Teknik Dergi, 2022 12577-12600, Paper 695

Comparing Performances of Machine Learning Techniques to Forecast Dispute Resolutions

Murat AYHAN[1]
Irem DIKMEN[2]
M. Talat BIRGONUL[3]

ABSTRACT

This paper compares classification performances of machine learning (ML) techniques for forecasting dispute resolutions in construction projects, thereby mitigating the impacts of potential disputes. Findings revealed that resolution cost and duration, contractor type, dispute source, and occurrence of changes were the most influential factors on dispute resolution method (DRM) preferences. The promising accuracy of the majority voting classifier (89.44%) indicates that the proposed model can provide decision-support in identification of potential resolutions. Decision-makers can avoid unsatisfactory processes using these forecasts. This paper demonstrated the effectiveness of ML techniques in classification of DRMs, and the proposed prediction model outperformed previous studies.

Keywords: Construction disputes, dispute resolution methods, multiclass classification, dispute management.

1. INTRODUCTION

Encountering conflicts is almost inevitable in construction projects particularly due to the complex, fragmented, and dynamic nature of the construction industry along with involvement of numerous parties usually in an adversarial relationship [1]. In case the parties in a conflict cannot reach a satisfactory outcome, the conflict may progress into a dispute [2]. Awwad et al. [3] stated that the construction industry is exceptionally susceptible to conflicts and disputes, and these may often escalate to lawsuits. At the same time, a growth in the number and severity of construction disputes were reported by several researchers [3, 4, 5]. Moreover, construction disputes can be detrimental as they have the potential to disrupt the

---

Note:
- This paper was received on April 29, 2021 and accepted for publication by the Editorial Board on September 20, 2021.
- Discussions on this paper will be accepted by November 30, 2022.

- https://doi.org/10.18400/tekderg.930076

1 Gazi University, Civil Engineering Department, Ankara, Turkey
muratayhan@gazi.edu.tr - https://orcid.org/0000-0002-2011-4190

2 Middle East Technical University, Civil Engineering Department, Ankara, Turkey
idikmen@metu.edu.tr - https://orcid.org/0000-0002-6988-7557

3 Middle East Technical University, Civil Engineering Department, Ankara, Turkey
birgonul@metu.edu.tr - https://orcid.org/0000-0002-1638-2926

workflow and lead to delayed schedules, budget overruns, poor communication, and damaged business relationships [6]. Therefore, it would be beneficial to avoid disputes; however, if the occurrence of disputes cannot be precluded, management personnel need to resolve them through various resolution processes [7]. Selecting an appropriate dispute resolution method (DRM) to resolve a dispute is crucially important as it paves the way for successful project completion [8]. However, management personnel have difficulties in reaching satisfactory outcomes out of disputed cases.

A best method that handles all disputes is not available as projects vary in scale, complexity, nature, and so forth [9]. Numerous interrelated factors should be considered to successfully manage disputes, making it a challenging decision-making problem. Contrarily, the construction industry relies on the experience and the level of knowledge of the decision-maker in such decisions [10]. On the other hand, a study on Turkish construction industry unveiled that the dispute management decision-making is characterized as an unconscious process, and the industry requires novel tools to overcome this deficit. It is also highlighted that there is a need for a more systematic approach to DRM selection instead of the industry's reliance on the current subjective approach [11].

The techniques available in the Artificial Intelligence (AI) domain has the potential to mitigate the subjectivity, which dominates dispute management decision-making, by providing systematical decision-support [8]. Solving an engineering problem via AI techniques involves learning from data while simulating underlying functional relationships that are difficult to rationalize, even if the interdependencies between inputs and outputs are unknown. Among various AI applications, machine learning (ML) domain focuses on developing systems capable of learning from data about a specific task automatically. It is possible to perform data classification tasks via ML techniques as these techniques can develop algorithms that utilize prespecified features to predict target labels [12].

This paper argues that appropriate DRM can be forecasted systematically, given the circumstances of the case, so that early-warnings of potential resolutions can be achieved. For this reason, ML techniques were utilized to develop classification models that forecast the occurrence of disputes and their potential resolutions, thereby mitigating the negative impacts of potential disputes. In Ayhan et al. [13], the effectiveness of ML techniques in early prediction of dispute occurrence was demonstrated, and promising classification accuracy results were obtained. This paper builds upon the work by Ayhan et al. [13] and applies multiclass classification techniques to forecast potential resolutions prior to dispute occurrence. For this reason, initially, the variables affecting dispute resolutions were identified by an extensive literature review, and the findings were used to develop a novel conceptual model that depicts the common factors influencing dispute resolutions. Considering that understanding the influential factors underlying a dispute determines the performance of a construction project [14], this conceptual model is the basis for the proposed study. Then, using the established conceptual model, past project data were collected via questionnaires with the decision-making authorities of the projects. Then, Chi-square tests of association were performed on the collected dataset to identify the relationships between the influential factors and DRMs. Based on the results of the Chi-square tests, the attributes, which were identified as statistically significantly associated with DRM preferences, were kept and remaining attributes were eliminated. This resulted in establishment of a classification model for forecasting dispute resolutions. The obtained classification model was experimented via alternative ML techniques and classifier performances were evaluated by 10-times repeated 10-fold cross-validation.

2. RESEARCH BACKGROUND

The main concern of the studies from the dispute resolution literature is avoiding ineffectual DRMs, which generally involve processes leading to settlement in courts [15]. In construction industry, litigation is the conventional method of providing involuntary and binding dispute resolution despite being costly and lengthy. Moreover, in many industries (i.e., Turkish construction industry) litigation is commonly used rather than seeking other resolutions despite the widespread dissatisfaction related to the litigation [14]. Arbitration was initially an inexpensive and efficient alternative to litigation; however, following the growing dissatisfaction, its categorization as an alternative dispute resolution (ADR) technique has been criticized [16]. Consequently, construction professionals resorted to ADR techniques due to their cost and time advantages, less adversarial nature, and lower legal requirements; indeed, common ADR techniques such as dispute review boards (DRB), mediation, and negotiation have gained popularity in the construction industry [7]. The disputed cases in this research's dataset were resolved through six different techniques as (1) litigation (LIT); (2) arbitration (ARB); (3) DRB; (4) mediation (MED); (5) senior executive appraisal (SEA); and (6) negotiation (NEG). Litigation and arbitration are considered as conventional DRMs, and the remaining methods are considered as ADR techniques. Technical and legal details of these techniques will exceed the scope of this paper.

A review of the literature reveals that researchers focused mainly on the most adopted DRMs in a specific region, or the implementation and potential advantages/disadvantages of specific DRMs [3]. For example, King et al. [17] conducted a questionnaire among experts in Malaysian construction industry to identify the most beneficial and resorted DRMs in terms of cost, time, and satisfaction. Focusing on Sri Lankan construction industry, Illankoon et al. [18] identified 15 dispute causes and 13 factors affecting DRM selection from the literature along with the most effective ADR method from perspectives of various parties in a project. Specific to disputes in Nepalese road construction projects, Kisi et al. [6] conducted surveys with experts to identify the preference frequencies of DRMs, so that various parties can comprehend the best practices related to a claim category. Sinha and Jha [19] identified the causes of commonly occurring disputes that are followed by litigation and causing delays in Public-Private-Partnership (PPP) road projects in India along with the causes leading to utilization of certain DRMs. The aforementioned studies provide valuable statistical frameworks and reflect current tendencies related to the DRM selection in various regions, rather than providing systematical decision-support systems for management personnel. On the other hand, AI applications can enable systematical selection of dispute resolutions and provide the necessary decision-support to obtain satisfactory outcomes [20].

Among studies that utilized AI techniques, Cheung et al. [8] developed a Case-Based Reasoning (CBR) model that retrieves similar dispute cases. Chen [21] proposed a model for construction professionals facing potential litigation from change order related disputes using the K-Nearest Neighbor (KNN) algorithm, and the model allows its users to select the most similar cases. Liu et al. [22] proposed a CBR system to extract experiences from past projects by retrieving similar cases. However, acting solely on similar cases may not be adequate because characteristics of disputes and possible resolutions differentiate, and finding a matching case is rather difficult.

