sustainability

MDPI

Article

Prediction of Risk Delay in Construction Projects Using a Hybrid Artificial Intelligence Model

Zaher Mundher Yaseen [1], Zainab Hasan Ali [2], Sinan Q. Salih [3,*] and Nadhir Al-Ansari [4,*]

[1] Sustainable Developments in Civil Engineering Research Group, Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam; yaseen@tdtu.edu.vn
[2] Civil Engineering Department, College of Engineering, University of Diyala, Baquba 32001, Iraq; zainabhasan222@gmail.com
[3] Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam
[4] Civil, Environmental and Natural Resources Engineering, Lulea University of Technology, 97187 Lulea, Sweden
* Correspondence: sinanq.salih@duytan.edu.vn (S.Q.S.); nadhir.alansari@ltu.se (N.A.-A.)

Received: 22 December 2019; Accepted: 14 February 2020; Published: 18 February 2020

Abstract: Project delays are the major problems tackled by the construction sector owing to the associated complexity and uncertainty in the construction activities. Artificial Intelligence (AI) models have evidenced their capacity to solve dynamic, uncertain and complex tasks. The aim of this current study is to develop a hybrid artificial intelligence model called integrative Random Forest classifier with Genetic Algorithm optimization (RF-GA) for delay problem prediction. At first, related sources and factors of delay problems are identified. A questionnaire is adopted to quantify the impact of delay sources on project performance. The developed hybrid model is trained using the collected data of the previous construction projects. The proposed RF-GA is validated against the classical version of an RF model using statistical performance measure indices. The achieved results of the developed hybrid RF-GA model revealed a good resultant performance in terms of accuracy, kappa and classification error. Based on the measured accuracy, kappa and classification error, RF-GA attained 91.67%, 87% and 8.33%, respectively. Overall, the proposed methodology indicated a robust and reliable technique for project delay prediction that is contributing to the construction project management monitoring and sustainability.

Keywords: delay sources; risk management; random forest-genetic algorithm; computer aid; construction project

1. Introduction

1.1. Research Background

The construction sector has a crucial role in improving the economics of developed countries [1]. The success of this sector is measured by time, cost and quality performance of construction projects. Prediction of construction durations represents a problem for both researchers and project managers. The construction process is subject to many factors and unpredicted variables that result from many sources. These sources prevent the completion of projects within the specified time and lead to a delay risk in the construction process [2,3].

Delay risk is considered as one of the major challenges tackled by construction firms [3,4]. Delay can be defined as an action or event that extends the time required to complete the project identified in a contract [3]. Project delay has an adverse effect on the project performance, which leads to cost overruns and productivity reduction. Its effect extends to include the owner, consultant and contractor

in terms of litigation, dispute and arbitration [5]. Delays are caused by many sources and factors such as the owner [6,7], designer [3,8], contractor [4,7], materials [4,7], project [7,8], labor [9] and external factors [3,10].

1.2. Literature Review

The prediction of project delay based on internal and external sources can help project managers to provide an accurate forecast of the project schedule, and this can assist a proactive management approach in the construction project [11]. Construction projects are dynamic and complex, included a huge number of project stockholders, feedback processes and non-linear relationships [12]. The existence of a delay problem is related to interdependent factors that affect the construction project and the complexity and uncertainty of construction activities. Thus, providing of an efficient tool for analyzing delay factors is key for estimating an accurate duration in construction projects [11].

By recalling previous studies, Chan (2001) used regression analysis to identify time–cost relationships for building projects in Malaysia [2]. This approach was developed for managers and owners to estimate the average time that is required for project delivery. Chan and Chan (2004) performed multiple regression exercises to analyze data related to the time performance of construction projects [13]. The results indicated that multiple regression was used as a useful method to predict time performance in construction projects. Rezaie et al. (2007) used Monte Carlo analysis to investigate the effects of uncertainties on the schedule performance [14]. The results revealed that this method is a good tool to simulate the relationship of uncertainties of construction projects, and represents many dimensions of the utility function. Hammad et al. (2008) developed a statistical model and sample tests to predict the cost and duration of construction projects in Jordan [15]. The proposed model showed reliable statistical results in the prediction of the time and cost performance of construction projects. Mohamed et al. (2009) introduced Analytical Hierarchy Processes (AHP) to predict time contingency in construction projects [16]. The model provided a reliable result in the estimation process. In another study, Abu Hammad et al. (2010) used statistical analysis to measure the risk impact on the time and cost of public building projects [17]. Dursun and Stoy (2011) modeled construction project duration by testing the statistical efficiency in classifying projects with respect to their locations [18]. The results demonstrated that the model was valid to select populations. Kokkaew and Wipulanusat (2014) presented a stochastic-based Monte Carlo approach for managing delay risk in construction projects [19]. This approach has been applied to two case studies, a commercial building and a build–operate–transfer (BOT) road. The results showed that this approach may enhance risk management in construction projects that are surrounded with uncertainties. However, although there have been several models introduced for delay risk simulation, these methods cannot explain and clarify the uncertainty and complexity of the construction process very well, and this can affect the prediction process of project duration.

Recently, artificial intelligence (AI) models have been widely applied in different fields of engineering and science [20]. Several studies used AI models in their research on construction project management. Elazouni (2006) Used an Artificial Neural Network (ANN) during the prequalification process before awarding a contract [21]. The model was used to classify contractors into several groups depending on their performance. The model demonstrated reliable result in representing contractors in four-dimensional space. Chao and Chien (2009) developed a neural network model to estimate the S-curve in a construction project, which represented by polynomial parameter [22]. The results displayed that the model can be useful for contractors and owners in the early planning phase of a construction project. Desai and Joshi (2010) used a decision tree to analyze labor productivity in construction projects [23]. Shin (2011) proposed the AdaBoost algorithm for the selection of a framework system in construction projects [24]. The algorithm was compared with an ANN model and the results revealed that the AdaBoost algorithm performed with better accuracy than the ANN model. In another study, Chou and Lin (2012) applied an ensemble method for disputing problems in public–private partnership projects [25]. The study used four regression and classification trees and two statistical

techniques to compare the performance of the utilized method. The study revealed that the ensemble models provided better accuracy that the individual models. Rudžianskaitė-Kvaraciejienė et al. (2015) Integrated a Random Forest algorithm with the modeling of public–private partnerships in infrastructure projects [26]. The method performed with good accuracy in the prediction process for public and private projects.

