Artificial Intelligence in Project Management

ADITYA JOSEPH[1], DR. K J GEORGE[2]

[1] *Student, St. Joseph's College of Engineering and Technology*

[2] *Professor, St. Joseph's College of Engineering and Technology*

Abstract- The main reason of this article is to assist members of the construction project team in understanding the elements that must be regularly checked in order to finish the work on time and on budget. As a result, the study's goal was to create a neural network model (ANN) model that could predict the performance of construction works based on the different speculation specified. Despite the frequency of delay studies, productive research to advance tools and capacity to address the fundamental problem is lacking. As this give idea of outlines the improvement of a machine learning algorithms for finding the risk of delay in high rise building. 36 delay risk variables were initially discovered in existing literature and then converted into questionnaires to examine the likelihood and consequences of the risk factors. A data collection for machine learning applications was created using 48 usable replies gathered from subject matter experts. The approaches of K-Nearest Neighbors (KNN), Neural Networks (ANN), Support Vector Machines (SVM), and Ensemble were investigated. The most important independent factors, according to feature subset selection, were "slowness in decision making," "delay in sub-contractors' work," "architects'/structural engineers' late issuing of instruction," and "waiting for approval of drawings and material supply." The model for finding the risk of delay was find out by ANN, and it was then finished with a classification accuracy of 93.75 percent. After the final model created in this study might help construction companies manage project risk on high rise building.

Indexed Terms- Factors affecting project performance, Artificial neural networks, risk assessment, KNN, ANN, SVM, ensemble methods.

I. INTRODUCTION

A structure plan is approved when it is finished on schedule, under budget, according to specifications, and to the satisfaction of all stakeholders. Functionality, contractor profitability, the lack of lawsuits and legal proceedings, and "fitness for purpose" for occupiers have all been used as indicators of project success. Delays are one of the most serious issues in building projects. Every construction project has delays, and the magnitude of such delays differs greatly between projects. Some works are only a few days behind schedule, while others were pushed back for over a year. Construction activities are generally completed with high costs, prolonged timelines (delays), and quality problems. Delay happens when the contractor, consultant, and client all contribute to the project failing to be finished within the initial, stated, or agreed-upon contract timeframe. Delays create job interruption and loss of productivity, late project completion, higher time-related costs, third-party claims, and contract abandonment or cancellation. It is critical that general management monitor project development in order to reduce the potential of delays occurring or identifying them at an early stage. The construction industry is one that has a lot of unpredictability in its day-to-day operations.

The construction engineering and management (CEM) industry within the architects, engineering, and construction (AEC) industry is beset with its own serious complications because it encompasses a wide range of construction-related activities and processes, as well as human factors and interactions. Construction, as a large sector of the economy, is critical to producing economic growth and long-term national development. According to a McKinsey Worldwide Institute survey performed in 2017, the worldwide construction sector contributes for around 13% of world GDP, with this figure anticipated to rise to 15% by 2020. Meanwhile, construction projects provide a varied range of job opportunities for 7% of the world's largest working population. Meanwhile, building projects provide a diverse range of employment prospects for 7% of the world's working population. Despite its economic significance, an evident concern is low labour productivity throughout the building process, which results in a waste of people, material resources, and financial resources. Because building operations contribute greatly to our society's economic well-being, implementing adequate construction management to increase product performance makes the most sense. It is anticipated that increasing construction productivity by 50 percent to 60 percent or more will add $1.6 trillion to the industry's value per year, significantly boosting global GDP.

In its day-to-day operations, the construction sector is riddled with unpredictability. Building projects are often finished with significant cost overruns, extended deadlines (delays), and quality difficulties, according to an assessment of existing research. The construction firm is approved as a main contribution to the international economy, accounting for 13% of international GDP and expected to increase up to 85% by 2030, with 3 major nations accounting for 57% of world market - Chinese, the U. S, and Indian (Robinson et.al, 2015). Woetzel et al. (2017) says that project that international infrastructure investment will be $3.4 trillion per year from 2013 to 2030, accounting for roughly 4% of international GDP.