It is observed that the literature is rich in studies on construction litigation, and there is a specific interest on predicting the outcomes of litigated cases. For example, Chau [23] aimed to generate insights on how a construction claim would be resolved if litigation were preferred by using a Particle Swarm Optimization (PSO) based Artificial Neural Network (ANN) model, which achieved 80.00% accuracy. Chen and Hsu [24] proposed a model that identifies the potential litigation probability of a case via ANN classifier with 84.61% accuracy. Using the same dataset of litigated cases filed in Illinois courts, several ML based models were developed to predict the outcomes of court rulings including ANN [25], Boosted Decision Trees (BDT) [26], and two hybrid systems [27, 28] that achieved 66.67%, 89.59%, 91.15%, and 96.02% accuracy, respectively. Specific to disputes caused by differences in site conditions, Mahfouz et al. [29] reviewed the links between 15 legal factors and litigation outcomes using several ML techniques, which led to the highest accuracy of 88.00% from the Naïve Bayes (NB) model. Although the advantages of predicting the outcomes prior to litigation are evident such that a party can keep away from courts upon identification of an unfavorable result, the mentioned studies do not offer any alternatives to litigation.

There are some other studies that are not limited with litigation and aim to provide decision-support during resolutions. Chong and Zin [2] utilized factor analysis approach to analyze DRM selection rationale of the decision-makers in the Malaysian construction industry. Chaphalkar et al. [30] claimed that if disputed parties can forecast the outcome with some certainty, they may prefer settling before conventional DRMs to avoid expenses and aggravation. For this reason, they developed a Multilayer Perceptron (MLP) model by using 204 variation claim cases resolved through arbitration processes in India, and classified these cases as accepted, rejected, or partly accepted based on 16 factors affecting decisions of the arbitrators. Although the model was significantly successful, it only targeted variation claims and arbitration cases.

Among other efforts, several ML models were developed by using a dataset of 152 PPP projects undertaken in Taiwan, and these cases were classified based on 15 features about the project and the dispute. Initially, Chou [31] performed DRM classification using single and ensemble ML models at two distinct phases as (1) project initiation; and (2) in the aftermath of dispute occurrence. The following study by Chou et al. [10] combined the capabilities of fuzzy logic, genetic algorithm, and Support Vector Machines (SVM) to forecast DRM selection. Once again by using several ML techniques, Chou et al. [32] discovered rule sets for classification of possible dispute resolutions. These studies can successfully forecast DRMs; however, the dataset was composed solely of instances with a certain project delivery system (i.e., PPP) from a certain construction industry (i.e., Taiwan).

Therefore, this study applied ML techniques to forecast the potential resolutions prior to dispute occurrence. The multiclass classification performances were compared with each other to select the best performing classifier for identification of potential resolutions, so that construction professionals can avoid unsatisfactory resolution processes, which will reduce the unnecessary costs, delays, and aggravation caused by using inconclusive processes. Moreover, management personnel can take the necessary precautions beforehand, thanks to the early-warnings of the proposed model.

3. RESEARCH METHODOLOGY

The methodology for this research is visualized in Figure 1, which involves three steps as (1) development of the conceptual model; (2) development of the classification model; and (3) finalization of the classification model.

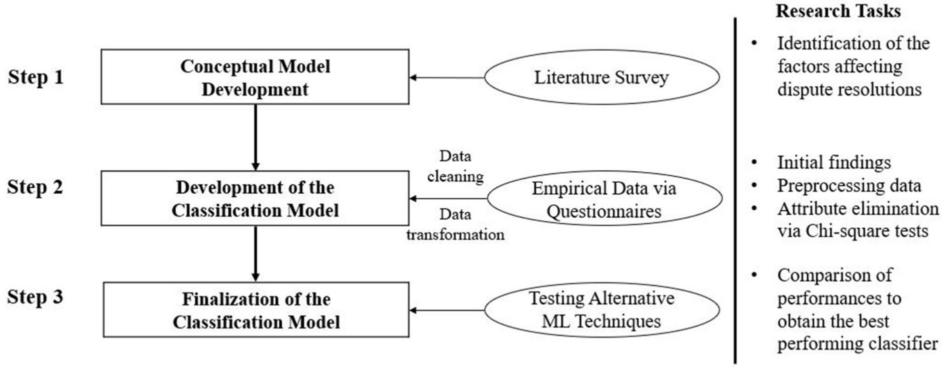

Figure 1 - Research methodology

3.1. The Conceptual Model

Numerous factors affecting dispute resolutions were identified from the literature and the most frequently perceived ones were picked for use in the conceptual model. The findings revealed 42 frequently perceived attributes and these factors can be grouped in six categories as (1) project characteristics (i.e., contract value); (2) changes or unexpected events; (3) delays; (4) characteristics of the disputed case (i.e., disputed extension of time (EoT) amount); (5) DRM characteristics (i.e., resolution duration); and (6) knowledge level on the DRM. These categories were based on the findings of several research including (1) İlter [11] that identified 16 factors affecting the recommendations of the legal professionals in DRM selection; (2) Chou et al. [10] that identified project attributes impacting the utilized DRM; (3) Awwad et al. [3] that listed 12 factors affecting the choice of ADR; (4) Lee et al. [7] that related 29 factors with DRM selection.

The aforementioned six groups were utilized in categorization of the influential factors into a conceptual model. A thorough discussion on identification of the factors from the literature is available in Ayhan [33], which proposed conceptual models composed of variables affecting the dispute occurrence and their resolutions that led to development of two distinct classification models. These conceptual models are depicted in Figure 2. The first model classified dispute occurrence of construction projects as disputed and undisputed projects [13], which demonstrated the effectiveness of ML techniques in early prediction of dispute occurrence. The second model, which is the subject of this paper, applies multiclass classification techniques to forecast potential resolutions prior to dispute occurrence.

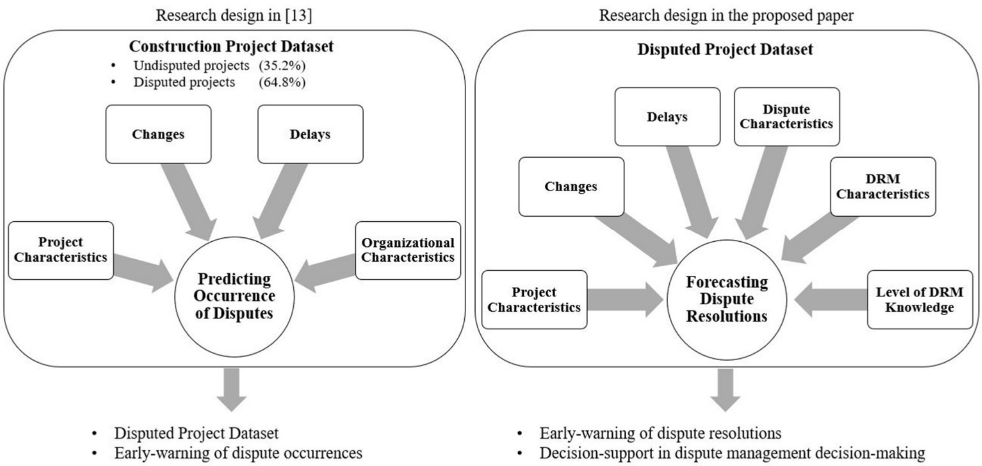

Figure 2 - The conceptual models

3.2. Development of the Classification Model

A questionnaire was prepared to collect past project data about the variables listed in the conceptual model. The questionnaire involved six distinct sections to collect (1) demographic information about the experts; (2) project specific information; (3) information to detect any variations or unexpected events during the course of the project; (4) dispute specific information; (5) information about the DRM characteristics; and (6) information about the level of knowledge of experts about certain DRMs. In section 5, participants were asked to rank the importance of DRM related features such as importance of preserving relationships, bindingness of the process, and so forth to understand what features the decision-makers consider during DRM selection. The questionnaire is available in the study by Ayhan [33].

Participants of the questionnaire raised their concerns in sharing disputed project data, and this was understandable considering the sensitive and confidential nature of dispute related data. Consequently, it was difficult to find participants willing to share such information. In order to overcome the difficulties in collecting dispute data, this study utilized a snowball sampling method.

The collected data was initially investigated, and noisy data was removed. Following the data cleansing, the next step was to understand the existence of associations among the variables because the effects of each attribute will be different. This research preferred to use Chi-square statistics, which a test of association among categorical variables [34]. Thus, at this point, the cleaned dataset was processed by converting numeric data to categories. The Chi-square analysis was performed by using the IBM SPSS Statistics, and this analysis enabled elimination of the attributes in the conceptual model that do not have a statistically significant association with dispute resolutions. As a result of attribute elimination, a simpler model for classification was obtained based on the collected dataset.