Heravi and Eslamdoost (2015) investigated the potential of an ANN model for the prediction of labor productivity in construction projects [27]. The results discovered that the ANN model showed better modeling of labor productivity. Gerassis et al. (2016) applied Bayesian networks to analyze the causes of accidents in embankment construction [28]. The study revealed that this method provided an accurate identification of embankment stability in civil engineering projects. By recalling the related literature review studies, AI model application is still a new methodology in the field of construction management research and delay risk prediction [29]. Few studies used AI models in risk prediction and classification.

Asadi et al. (2015) used a decision tree and a Naive Bayes model based on a questionnaire survey to predict delay in construction logistics. The authors evidenced the capacity of the decision tree has higher accuracy by 79.41% over the Naive Bayes model, which showed a lower accuracy value of 73.52% [30]. Naji et al. (2018) used a Bayesian decision tree model to predict the impact of contract changes on the time and quality performance of construction projects [31]. The model performed with good accuracy in the prediction process and caused an improvement in the project performance. Gondia et al. (2020) utilized Naive Bayes and decision tree models to predict the delay risk in construction projects. The study revealed the power of AI models in delay risk prediction and improving risk management strategies [11]. Based on the reported studies in the literature, the current research is established with the aim of providing a reliable methodology for delay risk prediction that will contribute to the baseline knowledge of construction management. Owing to the fact that standalone AI models experienced some limitations on tuning their internal parameters for an optimal learning process [32], the current study is adopted based on the integration of a nature-inspired optimization algorithm called Genetic Algorithm (GA) with a Random Forest (RF) model. The GA optimization approach was demonstrated as a reliable technique in tuning AI models for multiple engineering applications and thus it was selected for the current study [33–35].

1.3. Research Objectives

In the current study, the authors aim to explore and develop an effective tool to predict delay risk problems based on delay sources using previous construction projects data. The main contribution of the current investigation is to provide an accurate methodology that can assist in the prediction of future durations and monitor risk levels, based on these projects. This work can enhance a proactive approach in risk management. To achieve this aim, sources and factors of delay risk are extracted from the literature, and then related data to the delay risk problems are collected from previous construction projects. A questionnaire survey is adopted to measure the impact of various sources on the delay level in construction projects. Based on the complex nature of the construction process and the associated uncertainties of the delay sources, a hybrid model based on the integration of the Random Forest and Genetic Algorithm (RF-GA) is developed in order to analyze the data of completed previous projects. The performance of the developed model is studied statistically and discussed comprehensively. The potential of the proposed RF-GA model is validated against the classic RF model.

2. Research Methodology

2.1. Random Forest Model

The RF model was first developed by Breiman (2001) based on the combination of decision tree classifiers [36]. Each tree provides a prediction for the class label and the algorithm selects the classes that have the most choices. Random Forest is a very popular tool that uses the bootstrapping

method to train dataset samples and construct multiple random trees [37]. The algorithm gained significant important because it is invariant under scaling and it is robust to the inclusion of irrelevant features [38]. Several studies examined the application of Random Forest in engineering applications and demonstrated its feasibility in prediction processes [26,34,35]. Under the bootstrapping method, the data during the training phase are selected randomly and independently to develop an RF model, and the data that are not involved in the selection process are named "out-of-bag" [39]. The capacity of the random forest has been approved by several engineering problems such as [40,41]. During this process, the out-of-bag data are changed and the prediction error is measured to estimate the importance of input variables [41,42]. In the RF algorithm, overfitting does not occur due to large numbers of trees and the choice of the right type of random variables leads to accurate classification. Random Forests contain several parameters that need to be optimized, such as number of trees, minimum gain and maximum tree depth. In this study, these parameters were optimized by a genetic algorithm.

2.2. The Hybrid RF-GA Model

Genetic algorithm is a popular technique used to optimize problems in complex systems based on natural selection [43]. To solve a problem in a GA algorithm, random solutions are generated, and then the selection of the population is done to develop the model. The new solutions are developed by using selection, crossover and mutation. A string of bits or chromosomes is used to represent the solutions in the GA model. The position of bits is called the gene, and the gene contains many values that are named alleles. GA has been widely applied in different fields, such as image processing, pattern recognition and controlling systems [44]. The GA model was used in different research in construction management such as resources optimization [45], project scheduling [46,47], optimizing of time and cost in construction projects [48] and dispute classification [49]. In the present study, a GA model is presented to optimize parameters of the RF model, including number of trees, minimum gain and maximum tree depth. The description of the hybrid FR-GA model is illustrated in Figure 1.

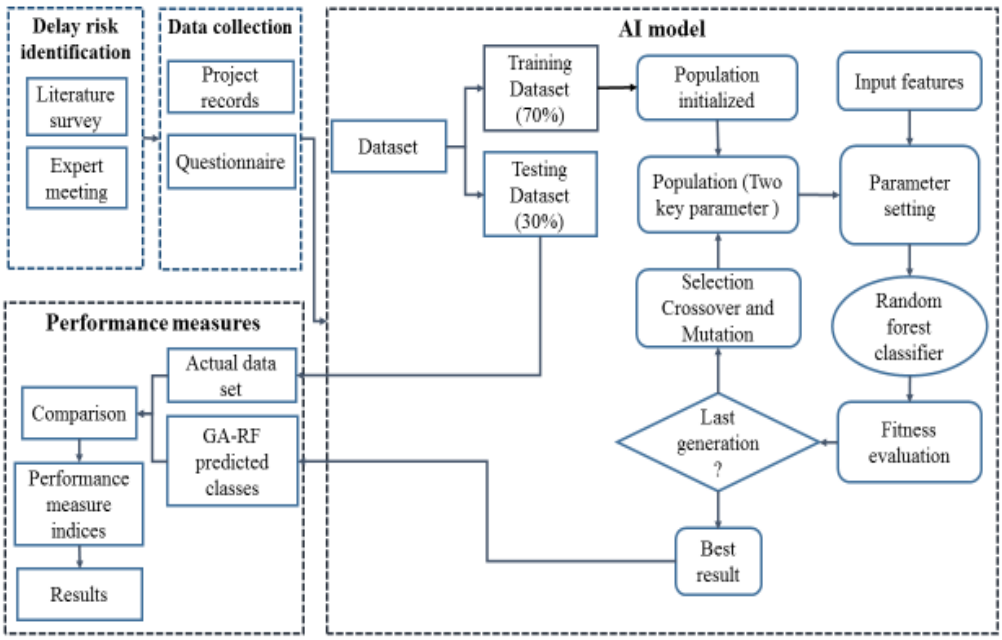


Figure 1. Hybrid artificial intelligence model Random Forest and Genetic Algorithm (RF-GA) structure.