For many decades, the worldwide construction sector has been plagued by underperformance difficulties, the most notable of which is the persistent incidence of delays (Mahamid et al. 2012). Claim, project cost and over time, productivity loss and income, and contract cancellation are all possible consequences of delays (Sambasivan et.al 2007). Surprisingly, the construction firm maintain to reach with the exponential rise of towering structures, particularly in Asian and Middle Eastern (Moon et al 2015). Tall structures have suffered from delayed completion periods, despite their exponential pace of expansion. The Council on Tall Buildings and Urban Habitat (CTBUH) reports that there is distributing increase in demolish tall building projects throughout the world (CTBUH 2014). Despite the fact that numerous studies on construction project delays exist (Assaf et al. 1995), these subjects are at best preparatory in nature and do not provide a practical solution to the underlying issue. Surprisingly, AlSehaimi et al. (2013) recommend that the issue of construction delays be addressed might be reduced by alternate research methodologies. The much-desired choice will focus on the creation of innovative tools that will help researchers put their results into operation more successfully, as a result, addressing the problem. Previously, research initiatives aimed to establish risk assessment tools as a feasible technique for mitigating building delays. Risk is described as the impact of uncertainty on an organization's ability to achieve its goals. As a result, risk analysis comprises identifying situations that may cause a project to deviate from its objectives. Kim et al. (2009) created a Bayesian Belief Network (BBN) model to evaluate the risk of construction project delays in Vietnamese. Hossen et al. (2015) proposed combining the Analytical Hierarchy Process (AHP) with the Relative Importance Index (RII) to evaluate the risk of nuclear power reactor project delay. In the Indian construction sector, Muneeswaran et al. (2018) employed the RII and fuzzy ranking to assess schedule delay and risk. Gunduz et al. (2015) suggested a method for estimating the possibility of construction delay in Turkish that is based on a mix of the RII technique and fuzzy logic.

Budayan et al. (2018) developed a computerized fuzzy-based system for assessing the risk of project delay based on past experience, competence, and judgement of users. Despite obvious research efforts, the construction industry is undergoing a technological transformation driven by the 4th industrial revolution (Industry 4.0). The current buzzword in the construction sector is "Construction 4.0," which promises to increase productivity through digitalization and automation (Garc a de Soto et al. 2019). Machine learning, a type of artificial intelligence, is one of the top ten technology powering Industry 4.0. (PricewaterhouseCoopers 2017). Although the industry's desire, as indicated by Construction 4.0, it lags behind other major sectors of latest technological adoption. Surprisingly, a random research search reveals that many research initiatives in other areas that use models to forecast delay risk dominate (Takeichi et al. 2017). Furthermore, there has been little research into the use of machine learning to decrease construction construction delays. Gondia et al. (2020) sought to evaluate the risk of building project delays using machine learning approaches in recent research. The study's main value proposition was identified as the use of factual data gathered from real-world case studies. Unfortunately, these data are hard to get since the construction industry is still significantly poor in capturing and disseminating data suitable for machine learning applications. The usage of subjective data acquired from experts and professionals is the consequence of such a shortcoming. Notably, et al. (2019) proposed using objective data to predict the percent delayed in Egyptian highway projects using a modular neural network technique. Although subjective data is inadequate for generalization of results, it has the potential to speed the usage and acceptance of ml algorithms to solve building difficulties. In view of the foregoing, the goal of this study is to develop a machine learning algorithm to help in the evaluation of delayed risk in tall construction works. The research is based on objective data gathered from experts working on high building projects in Gulf Co - operation Council (GCC) countries. To continue, the research identifies delayed project risks from the existing literature, and then used questionnaire surveys to evaluate the identified delayed risks in two zones: probability and impact. As per the expert's assessment, the obtained data was used to assess the risk of construction delays for the project under consideration. Following that, the data was formatted so that it could be used in machine learning techniques. The ml algorithms used in this research were selected with two factors associated: (1) methodologies often used in the problem's domain, such as order to complete the study of the research (2) Benz ecri's (1973) notion of "letting the data speak for itself," i.e., testing with a suite of algorithms and determining what works best for the dataset. These approaches have been utilized in similar building studies (Pe sko et al. 2017). Finally, the algorithms were compared using key performance indicators such as classification accuracy, accuracy, Cohen's correlation statistic, positive predictive value, and misclassification rate. The sections that follow go into the machine learning algorithms that were employed, the study's methodology, and the findings.