3.3. Finalizing the Classification Model

Alternative ML algorithms can be experimented on the developed classification model, and their performances can be compared to obtain the best classifier that will be proposed as the final classifier for forecasting dispute resolutions. The open-source WEKA 3.8.3. software was used for this purpose as it provides plenty of ready-to-use ML algorithms to its users. There is evidence in the literature that proves the software can generate stable results with equal or better performance compared with similar applications [35]. Moreover, rather than struggling with complicated computer codes, the optimization of algorithm hyperparameters can be conveniently performed via a simple graphical interface [36]. Therefore, WEKA can be confidently and easily used in data classification tasks in this research.

3.4. Machine Learning Techniques

The classification task in this research is a multiclass classification problem because the output is a multiple category variable. Due to differences in the characteristics of the data, an ML technique that can handle all data classification tasks do not exist [15]. To determine the best technique, the bias resulting from the ML algorithm should be coherent with the problem characteristics [36]. This can be achieved by experimenting promising single techniques on the dataset and comparing their performances with each other to select the best performer [27]. Therefore, this paper assessed the performances of several ML techniques.

The findings of a research identifying the top 10 data mining algorithms was the reference for the evaluated ML techniques in the proposed study [36]. Among the 10 algorithms, there were techniques for various purposes (i.e., data clustering); however, the task in this paper requires classification, and consequently, the classification algorithms were evaluated only. Apart from these algorithms, MLP was also tested because it is intensively preferred in construction domain. Within this context, the six ML techniques used in this paper are (1) NB; (2) KNN; (3) C4.5 Decision Tree (DT); (4) MLP; (5) Polynomial kernel SVM; and (6) Radial Basis Function (RBF) kernel SVM. The ML approach in this study is illustrated in Figure 3.

It should be noted that although there exists an enhanced release of C4.5 algorithm, called C5.0 algorithm, the researchers sticked with the C4.5 because it is freely available unlike the enhanced version. Moreover, it is revealed that C4.5 can still produce somehow equal or better performance compared with C5.0 [37, 38].

Binary classification capabilities of the ML algorithms may be extended for the multiclass problem in this research, except the SVM, which can only perform on binary tasks [39]. In the case of SVM algorithm, the problem should be decomposed into several binary classification tasks. WEKA supports four decomposition techniques as (1) one-vs-one (OvO); (2) one-vs-all (OvA); (3) random correction code (RCC); and (4) exhaustive correction code (ECC). In the OvA approach, for an output with k categories, there will be k binary classification tasks that aim to separate the instances belonging to a category from the combination of remaining categories, which will result in k classifiers [40]. Meanwhile, in the OvO technique, the instances belonging to a category are separated from only one other category. In other words, a classifier is trained to distinguish one class from another in a pairwise approach, which will result in $k(k-1)/2$ classifiers [41]. In this training scheme, a voting strategy, which is based on each classifier's classification decision, makes the final class assignment for a new instance [40]. Dietterich and Bakiri [42] proposed the error-correcting output codes (ECOC) method that decomposes a multiclass problem into a set of binary problems. In this decomposition, it is aimed to improve the classification performance by symbolizing each category of an output by bits of code words. WEKA has two extensions of the original ECOC, which are RCC and ECC. Although the ECC technique is more sophisticated than the RCC, an increment in the number of output categories causes an exponential growth in the number of classifiers to be generated in the ECC, which may result in infeasible solutions. In such cases, the RCC technique can be used where the only difference from the ECC is that there is randomization during the generation of the code word matrix at the beginning of operations [36]. In addition to solutions obtained from the ML techniques that can solve multiclass problems without using any further decomposition techniques, the multiclass problem in this research was solved by using the aforementioned decomposition techniques for all evaluated ML algorithms.

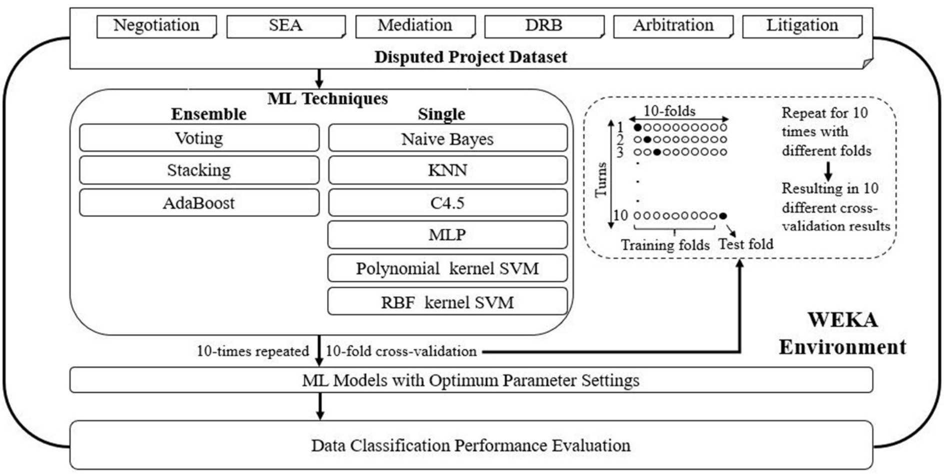

Figure 3 - The ML approach in this study

A convenient way to improve the performances of single ML techniques is to establish ensemble models that aim to compensate the errors in each single technique by synthesizing them [27]. In this paper, (1) voting; (2) stacked generalization; and (3) the AdaBoost algorithm were the techniques for ensemble model development. Voting is perceived as the easiest way of combining various classifiers [40], and due to this simplicity, it was utilized in this research. In voting, the class of an instance is the class that obtains the majority of the votes of the contained classifiers. However, when the minority decision is correct about the class of an instance, a misclassification problem may occur as the technique is not capable of determining which decision is the correct one. This shortcoming is addressed in stacking, where a meta-learner classifier is specifically trained to identify the reliable classifier (base-learner), and consequently, ensemble models obtained through stacking has the potential to perform better than voting [36]. The driving forces in developing ensemble models are (1) improving classification accuracy (as in stacking); (2) decreasing variance (as in bagging); and (3) decreasing bias (as in boosting). Unlike boosting where the next classifier is trained on the misclassifications of the former classifier systematically, developing complementary classifiers is by chance in bagging [40]. Therefore, boosting was utilized in this research, and the AdaBoost algorithm was specifically preferred because it is commonly used, easy to implement, and adaptable to various ML techniques [36].

Table 1 - Parameter configurations for the utilized ML techniques

Algorithm	Parameter	Search Range
NB	No parameter optimization	-
KNN	k neighbors	1-50 (increment by 1)
	Distance measurement function	Chebyshev, Euclidean, Manhattan
	Distance weighting method	Equal, Inverse, Similarity
C4.5	Pruning	Yes, No
	Reduced error pruning	Yes, No
	Subtree raising	Yes, No
	Threshold factor for pruning	[0.01-0.50] (increment by 0.01)
	Lowest number of instances at leaves	[1-10] (increment by 1)
	Number of folds for pruning	[2-5] (increment by 1)
	Laplace counts at leaves	Yes, No
MLP	Number of hidden layers	0, 1, 2, (total number of inputs and outputs) / 2
	Epochs (cycles)	500, 1000
	Momentum	[0.1-0.9] (increment by 0.1)
	Learning rate	[0.1-0.9] (increment by 0.1)
Poly. SVM	Penalty parameter	[2^{-2}-2^{15}] (increment exponentially by 1)
	Exponent	[1-10] (increment by 1)
RBF SVM	Penalty parameter	[2^{-2}-2^{15}] (increment exponentially by 1)
	Gamma	[2^{-15}-2^{4}] (increment exponentially by 1)
Voting	Combination rule	Majority Voting, Average of Probabilities

Table 1 - Parameter configurations for the utilized ML techniques (continue)

Algorithm	Parameter	Search Range
Stacking	Base-learner	Top 3 performing single classifiers in turns
	Meta-learner	Remaining 5 classifiers (excluding itself)
AdaBoost	Number of iterations	10
	Boosting mechanism	Resampling, Reweighting

Theoretical framework related to evaluated ML techniques exceeds the scope of this paper. However, it should be known that each ML technique has specific parameters that determines their success. Table 1 shows the parameters to be optimized for each algorithm, and the ranges to be searched. To determine the optimized value of a numeric parameter, there is a need to use a validation set or cross-validation method. WEKA has plenty of evaluators for this purpose such as cross-validated parameter selection and grid search.