2.3. Identification of Delay Sources and Factors in Construction Projects

The most important factors that affect delays in construction projects were identified from a literature survey, and these factors were categorized into different sources. These sources included owner, designer, contractor, project, material, equipment, labor and external factors. To obtain more

information on the delay problems and their factors in construction projects, interviews were held with 15 experts in construction work [11,50]. By this interview, the identified sources and factors and their relevance to the construction industry were confirmed. Based on the reviews and literature review, the most important delay factors and their sources were identified as shown in Table 1.

Table 1. Identified factors and sources in construction projects.

Delay Source	Delay Factors
1. Owner	1.1 Owner financial problems
	1.2 Payment delay by the owner
	1.3 Choosing of inefficient design team
	1.4 Inadequate experience of the owner
	1.5 Issuing of change orders by the owner
	1.6 Delay in location delivery to the contractor
	1.7 Choice of inefficient contractor
	1.8 Delay in decision making procedure
2. Designer	2.1 Inadequate experience of design members
	2.2 Delay in the preparation of design documents
	2.3 Defects in the design and ambiguity of design drawings
3. Contractor	3.1 Ineffective project planning
	3.2 Financial contractor difficulties
	3.3 Inadequacy of contractor
	3.4 Rework due to defects in executed work
	3.5 Ineffective supervision and site management
	3.6 Many changes in subcontractor parties
	3.7 Poor communication between contractor and project parties
4. Project	4.1 Awarding the contract to an inadequate contractor
	4.2 Disputes between project parties
	4.3 Period of contract is very short
	4.4 Errors in contract documents
5. Material	5.1 Deficiency of materials in the market
	5.2 Delay in supplying materials
	5.3 Ineffective quality of materials
	5.4 Poor storage of materials
6. Equipment	6.1 Poor efficiency of equipment
	6.2 Unsuitable type of equipment
7. Labor	7.1 Poor labor productivity
	7.2 Inadequacy of workforce
	7.3 Lack of labor
8. External factors	8.1 Political situation and terrorism
	8.2 Inflation
	8.3 Legislation changes in the country
	8.4 Unpredicted surface conditions
	8.5 Neighbor problems
	8.6 Bad weather conditions

2.4. Data Collection

The compiled data included 40 completed projects that had different degrees of time overrun. These projects were executed in Diyala city, Iraq. The collected data included historical documents of previous projects that were investigated to extract the measure of risk delay in construction projects. These documents included contract documents, specifications, change orders records and schedule baselines. To complete data collection, a questionnaire survey was arranged and constructed. Each questionnaire form contained a construction project and another nine variables. The first variable represented the delay level, and the other eight variables referred to the risk delay sources in the construction project. Each risk source was given scores depending on two scales. The first scale was the probability of risk to occur in the construction project and the second related to the impact of sources on the delay of the construction project, as shown in Table 2. The overall risk impact was evaluated by multiplying the two scales [3,43,44].

Table 2. Scales of probability and impact of risk delay in construction projects.

Scale	Probability	Impact
Very low	0.1	0.05
Low	0.3	0.1
Medium	0.5	0.2
High	0.7	0.4
Very high	0.9	0.8

The probability and impact of the variables were measured by using a five-point Likert scale with measures form very low to very high level [51]. The input variables were classified as: very low, low, medium, high and very high. The output variable (delay level) was also classified into three class measures. This method resulted in three categories of delay level that reduced the bias during the execution of the artificial intelligence model. Delay level was categorized as: <50% delay, 50%–100% delay and >100% delay. The questionnaire was allocated to a pilot study to measure the questionnaire reliability and to investigate the problems and determine the items that are more confusing than the others. The authors selected 40 parties for the pilot study as the size of the study was ranged between 30 and 50 parties [52]. To confirm the questionnaire reliability, Cronbach's alpha was adopted, and in this study the value of the alpha coefficient is 91.8%. The result of Cronbach's alpha confirms the reliability of the questionnaire.

2.5. Model Development Procedure

The questionnaire was distributed to 300 experts who worked in the collected projects. The experts were involved in different parties, which include client, engineer and the other experts of these projects. The collected projects were divided into two phases: 70% of the total projects (28 projects) were used for the training phase and 30% (12 projects) for the testing phase of the hybrid intelligence model. The genetic algorithm was applied to optimize the Random Forest classifier, and the hybrid RF-GA was used to predict the time performance in construction projects based on their risk levels by dividing the projects into a class label describing the predicted time delay. At first, the model describes the data with a set of records and variable values. The rows represent the individual projects and the columns represent the value of each project. The input variables included owner, designer, contractor, project, material, equipment, labor and external factors. The hybrid model was used for learning the dataset in order to predict the time delay at different levels of time overrun. The structure of the developed model is described in Figure 2.

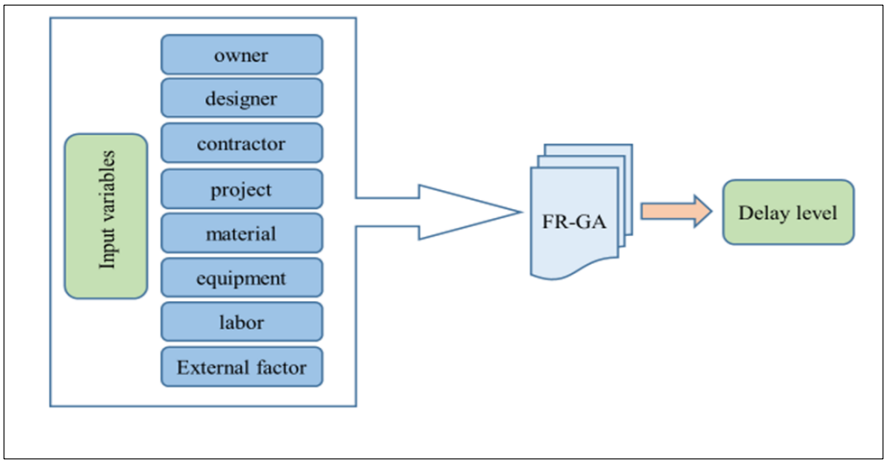


Figure 2. Structure of the delay prediction model.