II. MACHINE LEARNING TECHNIQUES

Machine learning is a developing discipline of A.I that is used for database design, which is the creation of statistical abstractions of information that computers can use to build a predictive model Supervised classification is an important subset of machine learning problems. Supervised classification has three main parts: (1) the incidence space, which contains a collection of explanatory variables; (2) the label space, which contains the dependent variable for each instance; and (3) the supervised ml method (Larranaga et al. 2018). This part offers an overview of the frequently adopted ml algorithms used in this study (ANN, KNN, and SVM). Detail mathematics explanations of these tactics may be found in suitable places (Wauters et.al 2017).

A. K-NEAREST NEIGHBORS (KNN)

Using a simple majority decision mechanism, the KNN classifier predicts new classes based on classes associated with the k-instances in the training set (Hodges et.al 1951). Using historical data, KNN is a common and simple method for discovering a data point's nearest neighbours (Vanhoucke et.al 2017). The forecast is built using the training examples that are closest to the current observation (Wauters et.al 2017). The prediction accuracy of the k-NN model is heavily influenced by the value of k. (Sethi et al. 2017). When an instance has k143 neighbours, for example, it is assigned to the class label of the three closest neighbours (Larranaga et al. 2018).

B. ARTIFICIAL NEURAL NETWORKS (ANN)

Artificial Neural Networks (ANNs) are data structures meant to imitate the behaviour of biological neural networks (Pitts et.al 1943). ANNs are seen as adaptive systems composed of connected "neurons" structured in a multi-layered network. It can gather, represent, and simulate complex nonlinear relationships between inputs and outputs by doing numerous concurrent calculations. An input layer, an output layer, and hidden layers comprise the layers. The learning method is based on altering the numerical values associated with the connecting edges among different artificial neurons on a regular basis (Sethi et al. 2017).

ANN's history may be split into three eras, which are as follows.

1) ANN during 1940s to 1960s
2) ANN during 1960s to 1980s
3) ANN from 1980s till Present

The notion of neural network is said to have originated with the work of biologist Warren McCulloch and mathematician Walter Pitts in 1943, when they constructed a rudimentary neural network using electrical circuits to depict how neurons in the brain may operate. The Organization of Behavior, written by Donald Hebb in 1949, proposed that repetitive stimulation of one neuron by another improves its strength each time they are utilised. Taylor proposed an associative memory network in 1956. Rosenblatt devised the Perceptron learning method for the McCulloch and Pitts neuron model in 1958. Bernard Widrow and Marcian Hoff created the "ADALINE" and "MADALINE" models in 1960.

In 1961 Rosenblatt made an unsuccessful attempt but proposed the “backpropagation” scheme for multilayer networks. In 1964 Taylor constructed a winner-take-all circuit with inhibitions among output units. In 1969 Multilayer perceptron MLP was invented by Minsky and Papert. In 1971 Kohonen developed Associative memories. In 1976 Stephen Grossberg and Gail Carpenter developed Adaptive resonance theory.

Hopfield's Energy method was a significant development in 1982. Ackley, Hinton, and Sejnowski invented the Boltzmann machine in 1985. Rumelhart, Hinton, and Williams proposed the Generalized Delta Rule in 1986. In 1988, Kosko introduced Binary Associative Memory (BAM) and the notion of Fuzzy Logic in ANN. According to the historical study, tremendous development has been accomplished in this subject. Neural network-based semiconductors are being developed, as are applications to complicated challenges. Certainly, neural network technology is in a moment of transition right now.