4. DATA COLLECTION AND DESCRIPTION

Initially, data about 151 construction projects were collected for this study. After removal of noisy and unrepresentative cases, there were 108 projects from 19 different countries. The data was collected by meetings with 78 experts individually, which represented 75 different companies. The participants were selected among professionals with decision-making authority. The average construction industry experience of the participants was 18 years, and 47.0% of them have worked for more than 15 years. Therefore, the opinions of experienced professionals were reflected in this study. Moreover, the dataset contained a broad array of projects to reflect the changes in the decision-making process to the research that result from varying characteristics of projects and disputes. In the dataset, it is observed that 38 projects were completed without any disputes (35.2%), while in 70 projects (64.8%), at least one disputed issue was experienced. These 70 disputed projects generated 82 distinct dispute cases, and the model was based on 54 dispute cases that the disputants reported satisfactory resolutions.

The attributes, their categories, and relative frequencies in the dataset are given in Table 2. This research used techniques such as Chi-square tests and NB algorithm that requires discrete data. Thus, all numeric attributes in the dataset were converted to categorical variables for computational purposes. This data transformation should be handled with care because the performance of the classification algorithm may be adversely affected if the distinctive features of the data are suppressed during discretization. With this consideration, WEKA provides an information gain-based supervised discretization method. In this method, discretization ranges can be defined based on the output so that the subjectivity during data conversion can be mitigated, and split points generating the maximum information gain can be determined so that the information loss can be diminished.

Table 2 - Attributes, categories, and relative frequencies

ID	Attribute	Categories and relative frequencies
PC1	Project location	Domestic (55.6%); International (44.4%)
PC2	Project value	< \$10 mil. (24.1%); \$10-\$100 mil. (35.2%); > \$100 mil. (40.7%)
PC3	Planned duration	< 1 year (14.8%); 1-2 years (31.5%); 2-3 years (25.9%); > 3 years (27.8%)
PC4	Type of construction	Housing (20.4%); Commercial (11.1%); Industrial (11.1%); Transportation (24.1%); Power plants/lines (3.7%); Medical (5.6%); Water supply/reservoir (7.4%); Sports, cultural and educational (9.3%);Public (5.6%); Soil works (1.9%)
PC5	Type of contractor	Single (79.6%); Joint venture (13.0%); Consortium (7.4%)
PC6	Type of employer	Public (46.3%); Private (40.7%); Public-Private-Partnership (13.0%)
PC7	Type of contract	Private (53.7%); Public procurement (14.8%); FIDIC red (20.4%); FIDIC silver/yellow (11.1%)
PC8	Payment method	Fixed (lump-sum) (46.3%); Unit price (53.7%)
PC9	Project delivery	Design-bid-build (63.0%); Design-build (18.5%); Engineering-procurement-construction (18.5%)
PC10	Design complexity	Very low (16.7%); Low (11.1%); Moderate (18.5%); High (35.2%); Very high (18.5%)
PC11	Construction complexity	Very low (9.3%); Low (13.0%); Moderate (14.8%); High (37.0%); Very high (25.9%)
C1	Changes	Yes (59.3%); No (40.7%)
D1	Delays (ratio)	0% (27.8%); 0%-20% (20.4%); 20%-40% (16.7%); > 40% (35.2%)
DC1	Disputant	Owner/employer (16.7%); Contractor (83.3%)
DC2	Phase	Construction (87.0%); Transfer/Repair/Maintenance (13.0%)
DC3	Dispute source	Cost of change orders (5.6%); Time & cost of change orders (31.5%); Measurement & valuation (11.1%); Delay in site handover (5.6%); Defects/errors/poor quality (11.1%); Imprudent contractor (11.1%); Late payments (5.6%); Errors/changes in quantities (7.4%); Site/soil investigation (9.3%); Interpretation of contract articles (1.9%)
DC4	Suspension of works	Yes (61.1%); No (38.9%)
DC5	Disputed amount	< 5 mil. \$ (44.4%); 5-25 mil. \$ (27.8%); 25-75 mil. \$ (14.8%); > 75 mil. \$ (13.0%)
DC6	Settled amount	\$0 (6.3%); < \$1 mil. (31.3%); \$1-\$5 mil. (31.3%); \$5-\$25 mil. (16.7%); > \$25 mil. (14.6%)
DC7	Success rate	0% (1.9%); 0%-25% (1.9%); 25%-50% (25.9%); 50%-75% (35.2%); > 75% (35.2%)
DC8	EoT claim occurrence	Yes (51.9%); No (48.1%)
DC9	Disputed EoT amount	0 days (51.9%); 0-6 months (18.5%); 6 months-1 year (11.1%); > 1 year (18.5%)
DC10	Settled EoT amount	0 days (53.7%); 0-6 months (16.7%); 6 months-1 year (14.8%); > 1 year (14.8%)
DC11	Success rate (EoT)	0% (53.7%); 0%-25% (0.0%); 25%-50% (0.0%);50%-75% (7.4%); > 75% (38.9%)

Table 2 - Attributes, categories, and relative frequencies (continued)

ID	Attribute	Categories and relative frequencies
DRMC1	Resolution cost	$0 (53.7%); < $100k (9.3%); $100k-$350k (13.0%); $350k-$1 mil.(14.8%), > $1 mil.(9.3%)
DRMC2	Resolution duration	< 2 weeks (20.4%); 2-4 weeks (18.5%); 1-3 months (29.6%); 3-6 months (3.7%); 0.5-2.5 years (14.8%); > 2.5 years (13.0%)
DRMC3	Preserving relationships	Rank-1 (33.3%); Rank-2 (5.6%), Rank-3 (5.6%); Rank-4 (7.4%); Rank-5 (16.7%); Rank-6 (9.3%); Rank-7 (5.6%); Rank-8 (0.0%); Rank-9 (5.6%); Rank-10 (11.1%)
DRMC4	Speed of process	Rank-1 (22.2%); Rank-2 (20.4%), Rank-3 (11.1%); Rank-4 (22.2%); Rank-5 (5.6%); Rank-6 (5.6%); Rank-7 (9.3%); Rank-8 (3.7%); Rank-9 (0.0%); Rank-10 (0.0%)
DRMC5	Cost of process	Rank-1 (9.3%); Rank-2 (14.8%), Rank-3 (16.7%); Rank-4 (16.7%); Rank-5 (5.6%); Rank-6 (5.6%); Rank-7 (16.7%); Rank-8 (11.1%); Rank-9 (3.7%); Rank-10 (0.0%)
DRMC6	Bindingness	Rank-1 (0.0%); Rank-2 (9.3%), Rank-3 (9.3%); Rank-4 (3.7%); Rank-5 (14.8%); Rank-6 (20.4%); Rank-7 (7.4%); Rank-8 (18.5%); Rank-9 (14.8%); Rank-10 (1.9%)
DRMC7	Confidentiality	Rank-1 (1.9%); Rank-2 (3.7%), Rank-3 (3.7%); Rank-4 (3.7%); Rank-5 (0.0%); Rank-6 (13.0%); Rank-7 (5.6%); Rank-8 (5.6%); Rank-9 (20.4%); Rank-10 (42.6%)
DRMC8	Fairness	Rank-1 (22.2%); Rank-2 (13.0%), Rank-3 (7.4%); Rank-4 (9.3%); Rank-5 (1.9%); Rank-6 (7.4%); Rank-7 (16.7%); Rank-8 (5.6%); Rank-9 (9.3%); Rank-10 (7.4%)
DRMC9	Flexibility in procedures	Rank-1 (1.9%); Rank-2 (1.9%), Rank-3 (16.7%); Rank-4 (13.0%); Rank-5 (11.1%); Rank-6 (5.6%); Rank-7 (14.8%); Rank-8 (18.5%); Rank-9 (7.4%); Rank-10 (9.3%)
DRMC10	Control over the process	Rank-1 (0.0%); Rank-2 (3.7%), Rank-3 (7.4%); Rank-4 (3.7%); Rank-5 (14.8%); Rank-6 (16.7%); Rank-7 (13.0%); Rank-8 (16.7%); Rank-9 (13.0%); Rank-10 (11.1%)
DRMC11	Reaching remedying solutions	Rank-1 (3.7%); Rank-2 (16.7%), Rank-3 (1.9%); Rank-4 (14.8%); Rank-5 (20.4%); Rank-6 (11.1%); Rank-7 (9.3%); Rank-8 (9.3%); Rank-9 (11.1%); Rank-10 (1.9%)
DRMC12	Willingness in reaching solutions	Rank-1 (5.6%); Rank-2 (11.1%), Rank-3 (20.4%); Rank-4 (5.6%); Rank-5 (9.3%); Rank-6 (5.6%); Rank-7 (1.9%); Rank-8 (11.1%); Rank-9 (14.8%); Rank-10 (14.8%)
K1	Litigation knowledge	Very low (7.4%); Low (11.1%); Moderate (16.7%); High (33.3%); Very high (31.5%)
K2	Arbitration knowledge	Very low (11.1%); Low (13.0%); Moderate (18.5%); High (38.9%); Very high (18.5%)
K3	DRB knowledge	Very low (25.9%); Low (5.6%); Moderate (20.4%); High (20.4%); Very high (27.8%)
K4	Mediation knowledge	Very low (5.6%); Low (5.6%); Moderate (18.5%); High (31.5%); Very high (38.9%)
K5	SEA knowledge	Very low (3.7%); Low (5.6%); Moderate (11.1%); High (20.4%); Very high (59.3%)
K6	Negotiation knowledge	Very low (0.0%); Low (1.9%); Moderate (5.6%); High (40.7%); Very high (51.9%)

