4
Practicality	3	5	4	3	5	4	4	3	5	5	4.1
Reliability	4	3	3	3	4	4	3	3	3	3	3.3

5. DISCUSSION

The construction industry continues to suffer from under productivity issues such as the occurrence of delays. Though the research domain is saturated with numerous delay studies, the majority of these studies are explorative, and hence do not provide solutions to the inherent problem. Few studies have sought to adopt constructive research methods that develop tools and models with practical and theoretical implications (Love et al. 2000; Ng et al. 2000; Abdul-Rahman 2008; AlSehaimi 2011; Motaleb 2014; Chai et al. 2015; Khair et al. 2018; Gunduz and Al-Naimi 2021). These studies have mainly viewed delay mitigation as a means of providing preventative measures or recommendations derived from lessons learned. Some studies have sought to control the estimated time based on data from previous cases, while others have proposed modified project management frameworks. Despite the inherent contribution of these studies to the body of knowledge on construction delay, there is a huge gap in research seeking the potential of computing and information technology for delay mitigation. Interestingly, the construction industry is moving towards digitization under the umbrella of Construction 4.0. Perhaps Construction 4.0 and its related technologies can transform the construction industry into a more efficient and transparent enterprise (Liao et al. 2017). A major enabling technology of Construction 4.0 is machine learning. Machine learning, although has predominated in other industries, has lacked rapid proliferation in the construction industry.

Therefore, this paper presents the development of a framework to mitigate delay in construction projects based on the application of machine learning. The specific advantage of the current study is in its adoption of modern digital technology such as machine learning to develop a framework that could serve as a tool for delay mitigation in construction. The use of historical data to develop data-driven predictive models is one of the key advantages of machine learning. Furthermore, this study promotes the concept of research triangulation, where multiple areas have been investigated as potential means of delay mitigation. Heale and Forbes (2013) define triangulation as “the use of multiple theories, data sources, methods or investigators within the study of a single phenomenon”. This is of significant value when compared to previous research works where methodologies proposed for delay mitigation have focused on specific areas such as time management and project management methods. The extant literature however reveals that there is a gamut of factors influencing construction delay, hence, tools that could control various project delay factors is of significant value. Notably, effective mitigation strategies for delay risk should holistically cover the project’s duration, cost, and performance (Galway, 2004; Meyer, 2015).

In light of the foregoing, the development of the framework was based on an identification of potential areas for machine learning application. Hence, existing studies in the domain of construction delays were consulted. As a result, the authors theorized three fundamental areas of machine learning application which is backed sufficiently by literature. The areas identified for delay mitigation formed the business understanding as described in the framework presented in figure 1. The framework was further implemented by obtaining relevant data from established databases, as well as construction professionals. The implementation of the framework is discussed in the results and findings section, where machine learning models were developed and were evaluated based on standard performance metrics as established in relevant literature.

The deployment of the framework requires its use in a real-life project. In this case, a robust and reliable dataset will be established and the ML models will be developed as previously discussed. The output of the developed ML models will be used to support decision-making in project planning and risk management exercises. Ultimately, the validity and performance of the framework compared with other delay mitigation approaches can be determined. However, this study investigated the framework’s validity for deployment through interviews with construction professionals. The validation process is a fundamental aspect of construction research, and so numerous studies in the construction domain have relied on interviews to validate the various models/frameworks

developed (Bryman and Bell, 2003; Gao and Low, 2014). Consequently, ten highly experienced construction professionals were interviewed to validate the framework. In selecting the professionals for the interview, demonstrated experience in project management, as well as education, were stipulated (see table 5). Their feedback generally reflects the concept of the framework as a welcome idea to the professional community. The professionals hinted that if such a framework were to be tactical developed into ICT tools or incorporated into existing project management tools, it will be a novel approach to mitigating construction delay. Professionals believe such a framework could serve as a baseline in the planning stage of construction projects. Professionals also hint that the continuous development of the framework through the documentation of data to increase the accuracy and reliability of the model's prediction is a tangible benefit. In general, the framework was viewed as a valuable contribution to construction project management.