C. SUPPORT VECTOR MACHINES (SVM)

Support vector machines are a hybrid of linear modelling and example-based learning (Witten et al. 2011). SVM is presently one of the most commonly used algorithms methods for regression and classification applications (Olatunji 2017). The theoretical underpinnings of SVM include structural reduced risk and statistics observational learning, with the goal of establishing the hyperplane (decision boundaries) that leads in efficient different classes (Sethi et al. 2017). The boundaries are formed by picking a small number of critical boundary occurrences known as support vectors. It is used to construct a logistic regression function that classifies them to the greatest extent possible. The instantiation method of SVM allows it to accommodate nonlinear components in the function, allowing it to construct higher order of polynomial decision boundaries (Witten et al. 2011). The "kernel trick" refers to this phenomenon, which is just a means of translating the input into high-dimensional feature spaces from which linear classification may be conducted (Cortes et al. 2017). Because of its ability to translate predictors into a higher feature space, SVM has the unique ability to manage intricate associations between predictors and outcomes (Olatunji 2017).

III. METHODOLOGY

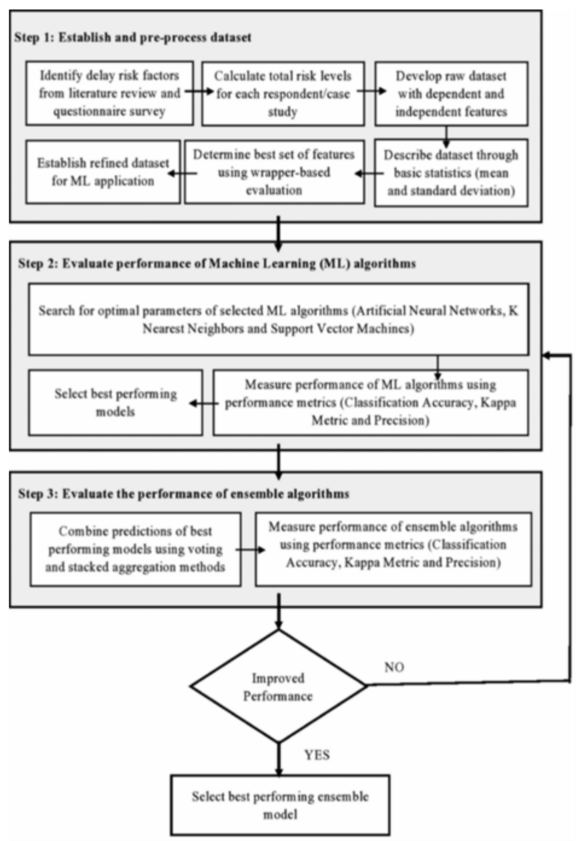

Fig I Development methodology for the proposed delayed risk prediction model

A. CREATING A DATASET

three stages. To begin, a detailed analysis of the reviewed studies was carried out in order to identify the most common delaying possible risks in the construction industry. A review of the literature revealed that there is no consensus on the number of development delay risk variables, with certain academics offering a list of over 200 delay risk indicators (Abdel-Hakam et.al 2016). As a result, our analysis identified 36 often examined delay risk variables culled from high-impact studies published in the previous 15 years (Aibinu et al. 2009). These publications were chosen because of their prominence in the construction research field, as indicated by their high number of citations. Table I shows how the risk factors were further coded and categorized. According to previous study (Gunduz et al. 2013), 9 risk categories were evaluated. Risk categories include product, labour, equipment, contractual relations, governance, financing, environmental considerations, modifications, and scheduling and regulating procedures.