5. RESULTS OF THE CHI-SQUARE TESTS

Table 3 - Results of Chi-square tests for attribute elimination

Identifier	Attribute	p-value	Selected for final model
PC1	Project location	0.236	NO
PC2	Project value	0.349	NO
PC3	Planned duration	0.221	NO
PC4	Type of construction	0.131	NO
PC5	Type of contractor	**0.003**	**YES**
PC6	Type of employer	0.581	NO
PC7	Type of contract	0.540	NO
PC8	Payment method	0.354	NO
PC9	Project delivery system	0.172	NO
PC10	Level of design complexity	0.601	NO
PC11	Level of construction complexity	0.342	NO
C1	Changes	**0.018**	**YES**
D1	Delays (ratio)	0.088	NO
DC1	Disputant	0.390	NO
DC2	Phase	0.406	NO
DC3	Dispute source	**0.014**	**YES**
DC4	Suspension of works	0.778	NO
DC5	Disputed amount	0.485	NO
DC6	Settled amount	0.668	NO
DC7	Success rate	0.910	NO
DC8	EoT claim occurrence	0.202	NO
DC9	Disputed EoT amount	0.976	NO
DC10	Settled EoT amount	0.709	NO
DC11	Success rate (EoT)	0.129	NO
DRMC1	Resolution cost	**0.000**	**YES**
DRMC2	Resolution duration	**0.000**	**YES**
DRMC3	Importance of preserving relationships	0.943	NO
DRMC4	Importance of speed of process	0.823	NO
DRMC5	Importance of cost of process	0.687	NO
DRMC6	Importance of bindingness	0.571	NO
DRMC7	Importance of confidentiality	0.521	NO

Table 3 - Results of Chi-square tests for attribute elimination (continue)

Identifier	Attribute	p-value	Selected for final model
DRMC8	Importance of fairness	0.069	NO
DRMC9	Importance of flexibility in procedures	0.308	NO
DRMC10	Importance of control over the process	0.468	NO
DRMC11	Importance of reaching remedying solutions	0.387	NO
DRMC12	Importance of willingness in reaching solutions	0.759	NO
K1	Knowledge level on litigation	**0.005**	**YES**
K2	Knowledge level on arbitration	**0.016**	**YES**
K3	Knowledge level on DRB	0.699	NO
K4	Knowledge level on mediation	0.480	NO
K5	Knowledge level on SEA	0.899	NO
K6	Knowledge level on negotiation	0.876	NO

Presence of insignificant attributes in the model causes an adverse impact on the performance of ML techniques and eliminating these attributes can improve the generalization performance [27]. Among attribute elimination techniques, Chi-square tests, which provide a practical method for revealing the relationships among categorical variables, was found to be suitable for this study's dataset due to its capabilities in handling attributes with multiple categories and diverse data distributions, unlike other alternative techniques [43]. Although the existence of association between variables can be identified by using this method, the strength of the association cannot be determined. Fortunately, association's strength can be detected among nominal variables by Cramer's V measure and among ordinals by Somers' d measure, where both measures can handle input and output variables with unequal numbers of categories [34].

Chi-square results were tabulated in Table 3 with their corresponding exact probability values (p-values). The p-value of a nominal variable was calculated by using the exact Pearson Chi-square statistics, while the p-value of an ordinal variable was calculated by using the Mantel-Haenszel linear association test. The p-values were evaluated using 95% confidence interval (CI) such that attributes were either eliminated from (p-value > 0.005) or selected for (p-value ≤ 0.005) the classification model. Chi-square results showed that statistical significance of the association with dispute resolutions can be proved in only seven attributes. Figure 4 shows the established classification model with seven attributes.

Table 4 is the contingency table showing Chi-square test results for the selected attributes in the classification model along with the strength of association values. For Cramer's V values exceeding 0.25, there exists a very strong association between input and output variables [44]. Among nominal attributes, all three selected attributes (PC5, C1, and DC3) had a very strong association with dispute resolutions. Without considering the signs in front of the value, Somers' d values exceeding 0.10 imply partially strong relationship, and values exceeding 0.40 imply a strong relationship [45]. Thus, DRMC1 and DRMC2 had strong association with dispute resolutions, while K1 and K2 had partially strong association.

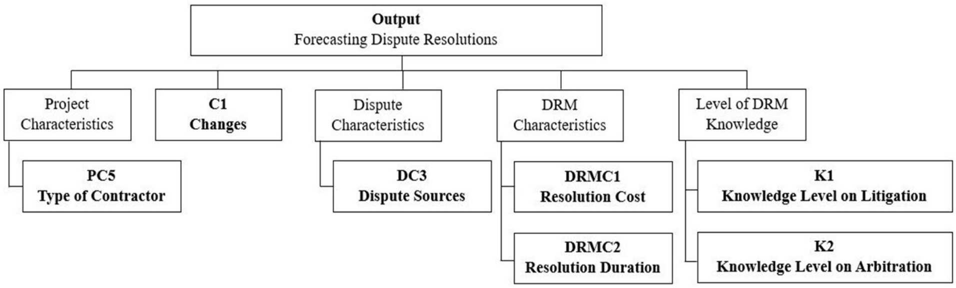

Figure 4 - The attributes in the classification model

Table 4 - Contingency table and strength of association values for the selected attributes

		Resolution method (Relative frequency (%))						Association
Attribute	Categories	LIT	ARB	DRB	MED	SEA	NEG	Strength
PC5 Type of contractor	Single	18.6	11.6	2.3	9.3	18.6	39.5	Cramer'sV 0.514
	JV	0.0	0.0	57.1	14.3	0.0	28.6	
	Consortium	25.0	25.0	0.0	0.0	50.0	0.0	
C1 Changes	Yes	6.3	18.8	9.4	12.5	25.0	28.1	Cramer'sV 0.491
	No	31.8	0.0	9.1	4.5	9.1	45.5	
DC3 Dispute source	Source 1	0.0	0.0	0.0	0.0	66.7	33.3	Cramer'sV 0.498
	Source 2	5.9	23.5	5.9	17.6	23.5	23.5	
	Source 3	50.0	0.0	16.7	0.0	33.3	0.0	
	Source 4	0.0	0.0	0.0	0.0	0.0	100.0	
	Source 5	33.3	0.0	0.0	16.7	0.0	50.0	
	Source 6	16.7	0.0	0.0	0.0	0.0	83.3	
	Source 7	33.3	0.0	0.0	0.0	33.3	33.3	
	Source 8	0.0	25.0	75.0	0.0	0.0	0.0	
	Source 9	20.0	20.0	0.0	20.0	0.0	40.0	
	Source 10	0.0	0.0	0.0	0.0	100.0	0.0	
DRMC1 Resolution cost	$0	0.0	0.0	0.0	0.0	34.5	65.5	Somers' d -0.909
	$0-$100k	0.0	0.0	0.0	100.0	0.0	0.0	
	$100k-$350k	28.6	0.0	71.4	0.0	0.0	0.0	
	$350k-$1mil.	62.5	37.5	0.0	0.0	0.0	0.0	
	> $1 mil.	40.0	60.0	0.0	0.0	0.0	0.0	