It is worthy to mention at this point, that machine learning, though provides an opportunity for modernizing the construction industry, also has some challenges. For instance, the construction industry has not been able to document data suitable for machine learning applications. Ballesteros-Perez et al., (2020) note that the size of data available is one of the limitations of applying artificial intelligence methods in construction management. Strikingly, Farmer (2016) stated "construction has not even made the transition to 'Industry 3.0' status which is predicated on large scale use of electronics and IT to automate production". Additionally, machine learning is not always applicable because it could be expensive or unnecessary if there are traditional engineering-based approaches capable of solving the problem effectively (Larrañaga et al., 2018). This study however shows that the construction industry has a lot to gain in seeking ways in which smart technologies could be used to solve some of its underperformance issues.

6. CONCLUSION

The construction industry has suffered for decades from underperformance issues such as delays. The slow adoption of emerging technologies is a crucial causal factor to under productivity in construction. The construction industry is currently witnessing a technological shift under the umbrella of Construction 4.0. One of the most significant technologies driving Construction 4.0 is machine learning-a subset of artificial intelligence. Construction 4.0 has ushered in a new age of construction digitization, where technologies such as machine learning will be used to solve age-long construction problems, and thus be of value to potential stakeholders. This study is the first to propose a delay mitigation framework as an adaptation of the CRISP-DM framework. The proposed delay mitigation framework was implemented with relevant data from past case studies, and the results, as well as the graphical representation of the framework, were presented to industry professionals. The proposed delay mitigation framework was viewed as a welcome development in the construction community. It was noted that the value of the framework is dependent on the reliability of the data used for machine learning applications. This study makes an important contribution to the construction industry, as it shows the value of adopting emerging computing technologies in solving age-long construction problems. Tools developed based on the framework could support decision-making at the planning stage of construction projects, with due consideration for the limitations consequential of the data used.

DATA AVAILABILITY STATEMENT

All data used in this study are available upon request from the corresponding author.

REFERENCES

- Abd El-Razek M. E., Bassioni H. A. and Mobarak A. M. (2008). Causes of delay in building construction projects in Egypt. *Journal of Construction Engineering and Management*, Vol. 134 No. 11, 831-841.
- Abdul Rahman H., Berawi M.A., Berawi A.R., Mohamed O., Othman M. and Yahya I.A. (2006). Delay mitigation in the Malaysian construction industry. *Journal of Construction Engineering and Management*, Vol. 132, No. 2, 125-133.
- Abdul-Rahman H., Yahya I. A., Berawi M. A. and Wah L. W. (2008). Conceptual delay mitigation model using a project learning approach in practice. *Construction Management and Economics*, Vol. 26, No. 1, 15-27.
- Abedi M., Fathi M.S. and Mohammad M.F. (2011). Major mitigation measures for delays in construction projects. Paper presented at the First Iranian Students Scientific Conference in Malaysia. Universiti Putra Malaysia, Malaysia, 9-10 April.
- Adeli H. (2020). Four decades of computing in civil engineering. In *CIGOS 2019, Innovation for Sustainable Infrastructure*. Springer, Singapore, (pp. 3-11).