2.6. Model Performance Measures

The performance of the predicted model was evaluated by using class performance and overall performance measures. Class performance was measured by precision, sensitivity and specificity [53,54]. The overall performance of the predicted model was evaluated by accuracy, classification error and kappa statistics. The kappa coefficient (k) was used in statistics to measure the quality of an item based on inter classifier agreement [55,56]. The equations of performance measures are explained as follows:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

$$Classification\ error = 1 - accuracy \tag{5}$$

$$k = \frac{p_o - p_e}{1 - p_e} \tag{6}$$

where:

TP means the number of positive classes that are correctly recognized by the algorithm;
FP represents the number of positive classes that are incorrectly classified by the algorithm;
TN means the number of negative classes that are correctly predicted by the algorithm;
FN represents the number of negative classed that are incorrectly recognized by the algorithm;
P_o means the observed agreement between rates; and
P_e represents the probability of chance agreement.

3. Results and Discussion

Analysis of collected data based on 40 projects was conducted to identify the sources of delay problems effectively. The properties of the complied data and the distribution of delay sources among the construction project are presented in Figures 3 and 4.

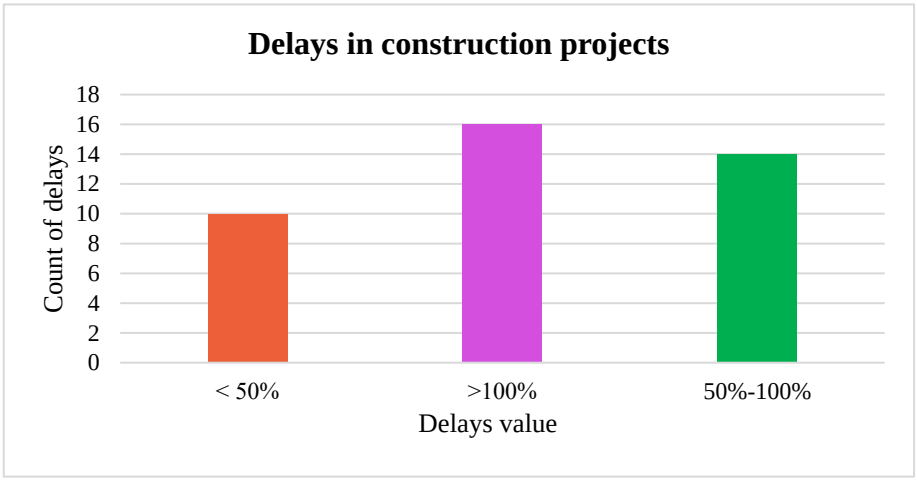
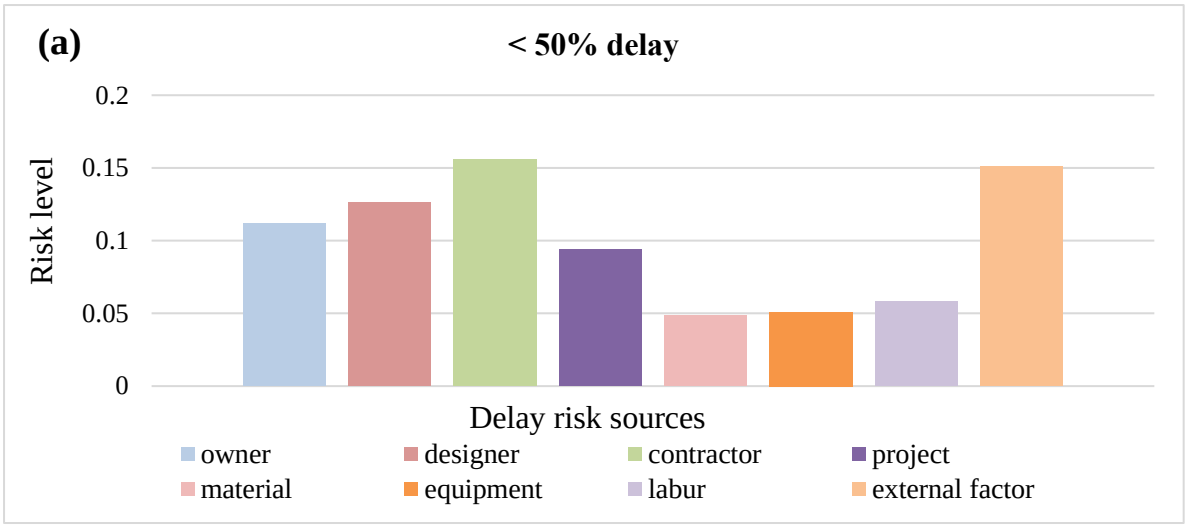


Figure 3. Frequency count of delays in construction projects.



