Article

Predicting the Impact of Construction Rework Cost Using an Ensemble Classifier

Fatemeh Mostofi [1][ID]**, Vedat Toğan** [1,*][ID]**, Yunus Emre Ayözen** [2] **and Onur Behzat Tokdemir** [3][ID]

1   Civil Engineering Department, Karadeniz Technical University, 61080 Trabzon, Türkiye
2   Strategy Development Department, Ministry of Transport and Infrastructure, 06338 Ankara, Türkiye
3   Civil Engineering Department, Istanbul Technical University, 34469 Istanbul, Türkiye
*   Correspondence: togan@ktu.edu.tr

Abstract: Predicting construction cost of rework (COR) allows for the advanced planning and prompt implementation of appropriate countermeasures. Studies have addressed the causation and different impacts of COR but have not yet developed the robust cost predictors required to detect rare construction rework items with a high-cost impact. In this study, two ensemble learning methods (soft and hard voting classifiers) are utilized for nonconformance construction reports (NCRs) and compared with the literature on nine machine learning (ML) approaches. The ensemble voting classifiers leverage the advantage of the ML approaches, creating a robust estimator that is responsive to underrepresented high-cost impact classes. The results demonstrate the improved performance of the adopted ensemble voting classifiers in terms of accuracy for different cost impact classes. The developed COR impact predictor increases the reliability and accuracy of the cost estimation, enabling dynamic cost variation analysis and thus improving cost-based decision making.

Keywords: construction rework; cost estimation; nonconformance report; voting classifier; ensemble learning; machine learning

1. Introduction

A successful construction project is delivered on time and within budget, conforming to the specified quality. To achieve this, potential construction errors and violations are managed by applying an adequate construction quality management (CQM) system. An indispensable procedure within CQM is quality control (QC), which involves ensuring construction activity delivery at a specified standard, appraising its conformance, and maintaining continuous quality improvement. In construction projects, the arrays of errors, omissions, negligence, changes, failures, and violations resulting from poor management, communication, and coordination, or the materialization of potential risks are solved through rework. Thus, it is necessary to put in place a construction QC mechanism that not only prevents the need for rework but also prepares for accepting, acting on, and coping with required rework. Hence, the cost of rework (COR) is an inseparable component of overall construction costs, and its reduction directly improves construction cost and quality performance.

Although construction rework has been addressed in the literature, it remains a widespread [1] and prevalent problem [2,3], and poses a real challenge [4–6]. Despite all the advances in philosophies such as lean and total quality management (TQM) in preventing construction errors, COR still accounts for a considerable portion of the total project cost [2,7–9] and affects the construction schedule and quality [10]. Construction rework directly impacts the contract value by 5% to 20% [2], which can lead to complete project failure. Measuring COR enables the CQM system to control the construction budget and improve cost performance while allowing construction professionals to better understand the magnitude of the rework, its causes, and decisions on rework prevention measures [9]. Identifying the impact of COR and its sources enables reductions in the amount of rework

and improvements in construction cost performance [11]. It is noteworthy that anticipating COR facilitates the utilization of QC techniques, such as Pareto analysis and pie charts. These QC techniques are dynamically used throughout the construction lifecycle to predict the construction rework items with a high-cost impact, which, in turn, allows for the timely adjustment of the associated construction schedule, budget, quality, human resources, and communication plans for the appropriate countermeasures. It is also noteworthy that obtaining COR is a key to understanding the cost of quality (COQ), i.e., the conformance costs, and the nonconformance costs, also referred to as cost of poor quality (COPQ) [12]. The ability of construction firms to measure COQ is essential for their survival in today's competitive environment [13].

While the construction management literature agrees on the important contribution of COR to total construction cost, it is not consistent with respect of the magnitude of COR's impact on overall construction cost, estimates of which vary between as much as 0.5% and 20% of the overall contract value [14]. This range reduces the practicality of COR in implementing effective countermeasures during the stages of a project. Additionally, in practice, countermeasures should focus on the construction work items with the higher impact on the total construction cost. It is not always feasible to implement preventive countermeasures or rework management strategies for all rework items associated with different construction activities. Evaluating the cost impact of construction rework supports early decision making on high-impact nonconformance items. For example, ranking the most impactful building defects offers construction companies insight for selecting the most appropriate strategy to continuously improve their construction activities and in turn to support sustainable decision making for the design and operation of buildings [15]. Furthermore, unless the COR for each construction activity is measured, it cannot be compared with the cost of the associated prevention or control plan. COR influences the construction budget, risk, and quality plans, which, in turn, affects the decision making associated with other project management knowledge areas. In order to improve the CQM, budget, and schedule plan, therefore, it is necessary to estimate the COR for each construction activity. This increases the error preparedness of construction organizations, which enhances decision-making resilience and plan accuracy while facilitating the prompt implementation of appropriate countermeasures, and thus also allows for appropriate contingency plans to be developed while helping the manager prevent later issues in other construction project phases [16].

In this study, COR is predicted for the total construction cost, which is a critical decision-making parameter. Experiments with advanced machine learning (ML) models, such as the ensemble method for predicting construction cost and COR, are lacking in the literature. Therefore, our work uses ensemble learning applied to the widely used construction nonconformity reports (NCRs) to ensure the robustness of the created COR predictor. As outlined in Figure 1, the main objective of this study is to assist construction quality managers and cost managers in including COR in their evaluations of different construction activities.

The remainder of this study is organized as follows. The next section discusses construction rework and its impact on overall construction cost. In addition, ensemble learning as a subdiscipline of ML is introduced, and, since the literature is limited to ML-based COR estimators, the ML applications for CQM and cost estimation are reviewed. The following section describes the NCRs obtained from different construction projects in Turkey. Then, the adopted methodology and ensemble COR predictor details are presented, and the benchmark ML predictor configuration is outlined. The next section gives the results that were obtained and discusses the practical implementations of the COR predictor that was developed and its contribution to the existing research on construction management. Finally, a brief conclusion reviews the study findings.
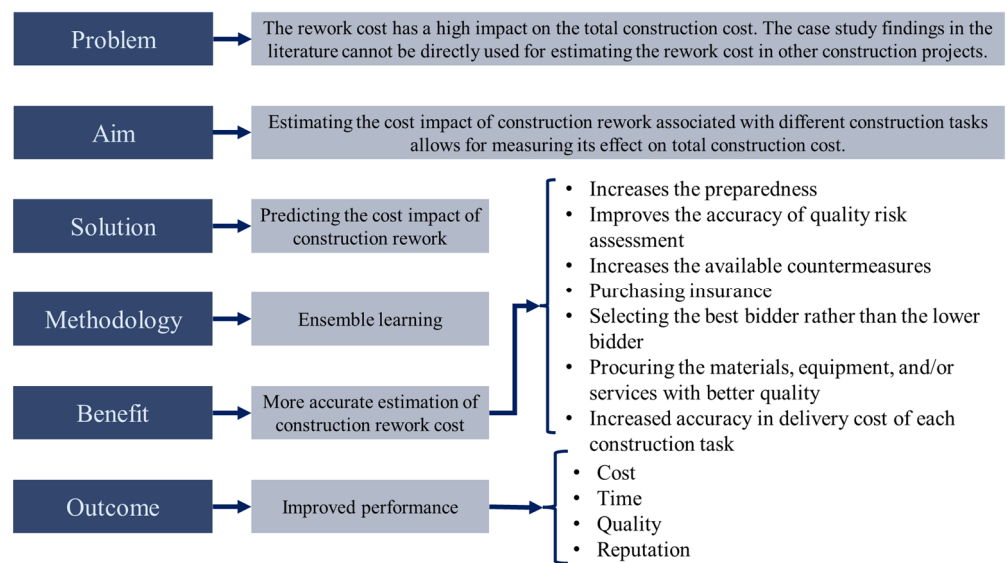
Figure 1. Study outline.