Table-I: Questionnaire result summary analysis

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

The second step comprised the development of a questionnaire survey to get feedback from construction experts on the impact and likelihood of the identified risk factors as they apply to their unique tall construction project. The questionnaire survey has two components (consequence and likelihood), each with a Likert scale that ranges from (1) to (5). Where (1) indicates Very Low and (2) indicates Low. (3) indicates medium, (4) indicates high, and (5) indicates pretty high. Respondents comprised specialists from four main groups operating across the life cycle of tall building projects (consultants, contractors, and clients' representatives/facility managers). The demographic profile of the respondents is shown in Table II. The questionnaire survey was distributed through a combination of means, included hand-delivered paper copies to managers at tall building construction sites and a web-based version issued through email to tall building specialists in the GCC countries. Among them are Saudi Arabia, the United Arab Emirates (UAE), Bahrain, Kuwait, Oman, and Qatar. Fellows and Liu (2015) advocated gathering at least 32 responses for "large number" statistics for determining sample size. As a consequence, 62 responses were obtained, 14 of which were eliminated due to multiple missing cells, possibly affecting the effectiveness of machine learning algorithms. As a consequence, 48 responses were judged appropriate for further development of the dataset. The mean values and standard deviations of the questionnaire findings for all 36 risk factors are shown in Table I.

The third step was to create the values of the parameters for the dataset. The likelihood ratings provided by the experts in the structured questionnaires selected the dataset's input variables. This is based on the premise that risk is proportionate to probability. As a consequence, Equations (1) and (2) were used to calculate the output variable (2). Where (RL1) is the risk level estimated for each respondent for a particular delay risk factor, and (RL2) is the total systematic risk for each respondent.

$$\text{Risk Level (RL1)} = \text{Consequence (C)} * \text{Likelihood(L)} \quad (1)$$
$$\text{Risk Level (RL2)} = \sum \text{(RL1)}/n \quad (2)$$

where n is the total number of delay risk factors. The risk levels were then assessed using the risk discretization matrix, as shown in Figure II. As a consequence, the whole dataset displayed three unbalanced groups, as shown in Figure III, with very high risks accounting for 19%, high risks accounting for 54%, and moderate risks accounting for 27%. Table III depicts the final structure of the created dataset.

B. PRE-PROCESSING OF DATA

Machine learning challenges involve the finding of the ideal set of attributes that improves a model's anticipated performance. This helps to the resolution of a frequent regular checks as the "curse of dimensionality." In this study, the Waikato Environment for Knowledge Analysis (WEKA 3.8.3) was used. The University of Waikato in New Zealand created this open-source machine learning programme based on Java (Witten et al. 2011). WEKA is a popular tool in the machine learning field because of its ease of use and wide selection of algorithms available on the platform (Larranaga et al. 2018). The ML algorithms used in this work in the WEKA environment are as follows: Ensemble voting: "meta.Vote," Ensemble stacking: "meta.stacking," k-NN is abbreviated as "IBk," ANN is abbreviated as "Multilayer Perceptron," SVM is abbreviated as "SMO," and Ensemble voting is abbreviated as "meta.Vote." The "Correlation AttributeEval" model was used to calculate the correlation and rank of various available features to the prediction output. The Recursive Feature Elimination (RFE) approach was then used to select attributes (Akande et al. 2015). The whole feature set (V) is split in half in RFE to extract the best V/2 features, and the poorest V/2 features are discarded. The splitting procedure is repeated until just one of the better features remains. Following that, the feature subset with the highest accuracy/or best performance measure is picked as the optimal subset to utilize. The RFE technique yielded the following results (Table IV). The next stage was to figure out which features would be most useful in the model creation process. The wrapper-based feature assessment method was used. This is when a strong classifier is used to compare all of the feature sets, and the best set of features is chosen based on relevant performance measures. SVM (SMO) was used as the wrapper-based classifier in this investigation. The feature set with the highest accuracy was the best V/8 features, as shown in Table 6. (Described in Table IV). In essence, the refined dataset for machine learning models contained four independent variables ("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"), as well as three classes as dependent variables (described in Figure III). The dataset was divided into a 66 percent to 34 percent train-test ratio.