Table 4 - Contingency table and strength of association values for the selected attributes (continue)

Attribute	Categories	Resolution method (Relative frequency (%))						Association Strength
		LIT	ARB	DRB	MED	SEA	NEG	
DRMC2 Resolution duration	< 2 weeks	0.0	0.0	0.0	36.4	0.0	63.6	Somers' d -0.667
	2-4 weeks	0.0	0.0	0.0	10.0	10.0	80.0	
	1-3 months	0.0	0.0	31.3	0.0	43.8	25.0	
	3-6 months	0.0	0.0	0.0	0.0	100.0	0.0	
	0.5-2.5 years	25.0	75.0	0.0	0.0	0.0	0.0	
	> 2.5 years	100.0	0.0	0.0	0.0	0.0	0.0	
K1 Litigation knowledge	Very low	0.0	0.0	0.0	0.0	50.0	50.0	Somers' d -0.309
	Low	0.0	0.0	16.7	0.0	16.7	66.7	
	Moderate	11.1	22.2	0.0	11.1	22.2	33.3	
	High	16.7	11.1	5.6	11.1	16.7	38.9	
	Very High	29.4	11.8	17.6	11.8	11.8	17.6	
K2 Arbitration knowledge	Very low	16.7	0.0	0.0	0.0	16.7	66.7	Somers' d -0.283
	Low	0.0	0.0	14.3	0.0	42.9	42.9	
	Moderate	40.0	0.0	0.0	10.0	10.0	40.0	
	High	4.8	19.0	4.8	19.0	19.0	33.3	
	Very High	30.0	20.0	30.0	0.0	10.0	10.0	

6. DATA CLASSIFICATION TESTS USING ML TECHNIQUES

In cases where the number of samples is limited, it is reasonable not to allocate instances to distinct sets for training, validation, and testing. Instead, all instances can be used for extracting knowledge to avoid loss of information in an already limited dataset. This can be achieved by cross-validation (CV), which aims to use all instances for training purposes, and then, the accuracy is obtained by resampling the dataset [46]. The *k*-fold CV is a commonly used version that is based on training and testing the model k-times randomly on different subsets of training data to generate an estimate of the performance of a classifier on new data [40]. The optimum value for *k* is put forth as 10 based on trials with diverse datasets and algorithms [20]. To avoid uneven representation among folds, stratification is used during resampling. Moreover, to decrease the high variance in CV results, the process was repeated 10 times and the final accuracy value is determined by averaging the results from each process [46]. Within this context, stratified 10-fold CV was utilized in this research by repeating the process 10 times.

Table 5 tabulates the outcomes from the 10-times repeated 10-fold CV analysis of the evaluated single ML techniques. All algorithms generated their best average classification results when ECC decomposition technique was used. The most successful classifiers are C4.5, NB, and MLP with 86.48%, 85.93%, and 83.33% average classification accuracy, respectively.

Table 5 - 10-times repeated 10-fold CV performance of single classifiers

Algorithm	Average Accuracy (%)	%95 CI Average Accuracy (%)
NB ECC	85.93	[84.50-87.35]
KNN ECC	74.63	[72.46-76.80]
C4.5 ECC	86.48	[85.08-87.88]
MLP ECC	83.33	[80.54-86.13]
Polynomial Kernel SVM ECC	82.04	[79.17-84.90]
RBF kernel SVM ECC	80.93	[79.84-82.02]

The top three algorithms are used as candidates during development of the ensemble classifiers. In voting, the ensemble classifier synthesized the classification decisions of these three algorithms. In stacking, two algorithms are merged as base-learner and meta-learner, where the learning process was performed on the complete dataset for base-learner, but the meta-learner can only access to the instances that are not misclassified by the base-learner. In this research, the aforementioned top three algorithms are used as base-learners in turns, and meta-learners were the remaining five techniques in turns, excluding the technique used as base-learner. Such an approach was preferred to avoid using classifiers of the same type during stacking [40]. This process brings out 15 stacked classifiers. The AdaBoost algorithm aims to transform weakly performing classifiers into successful ones and all six of the evaluated ML techniques are boosted via AdaBoost algorithm.

Although it is expected that ensemble models would improve the classification accuracy, they did not improve the performance at all times. Table 6 tabulates the outcomes from the 10-times repeated 10-fold CV analysis of the ensemble classifiers that performed better than their single counterparts. The stacked classifier combining C4.5 ECC and NB ECC classifiers achieved 86.67% average accuracy, while boosting of C4.5 ECC classifier by the AdaBoost algorithm generated 88.15% average accuracy. The most outstanding classification performance belonged to the classifier, which was generated by using the majority voting technique, and 89.44% average classification accuracy was achieved.

Table 6 - 10-times repeated 10-fold CV performance of ensemble classifiers

Algorithm	Average Accuracy (%)	%95 CI Average Accuracy (%)
Majority voting	89.44	[87.37-91.52]
Stacking: C4.5 ECC + NB ECC	86.67	[85.04-88.30]
AdaBoost: C4.5 ECC	88.15	[85.34-90.95]

7. DISCUSSION OF FINDINGS

Among PC attributes, the only selected attribute was the type of contractor (PC5). In this dataset, it is observed that when the contractor is a consortium, in 50.0% of the cases, they resorted to conventional DRMs. Meanwhile, when the contractor is a joint-venture (JV), conventional DRMs were never used and instead, DRB was the most common technique for resolution, which was used in 57.1% of the cases. This is in line with the study by Lingard et al. [47], which stated that when the participating firms remain as independent entities that do not have joint liability (i.e., consortium), it is possible for one company to gain while the other suffers; therefore, it is more likely to use conventional DRMs that fit to this nature. On the other hand, JVs have joint liability so that both rewards and penalties are shared among the participating companies. Considering that settling through ADR processes can generate win-win results unlike the conventional DRMs that declare a winner and a loser [16], it is more likely for JVs to resolve their disputes through ADR methods. Among eliminated PC attributes, considering that DRM is specified prior to dispute occurrence via contract documents, it was interesting that the type of contract (PC7) was not significantly associated. This is because parties prefer using alternative DRMs that are not specified in the contract with the aim of using the method that best suits their needs (i.e., via addendum to contract).

The occurrence of changes (C1) during the execution of a construction project was found to be an influential factor on DRM selection. Indeed, there is supporting evidence in the literature stating that the occurrence of changes is a common problem in the construction industry that can trigger problems during the execution of a construction project such as cost and time related conflicts [48]. In this study's dataset, it is revealed that the disputes resulting from occurrence of changes are resolved through ADR techniques mostly (75%). Considering that most disputes resulting from changes end up in courts [48], the proposed classification model can offer alternative and efficient ways to decision-makers for resolution rather than resorting to court involved unsatisfactory processes immediately.

Among DC attributes, the only selected attribute was the dispute source (DC3). The association of dispute sources with DRM preferences was also considered as an influential factor in other similar models in the literature [10, 31].

Among DRMC attributes, resolution cost (DRMC1) and duration (DRMC2) were the most influential attributes in the classification model with the highest strength of association values. Other DRMC attributes that reflected 10 different features of DRMs were ranked based on their importance; however, none of them were in the final model. This shows that experts based their DRM preferences mainly on the cost and the time they are willing to allocate. This was an expected outcome. For example, in the study of Cheung and Suen [9] that developed a multi-attribute utility theory model for resolution strategy selection, the highest utility factors were obtained from resolution duration and cost among all DRM selection criteria. Similarly, Illankoon et al. [18] identified time to reach a settlement as the most influential factor during DRM selection.

Among K attributes, K1 and K2, which represent the knowledge level of the decision-maker on litigation and arbitration, were in the classification model. In other words, the experience of the experts with litigation and arbitration shapes their DRM preferences. It is observed that litigation preference is increasing with the increasing level of knowledge on litigation. This is also valid for arbitration process. This reveals that experts that feel competent in litigation and/or arbitration prefer using these techniques over ADR methods.