2. Research Background

2.1. Construction Rework

The conventional construction rework procedure based on nonconformities raised within the NCRs is outlined in Figure 2.
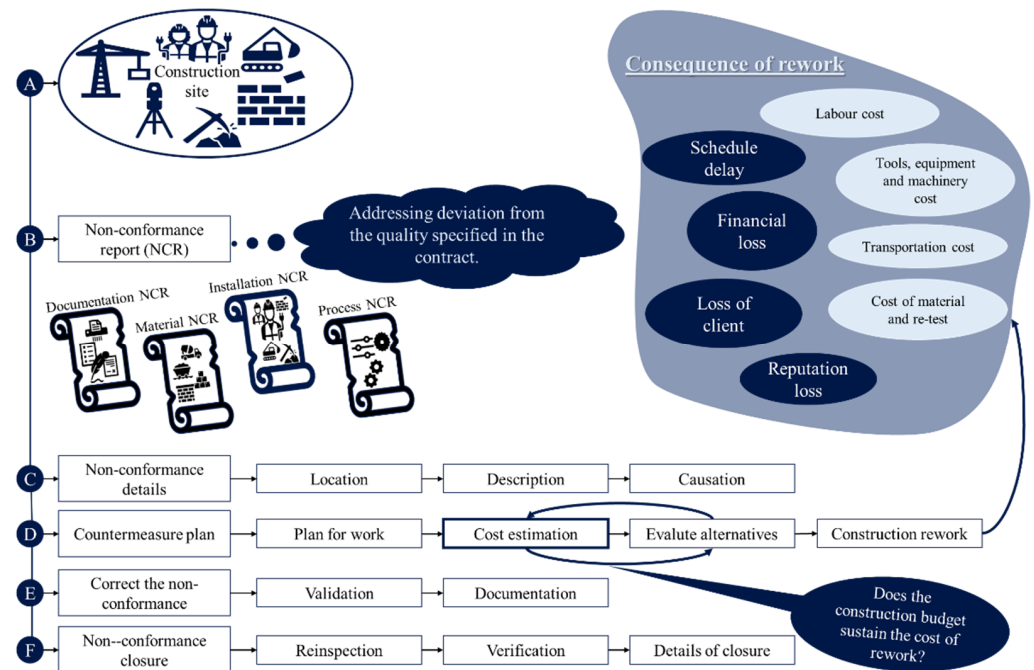
Figure 2. Construction rework procedure outline.

Site accidents, errors, failures, and violations are all causes of delay in the construction schedule and increase costs. Subsequent issues are often raised during the quality inspection of the construction activities performed, whereby the results are recorded in NCRs. The raised nonconformity should be addressed by the contractor, mainly in the form of rework. The topic of construction rework, including its causes, consequences, and prevention measures, has been widely addressed [17–20] using different terms and interpretations [9], such as quality deviation [21], construction nonconformance [22], defects [23], quality failure [24], and rework [25]. These have all emphasized the importance of factoring

in construction rework during the early stages of construction planning in order to mitigate its consequences, which are mainly cost overruns.

2.2. The Cost Impact of Construction Rework

Despite the attention given to construction cost estimation in the previous research, the prediction of construction cost overrun has received relatively little consideration [16]. Similarly, the estimation of cost overruns resulting from the cost of construction rework has not been adequately addressed. Accordingly, the literature on estimating the cost impact of COR is reviewed here, along with construction cost estimation methods within a broader framework. The associated literature [13,26–28] has covered the broader topics of COQ and COPQ. In the literature on CQM, Love and different co-authors have explored construction quality from different perspectives, including construction error [29–33] and rework management [32,34], its impact on construction safety [35–37], and cost [9,38]. Hall and Tomkins [39] included the prevention and appraisal costs required to achieve a 'complete' COQ for buildings in the UK, while Love and Li [40] extended their earlier work on rework causation to quantify the magnitude of COR for Australian construction projects [41].

The research agrees on the negative impact of COR on overall construction cost while attaining varying impact percentages for this according to the demographics and types of evaluated construction projects. Davis et al. [42] found the nonconformance cost to be responsible for over 12% of the total contract value. Love [8,9], through a questionnaire survey on different project types and procurement routes, identified the direct and indirect impact of rework on total construction cost as being 26% and 52%, respectively. Rework costs drag down construction productivity by damaging the associated plans related, for example, to time, cost, and human resources, and this causes financial and reputation loss for the project participants. Hwang et al. [11] evaluated the contribution of COR to the total construction cost of 359 projects, along with its impact on both client and contractor. They found that construction owners are absorbing twice as much impact from COR than contractors. To reduce the magnitude of this problem, contractors often apply an internal quality control and assurance system, and they also often implement proactive measures to anticipate possible rework and associated costs.

In addition to the negative effects of construction rework, there is a possible positive impact on the project cost and quality. Ye et al. [2] investigated 277 construction projects in China to identify the main areas of rework and showed that active rework can improve construction cost, time, and quality. This study further suggests that by implementing a reward strategy and value management tools, required rework can be identified early, enabling timely decision making about the rework, time, cost, and quality benefits for the construction project. A statistical evaluation of 78 data points obtained from construction professionals by Simpeh et al. [7] revealed a mean 5.12% contribution of COR to total contract value and a 76% probability of exceeding its average value. This study also found that rework prediction facilitates quantitative risk assessment and, subsequently, the identification of alternative countermeasures for rework prevention. A more recent study by Love and Smith [4] evaluated the literature and put the impact of COR at between less than 1% and more than 20% of the total contract value. The literature is not consistent in specifying the conditions according to which the impact of COR should be measured, which hinders its practical implementation. The most recent study stated that the COR can vary from 0.5% to 20% of the total contract value [14]. Thus, these studies have provided in-depth investigations on the cost impact of rework, but they are not consistent when it comes to the magnitude of that impact.

The literature on construction management has recorded different contribution percentages for the impact of COR on overall construction cost. Since studies are conducted on projects of different size and type, and within different demographics, the cost impact figures obtained cannot be directly extended to other projects. Although the literature shows the importance of the early identification of COR for improving construction cost

performance, the uncertainty about the magnitude of the impact hinders decision making when selecting the most advantageous countermeasures. Moreover, it is necessary to reach a different COR impact figure for each construction activity in order to prioritize activities with a higher cost impact, since it is not always feasible to implement preventive counter-measures or rework management strategies for all rework items. Furthermore, unless the COR for each work item is measured, it cannot be compared with the rework prevention or control cost.

Thus, to enhance the quality of decision making and quality planning, as well as to increase the chance of construction project success, it is important to estimate the COR for each work item. To translate the literature results on the impact of COR into the context of different construction projects, ML offers a data-oriented solution that can be utilized in different construction project contexts. ML approaches can predict COR by learning the complex patterns within the quality dataset.

2.3. ML for Construction Cost Prediction

ML uses historical evidence to offer a reliable solution that facilitates informed decision making. The literature on ML applications utilizing different types of datasets is growing in various fields [43–56]. Different ML approaches, such as artificial neural network (ANN), deep neural network (DNN), and support vector machine (SVM) are employed due to their ability to understand the complicated, non-linear patterns of real-world datasets. In this regard, the two ML approaches used for the cost estimation of construction projects were ANN [57,58] and SVM [55,59]. Even though other ML approaches, such as k-nearest neighbors (KNN) and decision trees (DT) share similarities with the ANN and SVM algorithms, they have yet to be investigated in the construction management literature [48]. Overall, construction cost estimation studies of more advanced ML approaches are scarce.

The literature on construction quality has mostly focused on quality assurance and quality control, using visual defect detection methodologies for a variety of tasks, including crack identification [60,61], damage localization on wooden building elements [62], and evaluation of pavement conditions [63]. ML approaches have also been used for the identi-fication of rework or defect construction items. To this end, Fan [64] recently constructed a hybrid ML model using association rule mining (ARM) and a Bayesian network (BN) approach identify quality determinants and gain more effective evaluations of defect risk and its occurrence. In a related study, Kim et al. [65] utilized SVM, random forest (RF), and logistic regression (LR) along with three natural language processing (NLP) methods on 310,000 defect cases from South Korea to assign defect items to the appropriate repair task. Shoar et al. [16] used RF to estimate the COR of engineering services in construction to be used for devising appropriate contingency plans. Their study found using RF as a cost estimator to be an efficient approach for screening and prioritizing from the standpoint of cost overrun within construction projects, and that it can be used to devise related contingency plans.

Regarding the present study, the most relevant study is that conducted by Doğan [66] to predict the cost impact of construction nonconformities using case-based reasoning (CBR). His results indicated that the ability of CBR to predict the cost impact of quality problems is higher in construction NCRs. Reviewing the construction management literature, one may say that the development of ML-based cost estimators is still at an early stage. There is a lack of advanced ML approaches, such as ensemble learning methods. Although studies have established the usefulness of these ML methods, they have not elaborated on the robustness of the developed estimators, that is, on the ability to use the systems developed for other datasets. Thus, there is a research gap in the implementation of advanced ML-based techniques for predicting the COR associated with different construction activities.