Table- II: Respondent demographics

Demographics	Category	Frequency (No.)	Percentage (%)
Professional organization	Contractors	14	30
	Contactors	17	35
	Clients' representatives/facility managers	17	35
Professional role	Executive director	6	13
	Director	7	15
	Senior project manager	15	32
	Civil engineer	9	19
	Facility manager	11	23
Years of experience	5 to 10	13	28
	10to 15	12	24
	> 15	23	48
Location	United Arab Emirates	20	42
	Saudi Arabia	15	32
	Kuwait	7	14
	Bahrain	3	6
	Oman	2	4
	Qatar	1	2

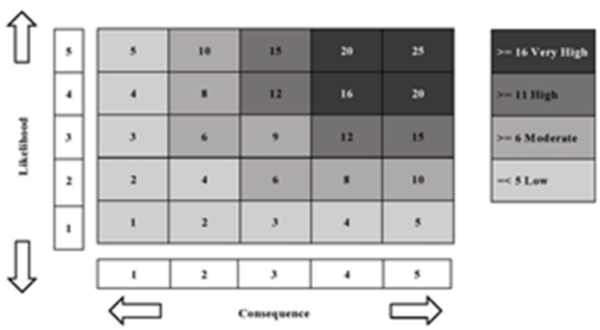

Fig II Discretization matrix adopted for delay risk assessment

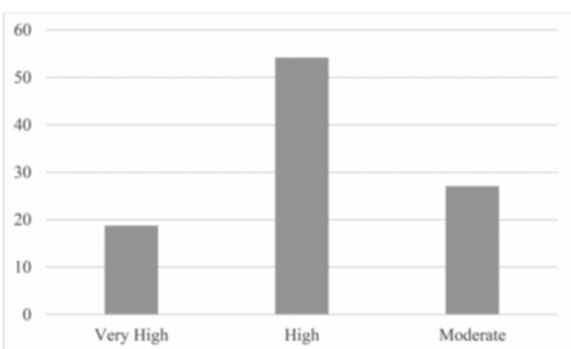

Fig III Unbalanced distribution of category of data

Table- III: Compiled data set structure

Project Case	Mat. 1	Mat. 2	Mat. 3	-	-	-	Sch. 6	Sch. 7	Sch. 8	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

C. ANALYZE MACHINE LEARNING ALGORITHMS

The setting of optimization hyperparameters affects the performance of machine learning algorithms. A systematic search was conducted in this work, with multiple sets of hyperparameters being used to gradually train the model until sufficient results were reached. Table V lists the optimization hyperparameters for the methods investigated in this work.

Table- IV: Description of feature sets based on Correlation AttributeEval

RFE process	No. of features	Description
V features	36	All features
Best V/2	18	Cont. 2; Cont. 6; Sch. 8; Mat. 3; Sch. 7; Cont. 1; Fin. 2; Cont. 4; Equip. 1; Chng. 6; Chng. 1; Chng. 3; Sch. 4; Equip. 2; Man. 1; Sch. 3; Man. 2; Gov. 2
Best V/4	9	Cont. 2; Cont. 6; Sch. 8; Mat. 3; Sch. 7; Cont. 1; Fin. 2; Cont. 4; Equip. 1
Best V/8	4	Cont. 2; Cont. 6; Sch. 8; Mat. 3
Best V/16	2	Cont. 2; Cont. 6
Best feature	1	Cont. 2

