Response to Reviewer 2 Comments

Point 1: Although the authors add some literatures to indicate motivation the core problem of innovation and feasibility has not been solved. The neural networks are data-driven methods based on large amounts of data that indicate certain unknown laws. For this paper authors obtained the data by questionnaire which is usually suitable for statistical theory. Just from the sample, different from the process system strictly governed by internal and external laws, construction disputes are more complexity and specific. Hundreds of samples are too small which will lead to a lack of generalizability. In fact a construction dispute is not just a technical issue of fitting by historical experience that is a way out. It is difficult to be convinced by the result of neural network fitting (Even get a higher sample fitting accuracy than empiricism). It is suggested to seek some methods of objective evaluation such as game theory, key indicators evaluation, analytic hierarchy process and so on.

Response 1: Thank you for your comments. The authors would attempt to address this comment with three points. First, with respect to the algorithms for predicting construction disputes, the authors saught for two representative cases for empirical and mechanistic models each. Multilayer perceptron and Bayesian belief network were the two justified selection based on their historical performance in the literature. The suggested algorithms such as AHP and key indicator evaluation are not capable of prediction. Game theory combined with a system dynamic approach may be a good choice of mechanistic model. However, its capability of predicting construction disputes has not been tested in the literature while that of BBN parameter learning has already been tested and agreed in the literature. Therefore, the authors would like to stick to our previous choice of MLP and BBN to realise our resarch aim.

Second, as for the sample size, the authors added a new paragraph to present our power analyses results in Section 4.3:

"Power analyses was conducted to understand the representativeness of the sam-ple. For Pearson chi-square tests, the highest degree of freedom encountered in this re-search was 16. In order to control the type I error below 0.05 and the type II error below 0.01, minimum sample size would be 150. For the Pearson's chi-square tests and Pearson's correlation test respectively, a total of 150 and 195 samples were re-quired for the power of 99%. Input and output parameters for the power analyses were available in Table 3. Therefore, 222 samples would well meet the power re-quirements."
(Lines 316-322)

"Table 3. Power analyses for sample size calculation" **(Line 323)**

	Input		**Output**	
Pearson's Chi-square test	Parameters	Value	Parameters	Value
	Effect size	0.5	Noncentrality parameter λ	37.5
	α err prob	0.05	Critical χ^2	26.3
	Power (1-β err prob)	0.99	Total sample size	150
	Df	16	Actual power	0.9903
Pearson's correlation test	Tails	Two	Lower critical r	-0.14
	Correlation ρ H_1	0.3	Upper critical r	0.14
	α err prob	0.05	Total sample size	195
	Power (1-β err prob)	0.99	Actual power	0.9903
	Correlation ρ H_0	0		

In the end, with respect to the question if construction disputes can be predicted with a number of identified factors, the authors do not intend to explain the every aspect of construction disputes with our models but to capture the majority of the issue. The very first predictive model for construction disputes in the literature proposed by Diekmann and Girard (1995) and then those models with improved accuracy such as Chaphalkar et al. (2015), Yousefi et al. (2016) and Alaloul et al. (2018) basically followed the same logic that models were constructed over limited number of identiried critical factors. The authors added some further explanation in the introduction:

"The prediction of issues such as construction litigation, construction disputes, and construction claims is sometimes arguable as it is believed by a genre of researchers that such issues are far too complicated to be predicted by a model. It is expected in this paper that certain abstraction can be realised to reflect the majority of the issues if not all and consequently the empirical and mechanistic algorithms for prediction can be compared."