2.4. Ensemble Learning

Single ML classifiers, such as SVM, KNN, NB, and DT, are trained with labeled datasets through various approaches to predict an output label class. Ensemble classifiers, however,

combine the best predictions of these single ML approaches to improve the final prediction accuracy with improved stability and robustness [67]. The ensemble methods vary according to how they combine the results of single ML classifiers, while their performance depends on the number of individual members along with their prediction accuracy [67]. There are three popular ensemble techniques: (i) stacked, (ii) voting classifiers, and (iii) tree-based. Kansara et al. [68] applied the stacked ensemble (XGBoost regression) and tree-based ensemble (RF) approaches to improve the price prediction accuracy for real estate datasets. However, stacked ensemble approaches have the disadvantages of additional complexity and high computational time. Thus, they are feasible only when other ensemble approaches are not applicable.

Overall, due to their improved accuracy [51,68,69], studies have adopted ensemble learning methods for different prediction activities within different fields. Therefore, ensemble predictors are expected to provide more accurate cost predictions. In addition, the mechanism of the ensemble classifier benefits from both strong and weak predictors, where the latter is used to improve the prediction of the underrepresented classes. The literature on construction quality has still not matured with respect to cost estimation using both single ML predictors and ensemble learning predictors. Furthermore, the COR for different construction activities is not addressed in the construction quality literature. Therefore, because of the superior performance of ensemble learning over single ML models [51,68,69], this study adopts two such techniques, referred to as *soft* and *hard voting classifiers*, and compares them with three conventional tree-based ensemble classifiers (RF, gradient boosting (GB), and AdaBoosting (AB)) along with four single ML classifiers (DT, naïve Bayes (NB9), Logistic Regression (LR), and SVM). Accordingly, Figure 3 presents a simplified form of the procedure for the hard and soft voting classifiers adopted in this study.
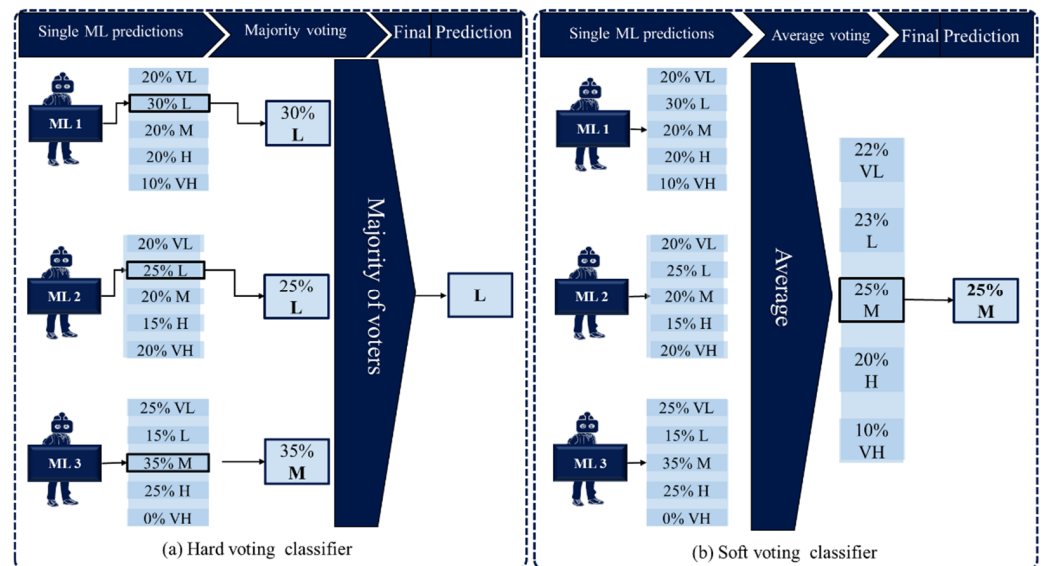

Figure 3. Simplified hard and soft voting COR classifiers.

As shown in Figure 3, each single classifier (ML 1–3) is referred to as a member that predicts an output class label, referred to as a vote. The hard voting classifier selects the label voted for by the majority of the members. The hard voting ensemble classifier uses the average of the predicted probabilities of all the members. For example, in Figure 3, ML 1 and ML 2 both classify the impact of COR as two, while ML 3 classifies it as three, so the hard voting model predicts the COR impact as two. Soft voting is less straightforward, since it uses the probability of each of the five classes and finds the average probability of all the classifiers within each class to select the final label. Voting classifiers can benefit from the voting of both single and ensemble classifiers. Tree-based ensemble approaches, such as

GB and AB, have been utilized to predict the rental price of apartments, showing better prediction accuracy than single ML approaches [69]. In addition to voting classifiers, bagging and boosting tree-based ensemble approaches are experimented with in this study. Figure 4 outlines the bagging and boosting mechanism within the tree-based ensemble approaches.
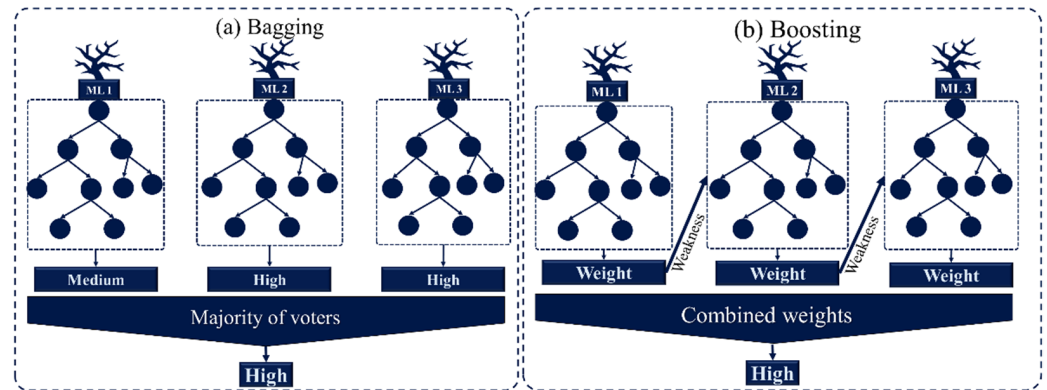

Figure 4. Simplified RF and boosting (AB and GB) ensemble mechanisms.

The bagging (i.e., RF) and boosting (i.e., AB and GB) mechanisms are the main ensemble approaches used within tree-based ensemble models, taking advantage of the best predictions of single DTs.

3. Data Description from Construction Nonconformance Report

This study uses the nonconformance items from diverse construction projects undertaken by international construction companies, collected in a study by Doğan [66] in 2021. The dataset comprises 2527 nonconformance items recorded by inspecting the different activities throughout the construction phase. A histogram associated with the construction activities and the frequency of recorded nonconformity is given in Figure 5, with activities having less than 20 occurrences aggregated under the 'other project activities' group. The collected nonconformance items were assigned to the different causation attributes through interviews.
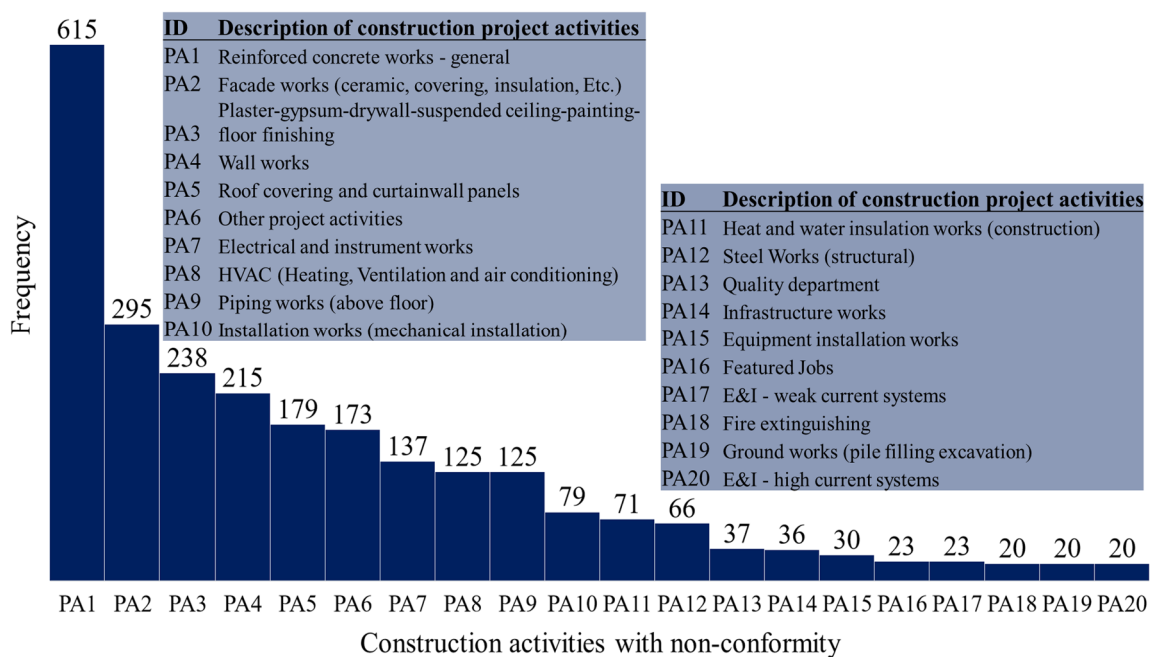

Figure 5. Details of construction activities registered in the NCRs.