- Aibinu A. and Jagboro G. (2002). The effects of construction delays on project delivery in Nigerian construction industry, *International Journal of Project Management*, Pergamon, Vol. 20, No. 8, pp. 593–599.
- Aibinu A. A. and Odeyinka H. A. (2006). Construction delays and their causative factors in nigeria. *Journal of Construction Engineering and Management*, Vol. 132, No. 7, 667–677.
- AlSehaimi, A. and Koskela, L. (2008). What can be learned from studies on delay in construction? In *16th Annual Conference of the International Group for Lean Construction*. International Group for Lean Construction (IGLC), Manchester, United Kingdom, pp. 95-106.
- AlSehaimi A. (2011). *Improving construction planning practice in Saudi Arabia by means of lean construction principles and techniques*. Doctoral dissertation, University of Salford, United Kingdom.
- AlSehaimi A., Koskela L. and Tzortzopoulos P. (2013). Need for alternative research approaches in construction management: Case of delay studies. *Journal of Management in Engineering*, Vol. 29, No. 4, 407-413.
- Assaf S. A., Al-Khalil M. and Al-Hazmi M. (1995). Causes of delay in large building construction projects, *Journal of Management in Engineering*, Vol. 11, No. 2, pp. 45–50.
- Aziz R. F. and Abdel-Hakam A. A. (2016) Exploring delay causes of road construction projects in Egypt, *Alexandria Engineering Journal*. Elsevier, Vol. 55, No. 2, pp. 1515–1539.
- Ballesteros-Perez P., Sanz-Ablanedo E., Soetanto R., Gonzalez-Cruz M.C., Larsen G.D. and Cerezo-Narvaez A., (2020). On the duration and cost variability of construction activities: an empirical study. *Journal of Construction Engineering and Management*, Vol. 146, No. 1, 04019093.
- Bassioni, H. A., Price, A. D. F., & Hassan, T. M. (2005). Building a conceptual framework for measuring business performance in construction: an empirical evaluation. *Construction management and economics*, 23(5), 495-507.
- Bryman, A., & Bell, E. (2003). Breaking down the quantitative/qualitative divide. *Business Research Methods*, 2(1), 465-478.
- Bunker, R. P., & Thabtah, F. (2019). A machine learning framework for sport result prediction. *Applied computing and informatics*, 15(1), 27-33.
- Chai, C. S., Yusof A. M. and Habil H. (2015). Delay mitigation in the Malaysian housing industry: A structural equation modelling approach. *Journal of Construction in Developing Countries*, Vol 20, No. 1, 65-83.
- Choi S., Kim Y. J., Briceno S. and Mavris D. (2016). Prediction of weather-induced airline delays based on machine learning algorithms. In *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)* (pp. 1-6). IEEE.
- Cortes C. and Vapnik V. (1995). Support-vector networks. *Machine learning*, Vol. 20, No. 3, 273-297.
- Conlin J. and Retik A. (1997). The applicability of project management software and advanced IT techniques in construction delays mitigation. *International Journal of Project Management*, Vol. 15, No. 2, 107-120.
- Doloi H., Sawhney A., Iyer K. C. and Rentala S. (2012). Analysing factors affecting delays in Indian construction projects. *International Journal of Project Management*, Vol. 30, No. 4, 479–489.
- El-Kholy A. M. (2019). Exploring the best ANN model based on four paradigms to predict delay and cost overrun percentages of highway projects. *International Journal of Construction Management*, 1-19.
- Egan J. (1998). *Rethinking construction*. Department of Environment, Transport and the Region, United Kingdom.
- Enshassi A., Al-Najjar J. and Kumaraswamy M. (2009) Delays and cost overruns in the construction projects in the Gaza Strip, *Journal of Financial Management of Property and Construction*, Vol. 14, No. 2, pp. 126–151.
- Faridi A. S. and El-Sayegh S. M. (2006). Significant factors causing delay in the UAE construction industry. *Construction Management and Economics*, Vol. 24, No. 11, 1167–1176.
- Farmer M. (2016). Modernise or Die—Time to decide the industry’s future. *Construction Leadership Council*.
- Forbes L. H. and Ahmed S. M. (2010). *Modern construction: lean project delivery and integrated practices*. CRC Press, Taylor and Francis, United States.
- Fugar F. D. and Agyakwah-Baah A. B. (2010). Delays in building construction projects in ghana. *Construction Economics and Building*, Vol. 10, No. (1–2), 103-116.
- Galway, L. (2004). Quantitative risk analysis for project management: A critical review. RAND Corporation.
- Gao S., Low S.P. (2014) Research Methodology. In: *Lean Construction Management*. Springer, Singapore.
- Gao Y. and Tian X. L. (2020). Prefabrication policies and the performance of construction industry in china. *Journal of Cleaner Production*, 120042.
- Gondia A., Siam A., El-Dakhakhni W. and Nassar, A. H. (2020). Machine learning algorithms for construction projects delay risk prediction. *Journal of Construction Engineering and Management*, Vol. 146, No. 1, 04019085.