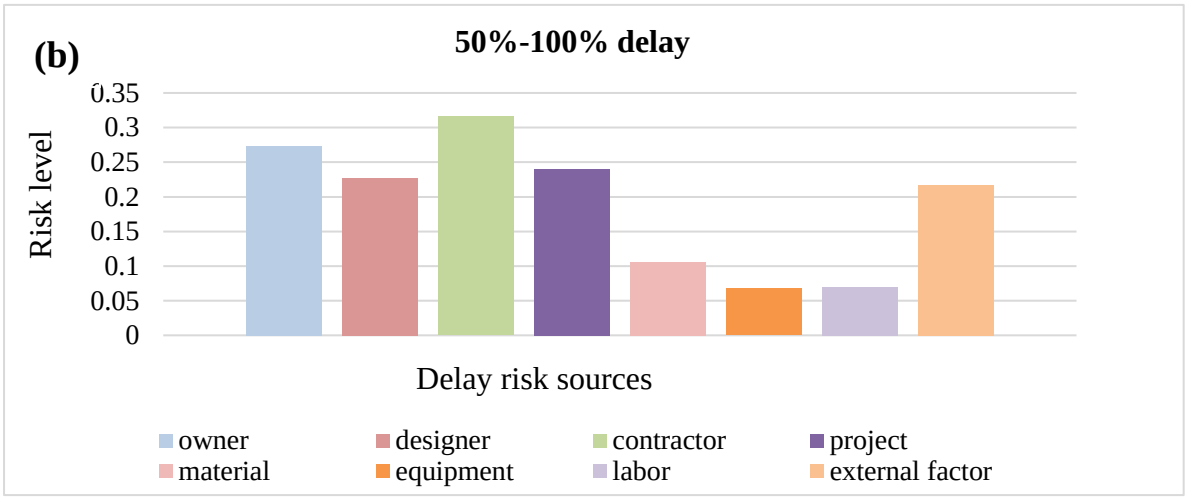


Figure 4. *Cont.*

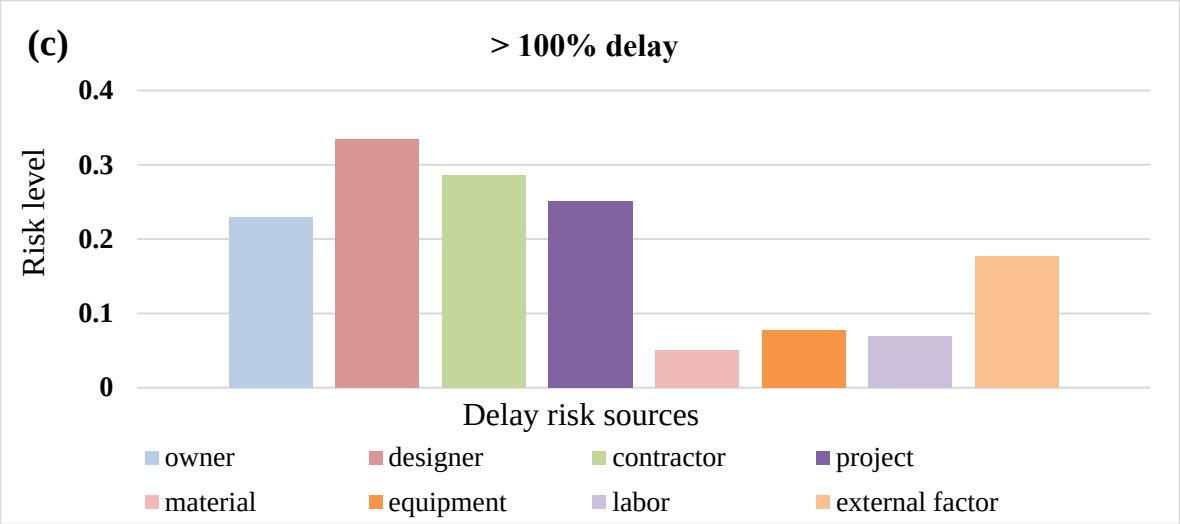


Figure 4. (**a**) Distribution of delay sources among projects with class <50% delay, (**b**) distribution of delay sources among projects with class 50%–100% delay, (**c**) distribution of delay sources among projects with class >100% delay.

Based on the reported results, Figure 3 shows the counts of projects with a <50% delay, 50%–100% delay and >100% delay were 10 (25%), 14 (35%) and 16 (40%), respectively. It can be seen that a high percentage of projects belongs to the class of >100% delay. Figure 4 demonstrates the distribution of delay sources among each class of delay problem, which was obtained from the historical records, pilot study and distributed questionnaire. These outcomes revealed the delay sources values of contractor, owner, designer, project and external factors have a higher impact that the other delay sources. Owner, designer, contractor and project are represented as the internal risk sources that have an impact on the project delay. External factors can be discussed by the special circumstances that are experienced in the studied region "Iraq" in a manner that severely affected the construction industry. These conditions have an enormous impact on the project stockholder and project performance. These conditions resulted in the stumbling and failure of many projects in the construction sector. On the other hand, the application of a robust predictive model can contribute to estimating an accurate duration in construction projects and analyzing delay risk sources that arise from the complex and dynamic nature of construction sector.

The statistical performance of the training and testing datasets of the proposed hybrid RF-GA model were evaluated based on the model performance measures against the classical Random Forest classifier. The performance measure metrics were evaluated based on the confusion matrix of the two classifiers. The confusion matrix is described in the performance of the classification model. The confusion matrix of the RF and RF-GA are displayed in Tables 3 and 4.

Table 3. Confusion matrix from RF classifier.

	Actual Class			
Predicted Class	**<50%**	**50%–100%**	**>100%**	**Total**
<50%	2	0	1	3
50%–100%	0	0	3	3
>100%	1	4	1	6
Total	3	4	5	

Table 4. Confusion matrix from RF-GA classifier.

	Actual Class			
Predicted Class	**<50%**	**50%–100%**	**>100%**	**Total**
<50%	2	0	0	2
50%–100%	1	5	0	6
>100%	0	0	4	4
Total	3	5	4	

The columns in the confusion matrix represent the actual classification within each class, while the rows correspond to the number of the predicted class. The correct predictors are located on the diagonal of the matrix. The confusion matrix of a high-performance model contains large numbers in its diagonal and the zero numbers outside the diagonal. The performance of the hybrid RF-GA and RF models during the training and testing phases was evaluated. Precision, sensitivity, specificity, accuracy, classification error and kappa statistics were computed and are presented in Tables 5 and 6.

Table 5. Comparison of two classifiers based on performance measures for the training phase (70%).

	RF			**RF-GA**		
Performance Index	**<50% Delay**	**50%–100% Delay**	**>100% Delay**	**<50% Delay**	**50%–100% Delay**	**>100% Delay**
Precision	87.5	100	90	87.5	100	100
Sensitivity	87.5	83.33	90	100	91.67	90
Specificity	95	100	94.44	95.2	100	100
Accuracy		92.86			96.43	
Classification error		7.41			3.57	
Kappa		89.2			94.6	

Table 6. Comparison of two classifiers based on performance measures for the testing phase (30%).