Since the dataset was collected during the construction phases, attributes related to the pre-construction, design, and tendering phases, such as those related to clients and subcontractors, are omitted. The obtained NCR dataset is described using the stacked histogram, which details different construction project types. The NCRs include details of the causation of each recorded item, divided into material, design, operation, and construction causation. In addition, the cost impact of COR (y) is assigned as an output feature column. This assigns each input feature a cost impact of between one and five, corresponding to very low (VL), low (L), medium (M), high (H), and very high (VH).

In addition, the output cost impact is recategorized into three, four and five cost impact classes to evaluate the class prediction accuracy of the adopted ensemble approaches (Figure 6).
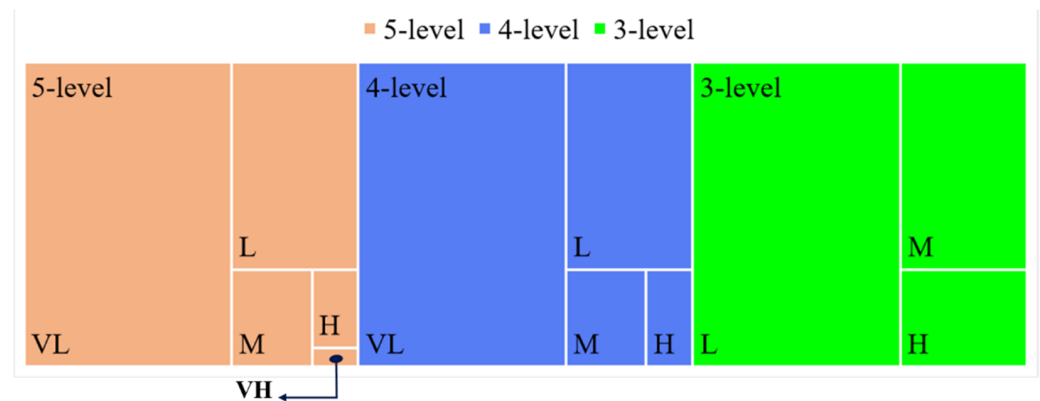

Figure 6. Cost impact classes.

As Figure 6 shows, critical nonconformance items with high-cost impact classes are underrepresented, there being few records of these compared with lower impact cost classes. Figure 7 shows the frequency of material-related nonconformance attributes used in this study. The observations from Figure 6 highlight the class imbalance among the different cost impact groups. The ability of ML to represent the under-represented cost impact classes is reduced.
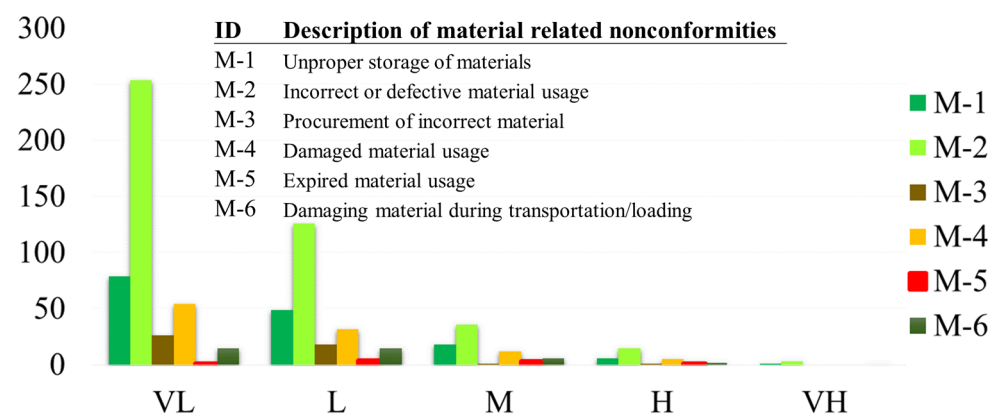
Figure 7. Material-related nonconformance attributes.

Furthermore, as shown in Figure 7, the collected NCRs also include the stage at which the nonconformance issue was initiated. Thus, each nonconformance attribute is linked to installation, documentation, material, or process damage. For example, 10 nonconformance issues with a cost impact of three are recorded as caused by damaged material usage (M-4), where the damage is initiated at the installation stage. Likewise, Figure 8 shows the nonconformity attributes related to design and operation. The design-related attributes

recorded during the construction phase are limited because the collected NCRs were only gathered from the construction site.

ID	Description of operational nonconformities
O-1	Problems with Purchasing Department
O-2	Problems with Warehouse (Labelling etc.)
O-3	Sending Wrong Material from Warehouse
O-4	Lack of Supervision
O-5	Problems with Documentation

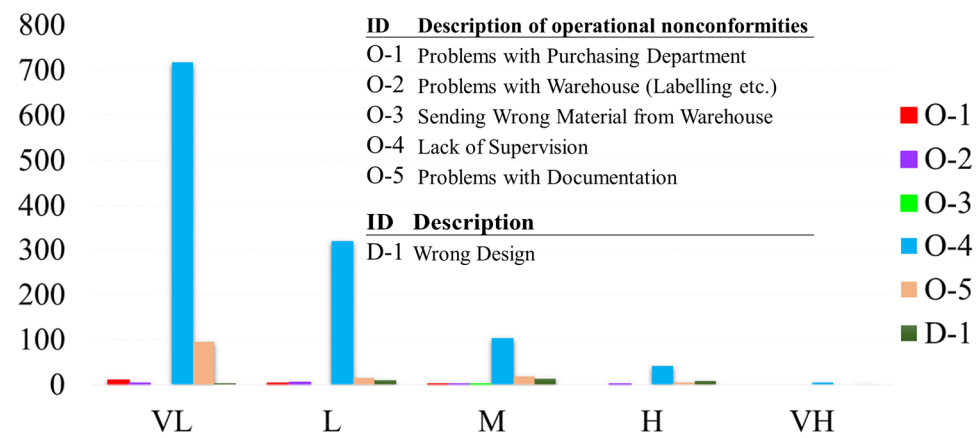| ID	Description
D-1	Wrong Design

Figure 8. Design- and operation-related nonconformance attributes.

As Figure 8 shows, the design-related nonconformance attributes mostly occurred during the processing stage, while the operation-related issues were mostly associated with a lack of supervision (O-4) during the installation phase. As this study uses NCRs from construction sites, the frequency of nonconformances within the installation stage significantly increases in construction-related attributes (Figure 9).
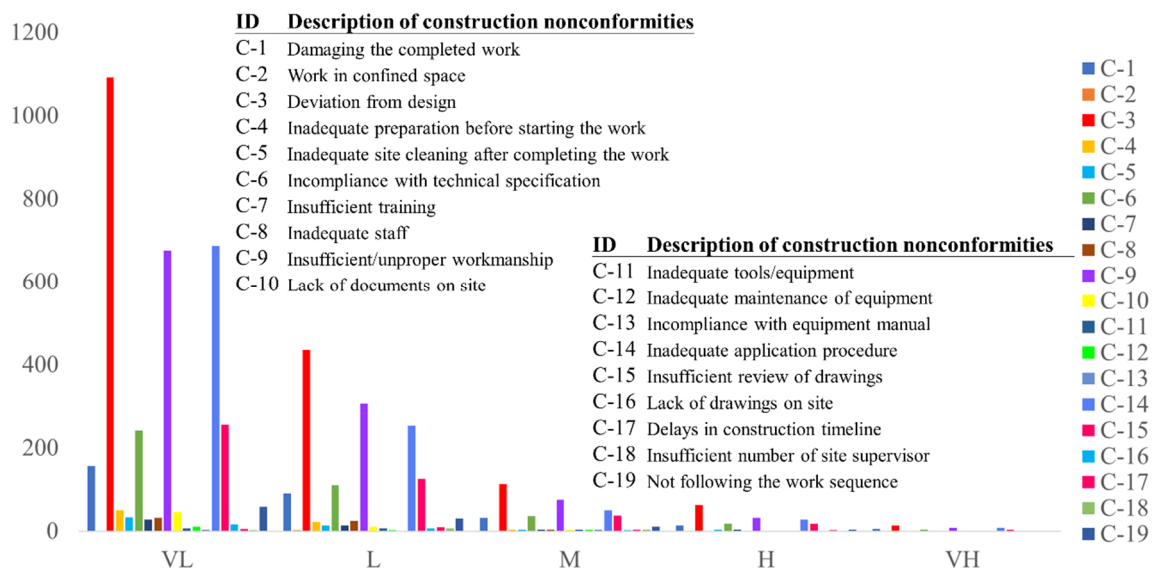