- Gündüz M., Nielsen Y. and Özdemir M. (2013). Quantification of delay factors using the relative importance index method for construction projects in turkey. *Journal of Management in Engineering*, Vol. 29, No. 2, 133–139.
- Gunduz, M., & Al-Naimi, N. H. (2021). Construction projects delay mitigation using integrated balanced scorecard and quality function deployment. *Engineering, Construction and Architectural Management*.
- Heale R. and Forbes D. (2013). Understanding triangulation in research. *Evidence-Based Nursing*, Vol. 16, No. 4, 98-98.
- Kagermann H., Wahlster W. and Helbig J. (2013). Securing the future of German manufacturing industry: Recommendations for implementing the strategic initiative INDUSTRIE 4.0. *Final report of the Industrie, 4(0)*, National Academy of Science and Engineering, München, Germany.
- Kasimu A.M. and Abubakar D.I. (2012). Causes of delay in Nigeria construction industry. *International Journal of Contemporary Research in Business*, Vol. 4, No. 2, 785–794.
- Keith K. (2018). The construction industry in the 21st century, *Communications of the ACM*, Vol. 61, No. 3, 18-20.
- Khair K., Mohamed Z., Mohammad R., Farouk H. and Ahmed M. E. (2018). A management framework to reduce delays in road construction projects in Sudan. *Arabian journal for science and engineering*, Vol. 43, No. 4, 1925-1940.
- Kim Y. J., Choi S., Briceno S., and Mavris D. (2016). A deep learning approach to flight delay prediction. In *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)* (pp. 1-6). IEEE.
- Larrañaga P., Atienza D., Diaz-Rozo J., Ogbechie A., Puerto-Santana C. E., and Bielza, C. (2018). *Industrial applications of machine learning*. CRC Press, Taylor and Francis, United States.
- Latham S.M. (1994). *Constructing the team*. Joint Review of Procurement and Contractual Arrangements in the United Kingdom Construction Industry, United Kingdom.
- Lek S. and Guégan, J.F. (1999). Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling*, Vol. 120, No. (2-3), pp.65-73.
- Liao Y., Deschamps F., Loures E.D.F.R. and Ramos L.F.P. (2017). Past, present and future of Industry 4.0-a systematic literature review and research agenda proposal. *International Journal of Production Research*, Vol. 55, No. 12, 3609-3629.
- Lo T.Y., Fung I.W. and Tung K.C. (2006). Construction delays in hong kong civil engineering projects. *Journal of Construction Engineering and Management*, Vol. 132, No. 6, 636–649.
- Love P.E.D., Li H., Irani Z., Treloar G.J. and Faniran O.O. (2000). MiDiCON: A model for mitigating delays in construction. Paper presented at the 1st International Conference on Systems Thinking in Management, ICSTM2000. Geelong, Australia, 8–10 November.
- Meyer, W. G. (2015). Quantifying risk: measuring the invisible. Paper presented at PMI® Global Congress 2015—EMEA, London, England. Newtown Square, PA: Project Management Institute.
- Miles M.B. and Huberman A.M. (1984). Qualitative data analysis: A sourcebook of new methods. In *Qualitative data analysis: a sourcebook of new methods*. Sage publications, New York, United States.
- Motaleb O.H. (2020). *Development of a risk response model to handle delays of construction projects in the United Arab Emirates*. Doctoral dissertation, Gordon University, United Kingdom.
- Ng L.F. (2007). Determinant factors of implementing build then sell in Malaysia: Housing developers' point of view. MSc dissertation, Universiti Sains Malaysia, Pulau Pinang, Malaysia.
- Ng S.T., Deng M.Z., Lam K.C. and Skitmore, M. (2000). A conceptual case-based decision model for mitigating construction delays. *International Journal of Construction Information Technology*, Vol. 8, No. 2, 1-20.
- Oesterreich T.D. and Teuteberg F. (2016). Understanding the implications of digitisation and automation in the context of Industry 4.0: A triangulation approach and elements of a research agenda for the construction industry. *Computers in industry*, Vol. 83, 121-139.
- Olatunji S.O., Al-Ahmadi M.S., Elshafei M. and Fallatah Y.A., (2013). Forecasting the Saudi Arabia stock prices based on artificial neural networks model. *International Journal of Intelligent Information Systems*, Vol. 2, No. 5, pp.77-86.
- Olatunji S.O., (2017). Extreme learning machines and support vector machines models for email spam detection. In *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)* (pp. 1-6). IEEE.
- Olawale Y. and Sun M. (2010). Cost and time control of construction projects: Inhibiting factors and mitigating measures in practice. *Construction Management and Economics*, Vol. 28, No. 5, 509–526.
- PricewaterhouseCoopers (2017). Innovation for the earth. Technical Report 161222-113251-LA-OS, World Economic Forum, Davos, Switzerland.
- Sambasivan M. and Soon Y.W. (2007). Causes and effects of delays in Malaysian construction industry,