	RF			**RF-GA**		
Performance Index	**<50% Delay**	**50%–100% Delay**	**>100% Delay**	**<50% Delay**	**50%–100% Delay**	**>100% Delay**
Precision	66.67	100	90	66.67	83.33	100
Sensitivity	50	50	90	80	100	80
Specificity	87.5	100	94.44	71.4	85.7	100
Accuracy		75			91.67	
Classification error		25			8.33	
Kappa		62.5			87	

Tables 5 and 6 demonstrate the comparison of the RF-GA and RF models based on performance measures for the overall and class performance. Based on these results, the attained values of accuracy, classification error, and Kappa for RF-GA were 96.43%, 3.57% and 94.6%, respectively for the training phase; whereas the RF model provided an accuracy value of 92.86%, a classification error of 7.41% and a Kappa statistics value of 89.2%. It can be noticed that the both models gave good results for the training phase. Based on the results of testing phase, RF-GA revealed a better performance than the RF

model. The provided values of accuracy, classification error and Kappa of RF-GA are 91.67%, 8.3%, and 87%, respectively.

With regards to performance measures, the RF-GA model exhibited a good performance in the prediction of delay in the construction sector. Based on the training phase, RF-GA achieved the minimum values of precision, sensitivity and specificity of 87.5, 90 and 95.2, respectively. The lowest values of RF in terms of precision, sensitivity and specificity were 87.5, 83.33 and 94.44, respectively. Based on the comparison between the two classifiers, it can be concluded that the RF-GA model outperformed the feasibility of the classical RF model in both the training and testing performances. Tables 5 and 6 revealed the superiority of the RF-GA classifier in terms of accuracy, classification error and Kappa statistics. This can be explained as due to the potential of the integration of the nature-inspired optimization algorithm (i.e., GA) that assisted in providing reliable hyperparameters optimization and thus attained a reliable learning process. The RF-GA model also provided higher values of precision, sensitivity and specificity in comparison with the RF model.

The RF-GA classifier showed an impressive performance in terms of overall and class measure indices. These results can be discussed by the ability of the genetic algorithm in solving optimization problems depending on the chromosome approach, and its capacity to solve the problems while dealing with multiple solutions [57]. It is even better to validate the current research results with the reported research over the literature. As compared with the previous results, it can be inferred that the RF-GA model demonstrated remarkable prediction superiority in comparison with the previous established studies as reported in Table 7. The capacity of the RF-GA model was compared with the best outcomes. The RF-GA model exceeded all of the reported related literature.

Table 7. Validation of the current research results against the reported related literature studies.

Author	Methods	Results
[30]	Questionnaire survey, decision tree and Naive Bayes	Accuracy of decision tree 79.41% is higher than Naive Bayes by 5.81%
[31]	Questionnaire and Bayesian decision tree	Bayesian decision tree gained an accuracy of 86.7%
[11]	Records of construction projects, meeting with experts, decision tree and Naive Bayes	Accuracy of Naive Bayes, 51.2%, is higher than decision tree by 4%
The current study	Records of construction projects, meetings and questionnaire survey, classical Random Forest, hybrid genetic Random Forest	Accuracy of genetic Random Forest, 91.76%, is higher than classical Random Forest by 16.67%

To summarize, a proactive management approach involves the identification of new risk delay sources and the monitoring of the sources that arise during the project lifecycle. As a result, the proposition of a reliable and robust methodology as an analysis tool that is able to mimic and comprehend the dynamic input variables is highly needed for this purpose. Hence, and based on the established methodology of the current research, the potential of the RF-GA model to be modified and set up for project duration prediction though the project lifecycle was evidenced. The RF-GA model was successfully developed for the investigated dynamic project delay risk prediction.

4. Conclusions

In this present study, an analysis tool that is capable of predicting the delay level in construction projects based on delay sources was proposed. To meet this goal, two approaches were adopted in this study. First, delay sources and factors were collected from a literature review and identified by an expert meeting. Data that are related to delay levels were compiled from 40 construction projects that are located in Diyala city, Iraq. The collected data included historical records of previous projects

that were investigated, and in order to extract the measure of delay risk in construction projects a questionnaire was prepared and distributed to 300 experts so as to extract the information about delay sources in construction projects. Risk sources were measured by computing the probability and the impact of each source. An analysis of data results and distribution of delay sources among the collected previous projects was implemented in order to better understand delay factors in the construction sector. Secondly, a hybrid RF-GA model was developed to deal with the complex and dynamic nature of data in the construction sector. The RF-GA model was evaluated by performance measure indices and compared with the classical RF model. Based on the analysis results, RF-GA revealed a better performance than the RF model. The RF-GA provided values of accuracy, classification error and Kappa were 91.67%, 8.3%, and 87%, respectively. These results reflect the ability of the model to handle the nonlinearity and complexity of data in the construction sector. The results also revealed the capability of the genetic algorithm in solving problems with multiple solutions.