ID	Description of construction nonconformities
C-1	Damaging the completed work
C-2	Work in confined space
C-3	Deviation from design
C-4	Inadequate preparation before starting the work
C-5	Inadequate site cleaning after completing the work
C-6	Incompliance with technical specification
C-7	Insufficient training
C-8	Inadequate staff
C-9	Insufficient/unproper workmanship
C-10	Lack of documents on site

ID	Description of construction nonconformities
C-11	Inadequate tools/equipment
C-12	Inadequate maintenance of equipment
C-13	Incompliance with equipment manual
C-14	Inadequate application procedure
C-15	Insufficient review of drawings
C-16	Lack of drawings on site
C-17	Delays in construction timeline
C-18	Insufficient number of site supervisor
C-19	Not following the work sequence

Figure 9. Construction-related nonconformance attributes.

4. Methodology

This study adopts ensemble ML classifiers to determine the impact of COR on overall project cost. The methodology is outlined in Figure 10.
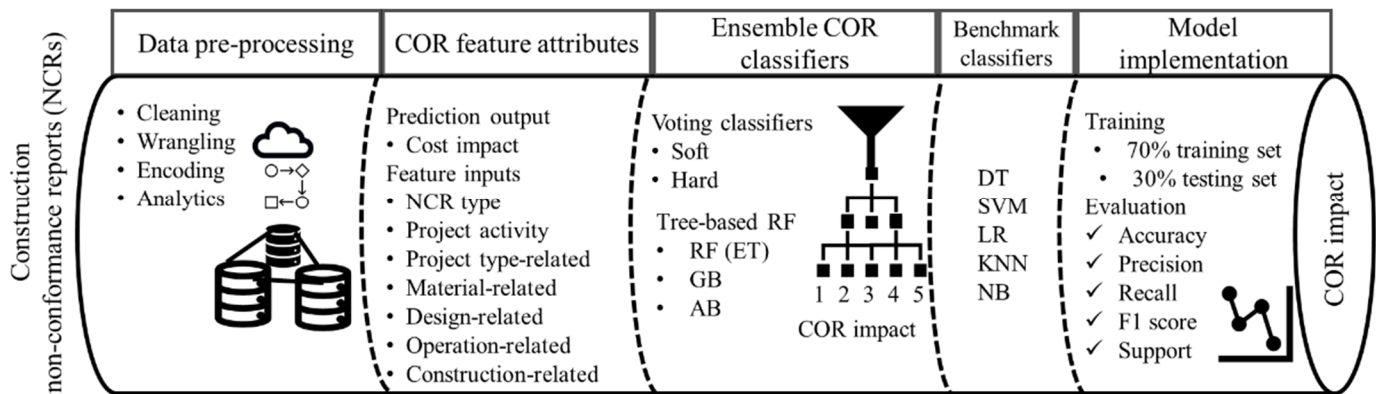
Figure 10. COR classifier methodology.

The methodology described was implemented within the Python programming environment. Most of the data preprocessing, analysis, and associated ML configuration was performed with widely used libraries, such as the Pandas [70], NumPy [70], and Scikit-Learn [71,72] packages. The nonconformance dataset was utilized with an ensemble classifier to predict the impact of COR on overall construction cost, while the results were compared with the single ML predictors.

4.1. Data Preprocessing

The dataset (*D*) was obtained from the construction NCRs with 39 feature columns and 2527 rows (Equation (1)):

$$D = \{x1,\ x2,\ \ldots,\ x31,\ x32,\ \ldots,\ x36, x37, x38,\ y\} \tag{1}$$

The material-, design-, operation-, and construction-related nonconformity items were used as 31 binary input feature columns. The project types were presented by five binary columns ($x32 - x36$), associated with industrial, hospital, high-rise, housing, and other building construction types. The NCR type column that shows the initiation area of the recorded nonconformity item was used as another input feature ($x37$) with four categories: installation, documentation, material inspection, and processes. To translate each category into the ML language, the dummy encoding method was used, which converts the NCR type into four columns, each showing a single category with either zero or one. For example, $\{1, 0, 0, 0\}$ shows the NCR type as installation, while $\{0, 0, 1, 0\}$ stands for material NCR type. In addition, the construction activity associated with each nonconformance item was used as a categorical input column ($x38$) with 20 categories, as depicted in Figure 5. Again, dummy encoding was used to translate the construction activity into binary format, this time within 20 feature columns. This resulted in an encoded dataset with 61 columns and 2527 nonconformance rows. Finally, 70% of the dataset was used for training and the rest (30%) was kept for performance evaluation.

4.2. Configuring Voting Classifiers

There is a wide range of single and ensemble ML algorithms that voting classifiers can use for a given prediction. This study explores different single and ensemble ML methods used in the literature to reach the best combination for the given study (Figure 11).
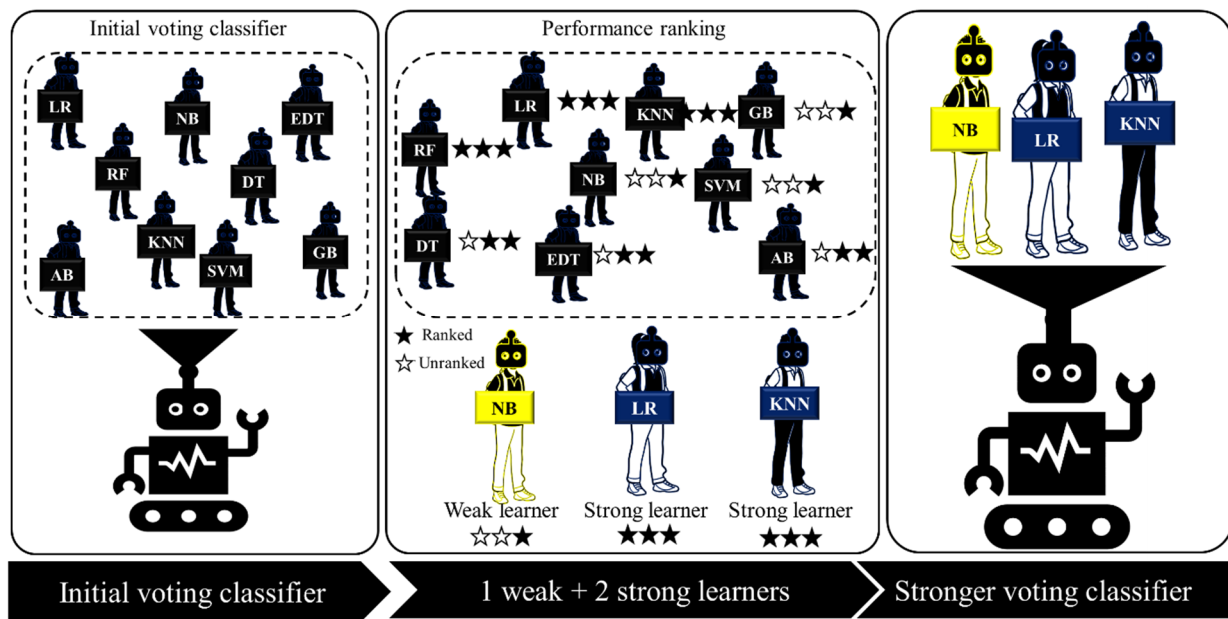
Figure 11. Configuration of soft and hard voting classifiers.