The experiments revealed that C4.5, NB, and MLP are the three-best single ML classifiers (when ECC decomposition technique is used) for the dataset in question, which generated average accuracy values of 86.48%, 85.93%, and 83.33%, respectively. Consequently, it is evident that the C4.5 classifier outperformed the competing classifiers in this research. The superiority of DT was expected as it provides an effective structure in which alternative decisions can be evaluated when complex information with several variables should be considered [26].

Following the experiments with single ML techniques, ensemble classifiers were developed to enhance the classification performance. It is experimentally revealed that the majority voting technique, which synthesized the classification decisions of the three top performing single classifiers, produced the highest average classification accuracy value as 89.44%. Therefore, the final classification model for forecasting dispute resolutions is the majority voting classifier. Classification accuracies of the compared single and ensemble classifiers were visualized in Figure 5.

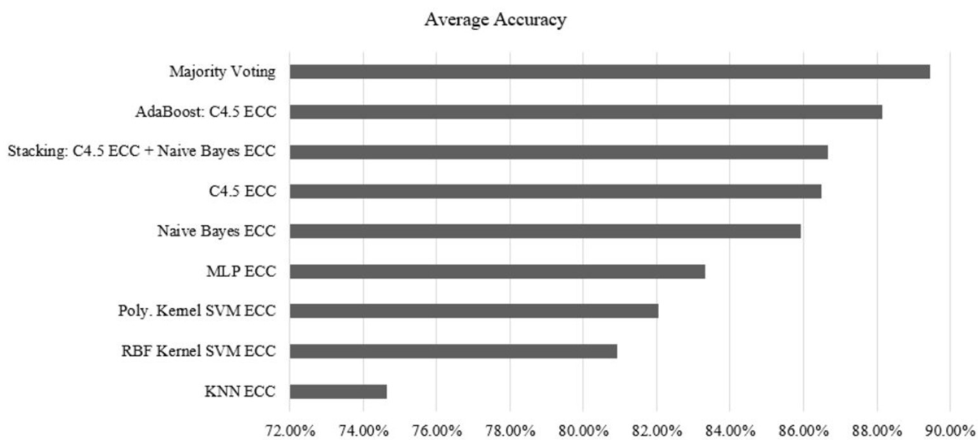

Figure 5 - Classification accuracies of the classifiers

Besides voting, 15 stacked classifiers were developed. Theoretically, when the classifiers that constitute the stacked classifier are diverse algorithms and their classification accuracies are high, the resultant ensemble model is expected to outperform the constituent classifiers [40]. However, this was not the case in many stacking trials. In this research, the most successful stacked classifier was developed by combining C4.5 ECC and NB ECC classifiers that achieved 86.15% average accuracy. A similar case was also observed for the classifiers boosted by the AdaBoost algorithm as some boosted classifiers showed weak performances. Theoretically, when the complexity of the single classifiers is high with respect to the amount of training instances, the outcome of the boosting is expected to be unsatisfactory [36]. The most successful AdaBoost model was obtained from boosting the C4.5 ECC classifier that resulted in an average accuracy value of 88.15%. In summary, the boosted classifier improved the base-learner (C4.5) performance by 1.67%, while the stacked classifier improved the base-learner (C4.5) by 0.19% and the meta-learner (NB) by 0.74%.

Among limited empirical research on forecasting dispute resolutions, Chou [31] achieved 83.82% test set accuracy during project initiation phase, and 69.05% test set accuracy in the aftermath of dispute occurrence. On the other hand, the ensemble models in the same study enhanced the accuracy on the test set during project initiation phase by achieving 84.65% accuracy. In Chou et al. [10], based on 10-fold CV results on the test set, an average accuracy of 61.75% was obtained from single SVM classifiers for DRM classification. This performance was improved by combining SVM with genetic algorithm and fuzzy logic to achieve 77.04% average 10-fold CV accuracy for the test set. In Chou et al. [32], the best average 10-fold CV result was obtained as 81.12% through SVM. Benchmarking these, it is evident that the performance of the proposed classification model is higher, and the results are encouraging.

During identification of the factors affecting dispute resolutions, it is observed that numerous subjective factors are effectual. Thus, the main limitation of this research is its dependence to subjective judgments of participating experts. Scarcity of sample projects is one other constraint. Even though the collected sample of construction projects is representative, the sample size is nonetheless limited because of the difficulties in acquiring such sensitive information. The sample size can be enlarged to improve the generalization of the presented model. However, it should be noted that such data scarcity problems were also encountered in other research since historical data is scarce in nature for construction industry [49]. Finally, although various ML techniques were compared in this study, the extent of the experimented techniques were limited and considerable classification techniques, which might offer potential improvements in the accuracy, were not evaluated in this research, that can be done as further research.

8. CONCLUDING REMARKS

In this research, dispute resolutions were forecasted by using alternative ML techniques, which included multiclass classification and ensemble models. A novel conceptual model was developed to identify the factors affecting dispute resolutions, and it is revealed that prediction models can be developed to provide decision-support that rely on the attributes in the conceptual model. The conceptual model can effectively guide decision-makers by highlighting factors to be considered during resolutions.

For construction professionals, the early-warnings of potential resolutions provided by the proposed model can enable avoiding the unnecessary costs, delays, and aggravation caused by using inconclusive resolution processes. The proposed model can help to reveal whether a selected DRM was appropriate or not, which may lead the decision-maker, upon identification of an inconclusive DRM, to settlement before deciding to implement certain resolution processes. For example, negotiation is generally the first choice in settling disputes, but if it is not the appropriate process for resolution, it would cause waste of time and money without reaching satisfactory outcomes. In such cases, the proposed model can be used to inform the users whether to give up negotiations, and resort to other DRMs.

In this research, the performances of various classifiers were compared with each other, and C4.5, NB, and MLP classifiers produced outstanding average classification accuracies of 86.48%, 85.93%, and 83.33% respectively, when the problem was decomposed into binary classification tasks using ECC decomposition technique. Moreover, three ensemble classifiers outperformed the single techniques. The first ensemble classifier was developed by stacking, which combined C4.5 ECC and NB ECC classifiers, and it achieved 86.67% average accuracy. The second one was developed by using AdaBoost algorithm on C4.5 ECC classifier, which achieved 88.15% average accuracy. The highest value was obtained as 89.44% from the majority voting model, which combined performances of C4.5, NB, and MLP classifiers. Therefore, the effectiveness of ML techniques in classification of DRMs has been demonstrated, and specifically, it is revealed that an appropriate combination of ML classifiers can further improve the classification performance. Moreover, the classification accuracy of the proposed prediction model outperformed previous studies.

Future work can focus on mitigation of data scarcity both by expanding the sample size, and by integrating the classification models with soft computing approaches. The improvement due to the proposed decision-support approach can be tested on other classification problems in construction management domain as a further study. Moreover, further improvement in classification accuracy can be pursued by using other classification techniques such as Random Forests and so forth.

References

[1] Alaloul, W. S., Hasaniyah, M. W., Tayeh, B. A., A Comprehensive Review of Disputes Prevention and Resolution in Construction Projects. 2nd Conference for Civil Engineering Research Networks, Bandung, Indonesia, 2019.

[2] Chong, H. Y., Zin, R. M., Selection of Dispute Resolution Methods: Factor Analysis Approach. Engineering Construction and Architectural Management, 19(4), 428–443, 2012.

[3] Awwad, R., Barakat, B., Menassa, C., Understanding Dispute Resolution in the Middle East Region from Perspectives of Different Stakeholders. Journal of Management in Engineering, 32(6), 2016.

[4] Parikh, D., Joshi, G. J., Patel, D.A., Development of Prediction Models for Claim Cause Analyses in Highway Projects. Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 11(4), 2019.

[5] Ustuner, Y. A., Tas, E., An Examination of the Mediation Processes of International ADR Institutions and Evaluation of the Turkish Construction Professionals' Perspectives on Mediation. Eurasian Journal of Social Sciences, 7(4),11–27, 2019.

[6] Kisi, K. P., Lee, N., Kayastha, R., Kovel, J., Alternative Dispute Resolution Practices in International Road Construction Contracts. Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 12(2), 2020.

[7] Lee, C. K., Yiu, T. W., Cheung, S.O., Selection and Use of Alternative Dispute Resolution (ADR) in Construction Projects - Past and Future Research. International Journal of Project Management, 34(3), 494–507, 2016.