Table- V: Description of ML algorithm's optimization parameters

ML Algorithm	Optimization parameters	
ANN (Multilayer Perceptron)	Learning rate	A value in the range [0,1] that changes the speed at which the weights of each connection between neurons is updated.
	Momentum	A value in the range [0,1] that uses the direction of previous weight updates to adjust the weight change speed.
	Network topology	Number of neurons in each hidden layer.
	Transfer function	It determines the output of each neuron given the input.
KNN (IBk)	k value	Number of nearest neighbors considered by the algorithm.
	Search algorithm	The manner in which the algorithm should find the nearest neighbors.
	Distance function	The distance in the feature space. The most common include Euclidean, Manhattan and Minkowski distances.
SVM (SMO)	Cost function, C	Also known as complexity constant, controls the trade-offs between errors and the margin size.
	Kernel function	A symmetric function of two arguments that returns the value of the inner product of the two mapped arguments.
	Tolerance parameter	Controls the amount of permissible SVM optimization problem-solving error
	Filter	Data transformation of the input variables before training.

Table- VI: Performance of SVM wrapper-based evaluation for various feature sets

ML Algorithm	Performance measure	All features	Best V/2	Best V/4	Best V/8	Best V/16	Best feature
SVM (SMO)	Classification Accuracy (%)	62.5	68.75	81.3	87.5	56.25	56.25
	Kappa Statistic	0.35	0.48	0.68	0.77	0.09	0
	Precision	0.67	0.75	0.89	0.82	0.6	0.56

D. PERFORMANCE EVALUATION

The confusion matrices were used to assess the performance of the algorithms deployed. True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN) are all represented by values in the confusion matrix (TN). In binary classification (2X 2 matrices), TP and TN represent the number of occurrences properly categorized as positive and negative, respectively, whereas FP and FN represent the number of instances incorrectly classified as positive and negative, respectively. In the case of multi-class confusion matrices, the TPs, FPs, FNs, and TNs are computed independently for each class and the weighted average is calculated. In this situation, TPs are the occurrences that have been accurately categorized along the diagonal of the matrix. For each class, TNs is the total of all categorized instances outside of the class in question. For each class, FNs are instances of that class that were wrongly classed as other classes. Finally, a class's FPs are the sums of all other classes' samples (Tharwat 2018). In a subsequent portion of this study, an exemplary case is offered. The classification accuracy of a model is defined as the ratio of correct classifications to the total number of classifications. The Cohen's kappa statistic is also seen to be particularly beneficial in circumstances of class imbalance, as in this research (see Figure 3). The performance measures utilized in this study are obtained using Equations (3)– (7)

Classification Accuracy = ((TP+ TN))/N (3)

where N= total number of samples

And Misclassification Error (1- Classification Accuracy)

$$\text{Cohen's kappa statistic} = \frac{\left(\frac{TP}{N}+\frac{TN}{N}-A\right)}{1-A} \quad (4)$$

$$A=\frac{(FN+TP)}{N}*\frac{(FP+TP)}{N}+\frac{(FP+TN)}{N}*\frac{(FN+TN)}{N}$$

$$\text{Precision (Positive Preditive Value)} = \frac{TP}{TP+FP} \quad (5)$$

$$\text{TP Rate (Sensitivity)} = \frac{TP}{TP+FN} \quad (6)$$

$$\text{FP Rate (Specificity)} = \frac{FP}{FP+TN} \quad (7)$$

E. EXAMINE ENSEMBLE ALGORITHMS

Ensemble approaches in machine learning can increase the performance of basic classifiers. This method combines the predictions of basic classifiers using a conventional mechanism. Voting and stacking ensemble approaches were examined in this work (Xia et al. 2011; Kuncheva and Rodr guez 2014). As the name indicates, voting entails the combining of two or more sub-models by some mechanism such as the mean or mode. As a result, each sub-model votes on the outcome. Stacking allows another algorithm to learn how to integrate the predictions of other sub-models as effectively as possible (Brownlee 2018)."