This study aimed at developing a COR impact predictor that achieves good prediction accuracy with simple implementation. Therefore, dimension reduction, class imbalances, and optimization techniques were not used to ascertain the best hyperparameters. Instead, the layers of different ML approaches were placed within a voting classifier and, based on the observed performance, a final voting classifier was configured with fewer ML members to accelerate the training procedure. The compulsory parameters, including the number of estimators (for RF, AB, and GB) and the number of neighbors (for KNN), were set roughly close to the benchmark models, while the other options were left as default. Although feature engineering and optimization techniques would have improved the performance of the ensemble predictor, their implementation was beyond the scope of the present study.

The radial basis function (RBF) was used as the SVM kernel while enabling a balanced class weight option. KNN was trained with 23 neighbors. The LR was adjusted with a LIBLINEAR optimizer and an l1 regularization. To ensure the creation of a weak learner, NB was used in its default form. The number of estimators for AB and RF was adjusted to 300 trees. The GB used 100 estimators while its learning rate was customized at 0.1. Afterwards, in order to reduce computational cost, the number of voting classifier members was reduced to three. In this respect, two strong learners were combined with a weak learner to simultaneously ensure accuracy and the elimination of bias. Combining strong and weak ML learners enhances the prediction accuracy of models for different rework cost impact classes while boosting the model's overall performance in terms of generalization ability and computational cost. Therefore, LR and KNN were used as strong learners, while NB was used as a weak learner for both the soft and hard voting classifiers.

4.3. Configuring Benchmark Classifiers

Unlike the voting classifiers, which were configured without any particular attention to the fine-tuning of their hyperparameters, each of the benchmark ML approaches was specifically fine-tuned to ensure a fair comparison between the ensemble voting classifiers and single ML predictions (Table 1).

Table 1. Hyperparameters of benchmark COR classifiers.

ML	Hyperparameters	Iterated Values	Selected Value
KNN	Number of neighbors	(2, 25)	23
LR	Optimizer	LIBLINEAR	LIBLINEAR
	Regularization	L1 L2	L1
SVM	Kernel	RBF	RBF
	C	1, 5, 10, 15	15
	Gamma	0.0001, 0.0005, 0.001, and 0.005	0.005
	Class weight	Imbalanced Balanced	Balanced
DT	Number of features	1589	15
	Tree depth	(1–37)	3
RF	Number of trees	50, 100, 150, 200, 300, 400	300
RF (ET)	Number of trees	115, 125, 135, 145, 155, 165, 175, 180, 185, 190, 195, 205	145
GB	Number of estimators (trees)	10, 20, 30, 40, 50, 60, 70, 80, 90, 100	10
	Learning rate	0.1, 0.01, 0.001	0.1
AB	Number of estimators (trees)	100, 150, 200	100
	Learning rate	0.01, 0.001	0.1

4.4. Evaluation Metrics

The prediction performance of the OCR impact predictors was evaluated using conventional accuracy, precision, accuracy, and F1 scores. For this, the number of correct predictions of each cost class (true positive (TP)) and correct assignments of the sample to the rest of the subclasses (true negative (TN)) was obtained. Likewise, the incorrect predictions within each subclass (false positive (FP), and false negative (FN)) were also recorded. Accuracy, F1 scores, precision, and recall [73] were obtained using Equations (2)–(5).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$F1\ score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \tag{3}$$

$$Percsion_{Multiclass} = \frac{\sum_{i=2}^{classes} TP_i}{\sum_{i=2}^{classes} TP_i + FP_i} \tag{4}$$

$$Recall_{Multiclass} = \frac{\sum_{i=2}^{classes} TP_i}{\sum_{i=2}^{classes} TP_i + FN_i} \tag{5}$$

Accuracy did not provide a satisfactory evaluation for an imbalanced dataset as it does not consider FP and FN. Thus, the F1 score was preferred for the imbalanced dataset as it combines precision and recall. However, for a reliable COR predictor, the accuracy of the subclasses also needed to be evaluated.

5. Results and Discussion

The prediction performance of the soft and hard voting classifiers along with the benchmark ML approaches are shown in Table 2.

Table 2. COR prediction performance results.

ML Type	OCR Classifier		Accuracy	Precision	Recall	F1 Score
Ensemble ML (voting)	Hard voting	5-level	0.59	0.49	0.51	0.59
		4-level	0.59	0.52	0.5	0.59
		3-level	0.61	0.54	0.53	0.61
	Soft voting	5-level	0.54	0.51	0.48	0.54
		4-level	0.53	0.71	0.48	0.53
		3-level	0.5	0.57	0.51	0.5
Single ML	KNN	5-level	0.6	0.48	0.5	0.6
		4-level	0.61	0.55	0.55	0.61
		3-level	0.59	0.54	0.55	0.59
	LR (lr)	5-level	0.61	0.47	0.5	0.61
		4-level	0.62	0.53	0.52	0.62
		3-level	0.62	0.54	0.53	0.62
	LR (l1)	5-level	0.55	0.43	0.48	0.55
		4-level	0.56	0.44	0.48	0.56
		3-level	0.6	0.56	0.55	0.6
	LR (l2)	**5-level**	0.63	0.52	0.56	**0.63**
		4-level	0.61	0.52	0.54	0.61
		3-level	0.62	0.49	0.49	0.62
	SVM	5-level	0.38	0.55	0.44	0.38
		4-level	0.42	0.55	0.47	0.42
		3-level	0.43	0.55	0.47	0.43
	DT	5-level	0.38	0.55	0.44	0.38
		4-level	0.62	0.49	0.48	0.62
		3-level	0.63	0.71	0.49	**0.63**
	NB	5-level	0.09	0.49	0.12	0.09
		4-level	0.09	0.6	0.09	0.09
		3-level	0.35	0.58	0.33	0.35
Ensemble ML (tree-based)	RF	5-level	0.62	0.61	0.61	0.62
		4-level	0.57	0.51	0.52	0.57
		3-level	0.57	0.48	0.51	0.57
	RF (ET)	5-level	0.56	0.47	0.5	0.56
		4-level	0.57	0.48	0.51	0.57
		3-level	0.57	0.51	0.53	0.57
	GB	5-level	0.63	0.62	0.55	0.63
		4-level	0.63	0.46	0.49	**0.63**
		3-level	0.62	0.39	0.48	0.62
	AB	5-level	0.62	0.38	0.47	0.62
		4-level	0.62	0.45	0.48	0.62
		3-level	0.62	0.45	0.49	0.62

As Table 2 shows, LR outperformed all predictors in terms of F1 score for the five-level cost impact prediction. Additionally, DT and GB displayed better F1 scores for the three- and four-level COR impact predictions. However, the accuracy and F1 scores can be misleading when working with an imbalanced dataset, so to investigate the practicality of the predictors, their ability in predicting each cost impact class needs to be investigated. This is best achieved by measuring the F1 score of each of the cost impact classes. Figure 12 presents the prediction performances of the best-performing classifiers within the benchmark model with the hard and soft voting classifiers.
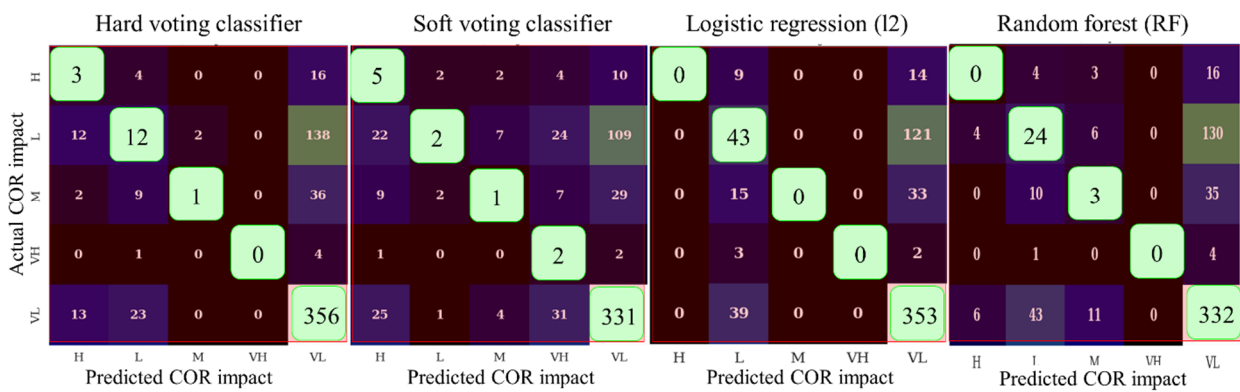
Figure 12. Prediction performance of 5-level classifiers.