Author Contributions: Conceptualization, Z.M.Y. and Z.H.A.; Data curation, Z.M.Y.; Formal analysis, Z.M.Y. and N.A.-A.; Funding acquisition, N.A.-A.; Investigation, Z.M.Y., Z.H.A. and S.Q.S.; Methodology, Z.H.A. and S.Q.S.; Project administration, S.Q.S.; Resources, Z.H.A.; Software, Z.H.A. and S.Q.S.; Supervision, N.A.-A.; Validation, Z.H.A. and S.Q.S.; Visualization, Z.H.A. and S.Q.S.; Writing—original draft, Z.M.Y., Z.H.A., S.Q.S. and N.A.-A.; Writing—review and editing, Z.M.Y. and N.A.-A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Vorakulpipat, C.; Rezgui, Y.; Hopfe, C.J. Value creating construction virtual teams: A case study in the construction sector. *Autom. Constr.* **2010**, *19*, 142–147. [CrossRef]
2. Chan, A.P. Time-cost relationship of public sector projects in Malaysia. *Int. J. Proj. Manag.* **2001**, *19*, 223–229. [CrossRef]
3. Assaf, S.A.; Al-Hejji, S. Causes of delay in large construction projects. *Int. J. Proj. Manag.* **2006**, *24*, 349–357. [CrossRef]
4. Sambasivan, M.; Soon, Y.W. Causes and effects of delays in Malaysian construction industry. *Int. J. Proj. Manag.* **2007**, *25*, 517–526. [CrossRef]
5. Aibinu, A.; Jagboro, G. The effects of construction delays on project delivery in Nigerian construction industry. *Int. J. Proj. Manag.* **2002**, *20*, 593–599. [CrossRef]
6. Odeh, A.M.; Battaineh, H.T. Causes of construction delay: traditional contracts. *Int. J. Proj. Manag.* **2002**, *20*, 67–73. [CrossRef]
7. Fugar, F.D.; Agyakwah-Baah, A.B. Delays in Building Construction Projects in Ghana. *Australas. J. Constr. Econ. Build.* **2010**, *10*, 103–116. [CrossRef]
8. Aziz, R.F. Ranking of delay factors in construction projects after Egyptian revolution. *Alex. Eng. J.* **2013**, *52*, 387–406. [CrossRef]
9. Al-Momani, A.H. Construction delay: a quantitative analysis. *Int. J. Proj. Manag.* **2000**, *18*, 51–59. [CrossRef]
10. Jing, W.; Naji, H.I.; Zehawi, R.N.; Ali, Z.; Al-Ansari, N.; Yaseen, Z.M. System Dynamics Modeling Strategy for Civil Construction Projects: The Concept of Successive Legislation Periods. *Symmetry* **2019**, *11*, 677. [CrossRef]
11. Gondia, A.; Siam, A.; El-Dakhakhni, W.; Nassar, A.H. Machine Learning Algorithms for Construction Projects Delay Risk Prediction. *J. Constr. Eng. Manag.* **2020**, *146*, 04019085. [CrossRef]
12. Mahamid, I.; Bruland, A.; Dmaidi, N. Causes of delay in road construction projects. *J. Manag. Eng.* **2011**, *28*, 300–310. [CrossRef]
13. Chan, A.P.C.; Chan, D.W. Developing a benchmark model for project construction time performance in Hong Kong. *Build. Environ.* **2004**, *39*, 339–349. [CrossRef]
14. Rezaie, K.; Amalnik, M.; Gereie, A.; Ostadi, B.; Shakhseniaee, M. Using extended Monte Carlo simulation method for the improvement of risk management: Consideration of relationships between uncertainties. *Appl. Math. Comput.* **2007**, *190*, 1492–1501. [CrossRef]

15. Hammad, A.A.A.; Ali, S.M.A.; Sweis, G.J.; Bashir, A. Prediction model for construction cost and duration in Jordan. *Jordan J. Civ. Eng.* **2008**, *2*, 250–266.
16. Mohamed, D.; Srour, F.; Tabra, W.; Zayed, T. A prediction model for construction project time contingency. In *Construction Research Congress 2009: Building a Sustainable Future*; ASCE: Reston, VA, USA, 2009; pp. 736–745.
17. Abu Hammad, A.; Ali, S.M.A.; Sweis, G.J.; Sweis, R. Statistical Analysis on the Cost and Duration of Public Building Projects. *J. Manag. Eng.* **2010**, *26*, 105–112. [CrossRef]
18. Dursun, O.; Stoy, C. Time–cost relationship of building projects: statistical adequacy of categorization with respect to project location. *Constr. Manag. Econ.* **2011**, *29*, 97–106. [CrossRef]
19. Kokkaew, N.; Wipulanusat, W. Completion delay risk management: A dynamic risk insurance approach. *KSCE J. Civ. Eng.* **2014**, *18*, 1599–1608. [CrossRef]
20. Brunette, E.S.; Flemmer, R.C.; Flemmer, C.L. A review of artificial intelligence. In Proceedings of the 2009 4th International Conference on Autonomous Robots and Agents, Wellington, New Zealand, 10–12 February 2009; pp. 385–392.
21. Elazouni, A. Classifying Construction Contractors Using Unsupervised-Learning Neural Networks. *J. Constr. Eng. Manag.* **2006**, *132*, 1242–1253. [CrossRef]
22. Chao, L.-C.; Chien, C.-F. Estimating Project S-Curves Using Polynomial Function and Neural Networks. *J. Constr. Eng. Manag.* **2009**, *135*, 169–177. [CrossRef]
23. Desai, V.S.; Joshi, S. Application of decision tree technique to analyze construction project data. In *Communications in Computer and Information Science, Proceedings of the International Conference on Information Systems, Technology and Management, Bangkok, Thailand, 11–13 March 2010*; Springer: Piscataway, NJ, USA, 2010; pp. 304–313.
24. Shin, Y.-S. Formwork System Selection Model for Tall Building Construction Using the Adaboost Algorithm. *J. Korea Inst. Build. Constr.* **2011**, *11*, 523–529. [CrossRef]
25. Chou, J.-S.; Lin, C. Predicting disputes in public-private partnership projects: Classification and ensemble models. *J. Comput. Civ. Eng.* **2012**, *27*, 51–60. [CrossRef]
26. Rudžianskaitė–Kvaraciejienė, R.; Apanaviciene, R.; Gelzinis, A. modelling the effectiveness of ppp road infrastructure projects by applying random forests. *J. Civ. Eng. Manag.* **2015**, *21*, 290–299. [CrossRef]
27. Heravi, G.; Eslamdoost, E. Applying Artificial Neural Networks for Measuring and Predicting Construction-Labor Productivity. *J. Constr. Eng. Manag.* **2015**, *141*, 04015032. [CrossRef]
28. Gerassis, S.; Martín, J.E.; García, J.T.; Saavedra, A.; Taboada, J. Bayesian decision tool for the analysis of occupational accidents in the construction of embankments. *J. Constr. Eng. Manag.* **2016**, *143*, 4016093. [CrossRef]
29. Bilal, M.; Oyedele, L.; Qadir, J.; Munir, K.; Ajayi, S.; Akinade, O.; Owolabi, H.A.; Alaka, H.A.; Pasha, M. Big Data in the construction industry: A review of present status, opportunities, and future trends. *Adv. Eng. Inform.* **2016**, *30*, 500–521. [CrossRef]
30. Asadi, A.; Alsubaey, M.; Makatsoris, C. A machine learning approach for predicting delays in construction logistics. *Int. J. Adv. Logist.* **2015**, *4*, 115–130. [CrossRef]
31. Hassan, Z.; Ibrahim, A.M.; Naji, H. Evaluation of Legislation Adequacy in Managing Time and Quality Performance in Iraqi Construction Projects- a Bayesian Decision Tree Approach. *Civ. Eng. J.* **2018**, *4*, 993. [CrossRef]
32. Yaseen, Z.; Mohtar, W.H.M.W.; Ameen, A.M.S.; Ebtehaj, I.; Razali, S.F.M.; Bonakdari, H.; Salih, S.Q.; Al-Ansari, N.; Shahid, S. Implementation of univariate paradigm for streamflow simulation using hybrid data-driven model: Case study in tropical region: Implementation of univariate paradigm for streamflow simulation using hybrid data-driven model: Case study in tropical region. *IEEE Access* **2019**, *7*, 74471–74481. [CrossRef]
33. Chou, J.-S.; Pham, A.-D. Hybrid computational model for predicting bridge scour depth near piers and abutments. *Autom. Constr.* **2014**, *48*, 88–96. [CrossRef]
34. Yaseen, Z.M.; Ehteram, M.; Hossain, S.; Chow, M.F.; Koting, S.; Mohd, N.S.; Jaafar, W.B.; Afan, H.A.; Hin, L.S.; Zaini, N.; et al. A Novel Hybrid Evolutionary Data-Intelligence Algorithm for Irrigation and Power Production Management: Application to Multi-Purpose Reservoir Systems. *Sustainability* **2019**, *11*, 1953. [CrossRef]