- International Journal of Project Management*, Vol. 25, No. 5, pp. 517–526.
- Sanni-Anibire M.O., Mohamad Zin R. and Olatunji S.O. (2020a). Causes of delay in the global construction industry: a meta analytical review. *International Journal of Construction Management*, 1-13.
- Sanni-Anibire, M. O., Zin, R. M., & Olatunji, S. O. (2020b). Machine learning model for delay risk assessment in tall building projects. *International Journal of Construction Management*, 1-10.
- Santoso D.S. and Soeng S. (2016). Analyzing delays of road construction projects in cambodia: causes and effects, *Journal of Management in Engineering*, Vol. 32, No. 6, p. 05016020.
- Sawhney A., Riley M. and Irizarry J. (2020). *Construction 4.0: An innovation platform for the built environment*. Routledge, London, United Kingdom.
- Sethi H., Goraya A. and Sharma V., (2017). Artificial intelligence-based ensemble model for diagnosis of diabetes. *International Journal of Advanced Research in Computer Science*, Vol. 8, No. 5.
- Shafique U., and Qaiser H. (2014). A comparative study of data mining process models (KDD, CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, Vol. 12, No. 1, 217-222.
- Schäfer, F., Zeiselmaier, C., Becker, J., & Otten, H. (2018). Synthesizing CRISP-DM and quality management: A data mining approach for production processes. In *2018 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD)* (pp. 190-195). IEEE.
- Schelling, X., & Robertson, S. (2020). A development framework for decision support systems in high-performance sport. *International Journal of Computer Science in Sport*, 19(1), 1-23.
- Smart Market Report (2011). *Mitigation of risk in construction: strategies for reducing risk and maximizing profitability*. McGraw Hill Construction, New York, USA.
- Sweis G., Sweis R., Abu Hammad A., and Shboul A. (2008). Delays in construction projects: The case of Jordan. *International Journal of Project Management*, Vol. 26, No. 6, 665–674.
- Takeichi N., Kaida R., Shimomura A., and Yamauchi T. (2017). Prediction of delay due to air traffic control by machine learning. In *AIAA Modeling and Simulation Technologies Conference*. Grapevine, Texas, USA, (p. 1323).
- Toor S.U.R., and Ogunlana S. (2008). Problems causing delays in major construction projects in Thailand. *Construction Management and Economics*, Vol. 26. No. 4, 395–408.
- Wauters M. and Vanhoucke M., (2017). A nearest neighbour extension to project duration forecasting with artificial intelligence. *European Journal of Operational Research*, Vol. 259, No. 3, pp.1097-1111.
- Weaver-Hart A. (1988). Framing an innocent concept and getting away with it. *UCEA Review*, Vol. 24, No. 2, 11-12.
- Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (Vol. 1). London, UK: Springer-Verlag.
- Woodley, C. (2019). Will digitalisation end construction disputes?. *Construction Research and Innovation*, 10(1), 15-17.
- Xu Z., Zayed T., and Niu Y. (2020). Comparative analysis of modular construction practices in mainland China, Hong Kong and Singapore. *Journal of Cleaner Production*, 245, 118861.
- Yaghini M., Khoshraftar M. M., and Seyedabadi M. (2013). Railway passenger train delay prediction via neural network model. *Journal of advanced transportation*, Vol. 47, No. 3, 355-368.
- Zou P.X., and Sunindijo R.Y. (2015). *Strategic safety management in construction and engineering*. John Wiley & Sons, Hoboken, New Jersey, United States.