As Figure 12 shows, both the soft and the hard voting classifiers were able to detect the high-cost impact items with only a 4% occurrence (support = 23). Among the COR classifiers, however, only one soft voting classifier detected the rework with a cost impact of five (VH) with a low occurrence of 1% (support = five).

A high accuracy with DT was only associated with nonconformance items with a very low cost impact, and it performed poorly for other underrepresented but more important cost impact classes (Figure 13). Likewise, voting classifiers exhibited better performance for medium- and high-impact cost estimation compared with GB, which completely failed to predict COR with a medium impact (Figure 14). Despite its poor class performance in the four-level classification, RF resulted in the best prediction for high-impact COR items without sacrificing overall accuracy. On the other hand, the soft voting classifier proved to be consistent in its precision accuracy for different classification levels.
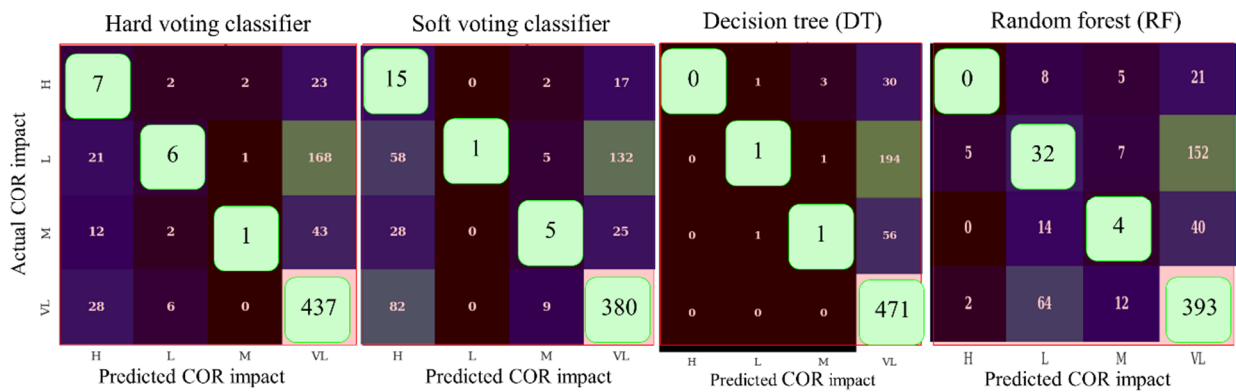

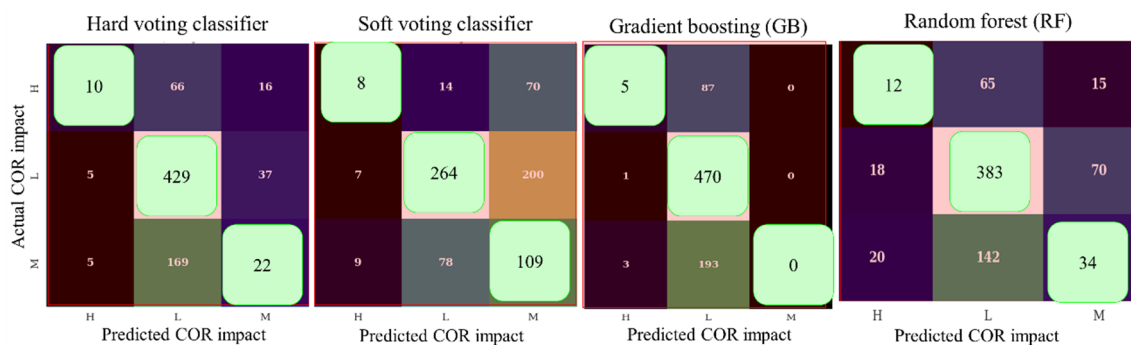**Figure 13.** Prediction performance in 4-level classifiers.

Figure 14. Prediction performance in 3-level classifiers.

To achieve a robust COR predictor, it is important to evaluate the ability of ML to solve the problem. Most single and tree-based ensemble ML predictors failed to estimate COR with high (four) or very high (five) impacts. The superior accuracy of benchmark ML approaches is due to their predicting low-cost impact rework items. However, they are incapable of predicting high-impact cost items. The prediction of low-cost impact rework items cannot reduce the deviation between the as-planned and as-built costs. On the other hand, voting classifiers are more successful in predicting high-impact COR items. The soft voting classifiers were the most robust for COR, displaying a significantly good performance in detecting underrepresented cost impact classes.

To better illustrate the practical implementation of the model, a trained soft voting classifier was used to predict an unseen user input from the test dataset. For example, the user may want to evaluate the material-related nonconformities of a high-rise building project using the available construction team status report and site conditions. The user defines a scenario in which negligence in the initial material inspection due to a lack of site supervision (O-4) results in the receipt of defective material (M-2) from the supplier. The defective material, accompanied by an insufficient review of the design documents (C-15), causes a deviation from the design (C-3). Thus, with respect to the scenario specified by the user, the system can predict the cost impacts of different construction activities under a user-defined scenario. Once the user specifies different activities, such as the facade works (ceramic, coating, insulation, etc.), the system uses the trained soft voting classifier to evaluate its impact on the overall construction cost. In this example, the soft voting classifier can accurately predict a cost impact of three (medium) on the overall construction cost.

6. Contribution to the Body of Knowledge

Despite its importance, construction organizations are rarely aware of the rework impact on their budgets and on safe and environmental performances [34]. Small-sized construction companies still do not appreciate the magnitude of profit loss due to poor quality as they do not usually allocate CQM within a budget [13]. CQM and QC are not limited to error and violation prevention measures, but can also contribute to coping strategies, such as planning alternative countermeasures. The ability to predict COR allows for timely decision making for the required countermeasures, which can also improve the construction time, cost, and quality performance [2]. If a construction budget cannot sustain COR, it is hard to correct a construction error.

In this context, an ensemble COR predictor allows for dynamic and fast cost impact estimations throughout a project and offers more reliable cost impact estimations than the existing, single ML approaches. This is especially useful for cost variation and content analysis using Pareto and pie chart techniques. Throughout the construction lifecycle, the accumulation of project experiences adds to the knowledge of the ensemble models, which can further enhance their cost impact prediction accuracy and thus facilitate enhanced strategic planning by prioritizing the quality control items with the greatest cost impact. Therefore, the proposed COR impact estimator can enhance decision making and the associated planning for construction professionals. Specifically, relationship-style construction contracting models, such as alliance contracts, incorporate an element of error though procuring the construction project under a 'no blame, no fault' culture.

Generally, estimating COR improves cost, schedule, and resource-allocation planning, enhances the creation of the associated contingency plans [16], and also increases the visibility of the expected failure scenarios when purchasing the rework insurance, which is usually added to the general liability policy. This study has focused on estimating the impact of COR on overall construction cost, so the ability of ensemble construction cost predictors to improve the cost impact estimation in these areas remains to be addressed in future research. From a technical perspective, the ensemble method adopted needs to be further improved using different engineering features and optimization techniques to enhance the model prediction accuracy, especially for the underrepresented cost classes.

7. Conclusions

The early estimation of COR offers several benefits for construction professionals in terms of increasing the preparedness of the construction budget for dealing with risk (i.e., COR). On the one hand, the construction quality management literature is still limited with regard to the ML-based COR predictors, while on the other hand, the developed single ML predictors within the other cost estimation fields are not responsive for underrepresented classes with limited data records. However, a COR predictor does not solve this problem unless it can predict all the cost impact classes. Therefore, this study has proposed a robust ensemble ML predictor for estimating the cost of construction rework.

The adopted ensemble voting classifiers proved to be more effective in predicting the underrepresented high-cost impact construction rework activities than the benchmark models. Both single ML and tree-based predictors failed to estimate COR with (very) high-cost impacts on overall construction budgets. Additionally, the soft voting classifier proved to be consistent in the accuracy of its prediction outcomes and was able to classify all the different COR impacts for three-, four-, and five-level classification tasks. The developed COR impact predictor increases the reliability and accuracy of the cost impact estimation, which, in turn, enables dynamic cost variation analysis and thus improves cost-based decision making.

COR has many undesirable effects, from cost fluctuations to the waste of material and labor and equipment hours. Ultimately, it is one of the crucial aspects of sustainability in construction. Thus, the early identification of high-cost impact rework items allows for a focus on countermeasures to prevent critical rework items. This in turn reduces the waste in construction flow, time, and material consumption while enhancing the different aspects of project performance, such as budget and quality performance. Finally, the discussed aspects of the project improve its overall sustainability level in terms of quality, economy, and waste criteria. Therefore, we recommend the further exploration of the use of different ML methods to predict and reduce COR.