[8] Cheung, S. O., Au-Yeung, R. F., Wong, V. W. K, A CBR Based Dispute Resolution Process Selection System. International Journal of IT in Architecture Engineering and Construction, 2(2),129-145, 2004.

[9] Cheung, S. O., Suen, H. C. H., A Multi-Attribute Utility Model for Dispute Resolution Strategy Selection. Construction Management and Economics, 20(7), 557–568, 2002.

[10] Chou, J. S., Cheng, M. Y., Wu, Y. W., Improving Classification Accuracy of Project Dispute Resolution using Hybrid Artificial Intelligence and Support Vector Machine Models. Expert Systems with Applications, 40(6), 2263–2274, 2013.

[11] İlter, D., Opinions of Legal Professionals Regarding the Selection of Appropriate Resolution Method in Construction Disputes. RICS COBRA Annual Construction Building and Real Estate Research Conference, Paris, France, 2010.

[12] Siam, A., Ezzeldin, M., El-Dakhakhni, W., Machine Learning Algorithms for Structural Performance Classifications and Predictions: Application to Reinforced Masonry Shear Walls. Structures, 22, 252–265, 2019.

[13] Ayhan, M., Dikmen, I., Birgonul, M. T., Predicting the Occurrence of Construction Disputes using Machine Learning Techniques. Journal of Construction Engineering and Management, 147(4), 2021.

[14] Çevikbaş, M., Köksal, A., An Investigation of Litigation Process in Construction Industry in Turkey, Teknik Dergi, 29(6), 8715–8729, 2018.

[15] Pulket, T., Arditi, D., Construction Litigation Prediction System using Ant Colony Optimization. Construction Management and Economics, 27(3), 241–251, 2009.

[16] Harmon, K. M. J., Resolution of Construction Disputes: A Review of Current Methodologies. Leadership and Management in Engineering, 3(4), 187–201, 2003.

[17] King, L. S., Kamarazaly, M. A. H., Hashim, N., Yaakob, A. M., Man, N.H., Analysis on the Issues of Construction Disputes and the Ideal Dispute Resolution Method. Malaysian Construction Research Journal, 7(2), 153–165, 2019.

[18] Illankoon, I. M. C. S., Tam, W. V. Y., Le, N. K., Ranadewa, K. A. T. O., Causes of Disputes, Factors Affecting Dispute Resolution and Effective Alternative Dispute Resolution for Sri Lankan Construction Industry. International Journal of Construction Management, 1–11, 2019.

[19] Sinha, A. K., Jha, K. N., Dispute Resolution and Litigation in PPP Road Projects: Evidence from Select Cases. Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 12(1), 2020.

[20] Chou, J. S., Cheng, M. Y., Wu, Y. W., Pham, A. D., Optimizing Parameters of Support Vector Machine using Fast Messy Genetic Algorithm for Dispute Classification. Expert Systems with Applications, 41(8), 3955–3964, 2014.

[21] Chen, J. H., KNN Based Knowledge-Sharing Model for Severe Change Order Disputes in Construction. Automation in Construction, 17(6), 773–779, 2008.

[22] Liu, J., Li, H., Skitmore, M., Zhang, Y., Experience Mining Based on Case-Based Reasoning for Dispute Settlement of International Construction Projects. Automation in Construction, 97, 181–191, 2019.

[23] Chau, K. W., Application of PSO-Based Neural Network in Analysis of Outcomes of Construction Claims. Automation in Construction, 16(5), 642–646, 2007.

[24] Chen, J. H., Hsu, S. C., Hybrid ANN-CBR Model for Disputed Change Orders in Construction Projects. Automation in Construction, 17(1), 56–64, 2007.

[25] Arditi, D., Oksay, F. E., Tokdemir, O. B., Predicting the Outcome of Construction Litigation using Neural Networks. Computer-Aided Civil and Infrastructure Engineering, 13(2), 75–81, 1998.

[26] Arditi, D., Pulket, T., Predicting the Outcome of Construction Litigation using Boosted Decision Trees. Journal of Computing in Civil Engineering, 19(4), 387–393, 2005.

[27] Arditi, D., Pulket, T., Predicting the Outcome of Construction Litigation using an Integrated Artificial Intelligence Model. Journal of Computing in Civil Engineering, 24(1), 73–80, 2010.

[28] Pulket, T., Arditi, D., Universal Prediction Model for Construction Litigation. Journal of Computing in Civil Engineering, 23(3), 178–187, 2009.

[29] Mahfouz, T., Kandil, A., Davlyatov, S., Identification of Latent Legal Knowledge in Differing Site Condition (DSC) Litigations. Automation in Construction, 94, 104–111, 2018.

[30] Chaphalkar, N. B., Iyer, K. C., Patil, S. K., Prediction of Outcome of Construction Dispute Claims using Multilayer Perceptron Neural Network Model. International Journal of Project Management, 33(8), 1827–1835, 2015.

[31] Chou, J. S., Comparison of Multilabel Classification Models to Forecast Project Dispute Resolutions. Expert Systems with Applications, 39(11), 10202–10211, 2012.

[32] Chou, J. S., Hsu, S. C., Lin, C. W., Chang, Y. C., Classifying Influential Information to Discover Rule Sets for Project Disputes and Possible Resolutions. International Journal of Project Management, 34(8), 1706–1716, 2016.

[33] Ayhan, M., Development of Dispute Prediction and Resolution Method Selection Models for Construction Disputes. Ph.D. Thesis, Middle East Technical University, Ankara, 2019.

[34] Weisburd, D., Britt, C., Statistics in Criminal Justice, 3rd ed, Boston. Springer, 2007.

[35] Arasu, B. S., Seelan, B. J. B., Thamaraiselvan, N., A Machine Learning-Based Approach to Enhancing Social Media Marketing. Computers & Electrical Engineering, 86, 2020.

[36] Witten, H. W., Frank, E., Hall, M. A., Pal, C. J., Data Mining: Practical Machine Learning Tools and Techniques, 4th ed, Burlington. Morgan Kaufmann, 2016.

[37] Hssina, B., Merbouha, A., Ezzikouri, H., Erritali, M., A Comparative Study of Decision Tree ID3 and C4.5. International Journal of Advanced Computer Sciences and Applications, 4(2), 13–19, 2014.

[38] Febriantono, M. A., Pramono, S. H., Rahmadwati, R., Naghdy, G.), Classification of Multiclass Imbalanced Data using Cost-Sensitive Decision Tree C5.0. IAES International Journal of Artificial Intelligence, 9(1), 65–72, 2020.

[39] Cortes, C., Vapnik, V., Support-vector networks. Machine Learning, 20(3), 273–297, 1995.

[40] Alpaydin, E., Introduction to Machine Learning, 2nd ed, Cambridge. MIT Press, 2010.

[41] Hsu, C. W., Lin, C. J., A Comparison of Methods for Multiclass Support Vector Machines. IEEE Transactions on Neural Networks, 13(2), 415–425, 2002.

[42] Dietterich, T. G., Bakiri, G., Solving Multiclass Learning Problems via Error-Correcting Output Codes. Journal of Artificial Intelligence Research, 2, 263–286, 1994.

[43] McHugh, M. L., The Chi-Square Test of Independence. Biochemia Medica, 23(2), 143–149, 2013.

[44] Akoglu, H., User's Guide to Correlation Coefficients. Turkish Journal of Emergency Medicine, 18(3), 91–93, 2018.

[45] Pollock III, P.H., An SPSS Companion to Political Analysis, 4th ed, Washington, DC. CQ Press, 2011.

[46] Vanwinckelen, G., Blockeel, H., On Estimating Model Accuracy with Repeated Cross-Validation. 21st Belgian-Dutch Conference on Machine Learning, Ghent, Belgium, 2012.

[47] Lingard, H., Brown, K., Bradley, L., Bailey, C., Townsend, K., Improving Employees' Work-Life Balance in the Construction Industry: Project Alliance Case Study. Journal of Construction Engineering and Management, 133(10), 807–815, 2007.

[48] İlter, O., Çelik, T., Investigation of Organizational and Regional Perceptions on the Changes in Construction Projects. Teknik Dergi, 32(6), 2021.

[49] Yu, W. D., Hybrid Soft Computing Approach for Mining of Complex Construction Databases. Journal of Computing in Civil Engineering, 21(5), 343–352, 2007.