35. Yaseen, Z.M.; Ebtehaj, I.; Kim, S.; Sanikhani, H.; Asadi, H.; Ghareb, M.I.; Bonakdari, H.; Mohtar, W.H.M.W.; Al-Ansari, N.; Shahid, S. Novel Hybrid Data-Intelligence Model for Forecasting Monthly Rainfall with Uncertainty Analysis. *Water* **2019**, *11*, 502. [CrossRef]
36. Breiman, L.; Cutler, A. State of the art of data mining using Random forest. In Proceedings of the Salford Data Mining Conference, San Diego, CA, USA, 24–25 May 2012.
37. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
38. Ruppert, D. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *J. Am. Stat. Assoc.* **2004**, *99*, 567. [CrossRef]
39. Catani, F.; Lagomarsino, D.; Segoni, S.; Tofani, V. Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 2815–2831. [CrossRef]
40. Naghibi, S.A.; Ahmadi, K.; Daneshi, A. Application of Support Vector Machine, Random Forest, and Genetic Algorithm Optimized Random Forest Models in Groundwater Potential Mapping. *Water Resour. Manag.* **2017**, *31*, 2761–2775. [CrossRef]
41. Alipour, M.; Harris, D.K.; Barnes, L.E.; Ozbulut, O.; Carroll, J. Load-Capacity Rating of Bridge Populations through Machine Learning: Application of Decision Trees and Random Forests. *J. Bridg. Eng.* **2017**, *22*, 04017076. [CrossRef]
42. Liaw, A.; Wiener, M. Classification and regression by randomforest. *R News* **2002**, *2*, 18–22.
43. Holland, J.H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*; MIT Press: Cambridge, MA, USA, 1992.
44. Azar, A.T.; Vaidyanathan, S. *Computational Intelligence Applications in Modeling and Control*; Springer: Piscataway, NJ, USA, 2015.
45. Kandil, A.; El-Rayes, K. Parallel Genetic Algorithms for Optimizing Resource Utilization in Large-Scale Construction Projects. *J. Constr. Eng. Manag.* **2006**, *132*, 491–498. [CrossRef]
46. Senouci, A.; Eldin, N.N. Use of Genetic Algorithms in Resource Scheduling of Construction Projects. *J. Constr. Eng. Manag.* **2004**, *130*, 869–877. [CrossRef]
47. Long, L.D.; Ohsato, A. A genetic algorithm-based method for scheduling repetitive construction projects. *Autom. Constr.* **2009**, *18*, 499–511. [CrossRef]
48. Rogalska, M.; Bozejko, W.; Hejducki, Z. Time/cost optimization using hybrid evolutionary algorithm in construction project scheduling. *Autom. Constr.* **2008**, *18*, 24–31. [CrossRef]
49. 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 Syst. Appl.* **2014**, *41*, 3955–3964. [CrossRef]
50. Xia, N.; Zhong, R.; Wu, C.; Wang, X.; Wang, S. Assessment of Stakeholder-Related Risks in Construction Projects: Integrated Analyses of Risk Attributes and Stakeholder Influences. *J. Constr. Eng. Manag.* **2017**, *143*, 04017030. [CrossRef]
51. Ismail, I.; Memon, A.H.; Rahman, I.A. Expert opinion on risk level for factors affecting time and cost overrun along the project lifecycle in Malaysian construction projects. *Int. J. Constr. Technol. Manag.* **2013**, *1*, 2289.
52. Thomas, S.J. *Using Web and Paper Questionnaires for Data-Based Decision Making: From Design to Interpretation of the Results*; Corwin Press: Thousand Oaks, CA, USA, 2004.
53. Helmer, G.; Wong, J.; Honavar, V.G.; Miller, L. Automated discovery of concise predictive rules for intrusion detection. *J. Syst. Softw.* **2002**, *60*, 165–175. [CrossRef]
54. Davis, J.; Goadrich, M. The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd International Conference on Machine Learning*; ACM: New York, NY, USA, 2006; pp. 233–240.
55. Smeeton, N.C. Early history of the kappa statistic. *Biometrics* **1985**, *41*, 795.
56. Pontius, R.G.; Millones, M. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *Int. J. Remote Sens.* **2011**, *32*, 4407–4429. [CrossRef]
57. Tabassum, M.; Mathew, K. A genetic algorithm analysis towards optimization solutions. *Int. J. Digit. Inf. Wirel. Commun.* **2014**, *4*, 124–142. [CrossRef]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).