CONCLUSION

The building business is undergoing a technological transformation that will increase productivity. Project delays and abandonment, which have become a threat in tall construction projects, are a serious under-productivity concern. Numerous research on the causes of building delays have previously been conducted; however, the present difficulty is establishing prescriptive methods to address the issue. This work contributes significantly to this goal by establishing a model that might aid in the assessment of delay risk in tall construction projects. 36 delay risk factors were considered in the development of the delay risk model, and further analysis revealed that "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" were the most influential factors. Subjective data on these characteristics gathered from industry specialists was then utilized to create a classification model using ANN, with the results revealing an exceptional performance level of 93.75 percent classification accuracy. According to the findings of this study, machine learning might be a viable technique for constructing models that properly estimate the risk of delays in tall construction projects. The model created in this study may be inferred to actively enhance risk-based decision making in tall construction projects. Despite the fact that the model is data was collected from tall building experts in the Gulf Cooperation Council (GCC), the approach used may be adapted to a variety of construction settings across the world. Machine learning-based delay risk assessment models might be included into cutting-edge construction risk management systems in the future.

REFERENCES

[1] Muizz O. Sanni-Anibire, Rosli Mohamad Zin & Sunday Olusanya Olatunji, “Machine learning model for delay risk assessment in tall building projects (2020), pp 1-11.

[2] Abd El-Razek ME, Bassioni HA, Mobarak AM, “Causes of delay in building construction projects in Egypt (2008), vol. 134, Issue11 pp 931-941.

[3] Aibinu AA, Odeyinka HA, “Construction delays and their causative factors in Nigeria (2006), vol. 132, Issue7 pp 667-677.

[4] Akande KO, Owolabi TO, Olatunji SO, “Investigating the effect of correlation- based feature selection on the performance of support vector machines in reservoir characterization (2015), vol. 22 pp 515-522

[5] AlSehaimi A, Koskela L, Tzortzopoulos P, “Need for alternative research approaches in construction management: case of delay studies (2013), vol. 29, Issue4 pp 407-413.

[6] Aziz RF, Abdel-Hakam AA, “Exploring delay causes of road construction projects in Egypt (2016), vol. 55, Issue2 pp 1515–1539.

[7] Bayram S, “Duration prediction models for construction projects: In terms of cost or physical characteristics? KSCE (2017), vol. 29, Issue4 pp 2049–2060.

[8] Budayan C, Dikmen I, Talat Birgonul M, Ghaziani A., “A computerized method for delay risk assessment based on fuzzy set theory using MS Project (2018), vol. 22, Issue8 pp 2714–2712.

[9] Choi S, Kim YJ, Briceno S, Mavris D, “Prediction of weather-induced airline delays based on machine learning algorithms. In 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC). IEEE. (2016), pp 1-6.

[10] Darko A, Chan AP, Adabre MA, Edwards DJ, Hosseini MR, Ameyaw EE, “Artificial intelligence in the AEC industry: Scientometric analysis and visualization of research activities (2020), vol. 112, Issue4 pp 081-103. .

[11] Doloi H, Sawhney A, Iyer KC, Rentala S, “Analysing factors affecting delays in Indian construction projects (2012), vol. 30, Issue4 pp 479–489.

[12] Faridi AS, El-Sayegh SM, “Significant factors causing delay in the UAE construction industry (2006), vol. 24, Issue11 pp 1167–1176.

[13] Hossen MM, Kang S, Kim J, “Construction schedule delay risk assessment by using combined AHP-RII methodology for an international NPP project (2015), vol. 47, Issue3 pp 362–379.

[14] Kaming PF, Olomolaiye PO, Holt GD, Harris FC, “Factors influencing construction time and cost overruns on high-rise projects in Indonesia (1997), vol. 15, Issue1 pp 83–94.

[15] Kim S-Y, Tuan NV, Ogunlana SO, “Quantifying schedule risk in construction projects using Bayesian belief networks (2009), vol. 27, Issue4 pp 39–50.

[16] Sovan Lek, J.F. Gue´gan, “Artificial neural networks as a tool in ecological modelling, an introduction. Ecol Model (1999), vol. 120, Issue (2-3) pp 65–73.