References

1. Mohamed, H.H.; Ibrahim, A.H.; Soliman, A.A. Toward Reducing Construction Project Delivery Time under Limited Resources. *Sustainability* **2021**, *13*, 11035. [CrossRef]
2. Ye, G.; Jin, Z.; Xia, B.; Skitmore, M. Analyzing Causes for Reworks in Construction Projects in China. *J. Manag. Eng.* **2015**, *31*, 04014097. [CrossRef]
3. Hwang, B.-G.; Zhao, X.; Goh, K.J. Investigating the Client-Related Rework in Building Projects: The Case of Singapore. *Int. J. Proj. Manag.* **2014**, *32*, 698–708. [CrossRef]
4. Love, P.; Smith, J. Unpacking the Ambiguity of Rework in Construction: Making Sense of the Literature. *Civ. Eng. Environ. Syst.* **2018**, *35*, 180–203. [CrossRef]
5. Asadi, R.; Wilkinson, S.; Rotimi, J.O.B. Towards Contracting Strategy Usage for Rework in Construction Projects: A Comprehensive Review. *Constr. Manag. Econ.* **2021**, *39*, 953–971. [CrossRef]
6. Al-Janabi, A.M.; Abdel-Monem, M.S.; El-Dash, K.M. Factors Causing Rework and Their Impact on Projects' Performance in Egypt. *J. Civ. Eng. Manag.* **2020**, *26*, 666–689. [CrossRef]
7. Simpeh, E.K.; Ndihokubwayo, R.; Love, P.E.D.; Thwala, W.D. A Rework Probability Model: A Quantitative Assessment of Rework Occurrence in Construction Projects. *Int. J. Constr. Manag.* **2015**, *15*, 109–116. [CrossRef]

8. Love, P.E.D.; Sing, C.-P. Determining the Probability Distribution of Rework Costs in Construction and Engineering Projects. *Struct. Infrastruct. Eng.* **2013**, *9*, 1136–1148. [CrossRef]

9. Love, P.E.D. Influence of Project Type and Procurement Method on Rework Costs in Building Construction Projects. *J. Constr. Eng. Manag.* **2002**, *128*, 18–29. [CrossRef]

10. Khalesi, H.; Balali, A.; Valipour, A.; Antucheviciene, J.; Migilinskas, D.; Zigmund, V. Application of Hybrid SWARA–BIM in Reducing Reworks of Building Construction Projects from the Perspective of Time. *Sustainability* **2020**, *12*, 8927. [CrossRef]

11. Hwang, B.-G.; Thomas, S.R.; Haas, C.T.; Caldas, C.H. Measuring the Impact of Rework on Construction Cost Performance. *J. Constr. Eng. Manag.* **2009**, *135*, 187–198. [CrossRef]

12. Schiffauerova, A.; Thomson, V. A Review of Research on Cost of Quality Models and Best Practices. *Int. J. Qual. Reliab. Manag.* **2006**, *23*, 647–669. [CrossRef]

13. Abdelsalam, H.M.E.; Gad, M.M. Cost of Quality in Dubai: An Analytical Case Study of Residential Construction Projects. *Int. J. Proj. Manag.* **2009**, *27*, 501–511. [CrossRef]

14. Love, P.E.D.; Matthews, J.; Sing, M.C.P.; Porter, S.R.; Fang, W. State of Science: Why Does Rework Occur in Construction? What Are Its Consequences? And What Can Be Done to Mitigate Its Occurrence? *Engineering*, 2022; *in press*. [CrossRef]

15. Milion, R.N.; Alves, T.; da C. L. Alves, T.; Paliari, J.C.; Liboni, L.H.B. CBA-Based Evaluation Method of the Impact of Defects in Residential Buildings: Assessing Risks towards Making Sustainable Decisions on Continuous Improvement Activities. *Sustainability* **2021**, *13*, 6597. [CrossRef]

16. Shoar, S.; Chileshe, N.; Edwards, J.D. Machine Learning-Aided Engineering Services' Cost Overruns Prediction in High-Rise Residential Building Projects: Application of Random Forest Regression. *J. Build. Eng.* **2022**, *50*, 104102. [CrossRef]

17. Knyziak, P. The Impact of Construction Quality on the Safety of Prefabricated Multi-Family Dwellings. *Eng. Fail. Anal.* **2019**, *100*, 37–48. [CrossRef]

18. Love, P.E.D.; Ika, L.; Luo, H.; Zhou, Y.; Zhong, B.; Fang, W. Rework, Failures, and Unsafe Behavior: Moving Toward an Error Management Mindset in Construction. *IEEE Trans. Eng. Manag.* **2022**, *69*, 1489–1501. [CrossRef]

19. Fayek, A.R.; Dissanayake, M.; Campero, O. Developing a Standard Methodology for Measuring and Classifying Construction Field Rework. *Can. J. Civ. Eng.* **2004**, *31*, 1077–1089. [CrossRef]

20. Sugawara, E.; Nikaido, H. Properties of AdeABC and AdeIJK Efflux Systems of Acinetobacter Baumannii Compared with Those of the AcrAB-TolC System of Escherichia Coli. *Antimicrob. Agents Chemother.* **2014**, *58*, 7250–7257. [CrossRef]

21. Burati, J.L.; Farrington, J.J.; Ledbetter, W.B. Causes of Quality Deviations in Design and Construction. *J. Constr. Eng. Manag.* **1992**, *118*, 34–49. [CrossRef]

22. Abdul-Rahman, H. The Cost of Non-Conformance during a Highway Project: A Case Study. *Constr. Manag. Econ.* **1995**, *13*, 23–32. [CrossRef]

23. Josephson, P.E.; Hammarlund, Y. Causes and Costs of Defects in Construction a Study of Seven Building Projects. *Autom. Constr.* **1999**, *8*, 681–687. [CrossRef]

24. Barber, P.; Graves, A.; Hall, M.; Sheath, D.; Tomkins, C. Quality Failure Costs in Civil Engineering Projects. *Int. J. Qual. Reliab. Manag.* **2000**, *17*, 479–492. [CrossRef]

25. Ashford, J.L. *The Management of Quality in Construction*; Routledge: London, UK, 2002; ISBN 9781135833862.

26. Campanella, J. Principles of Quality Costs: Principles, Implementation, and Use. In Proceedings of the Annual Quality Congress Proceedings, Milwaukee, WI, USA, 4–8 May 1998; p. 507.

27. Bajpai, A.K.; Willey, P.C.T. Questions about Quality Costs. *Int. J. Qual. Reliab. Manag.* **1989**, *6*, 6. [CrossRef]

28. Ziegel, E.R. Juran's Quality Control Handbook. *Technometrics* **1990**, *32*, 97–98. [CrossRef]

29. Love, P.E.D.; Smith, J. Error Management: Implications for Construction. *Constr. Innov.* **2016**, *16*, 418–424. [CrossRef]

30. Love, P.E.D.; Lopez, R.; Edwards, D.J. Reviewing the Past to Learn in the Future: Making Sense of Design Errors and Failures in Construction. *Struct. Infrastruct. Eng.* **2013**, *9*, 675–688. [CrossRef]

31. Love, P.E.D. Creating a Mindfulness to Learn from Errors: Enablers of Rework Containment and Reduction in Construction. *Dev. Built Environ.* **2020**, *1*, 100001. [CrossRef]

32. Love, P.E.D.; Matthews, J.; Fang, W. Rework in Construction: A Focus on Error and Violation. *J. Constr. Eng. Manag.* **2020**, *146*, 1901. [CrossRef]

33. Love, P.E.D.; Smith, J.; Teo, P. Putting into Practice Error Management Theory: Unlearning and Learning to Manage Action Errors in Construction. *Appl. Ergon.* **2018**, *69*, 104–111. [CrossRef]

34. Matthews, J.; Love, P.E.D.; Porter, S.R.; Fang, W. Smart Data and Business Analytics: A Theoretical Framework for Managing Rework Risks in Mega-Projects. *Int. J. Inf. Manag.* **2022**, *65*, 102495. [CrossRef]

35. Love, P.; Ika, L.; Matthews, J.; Fang, W.; Carey, B. The Duality and Paradoxical Tensions of Quality and Safety: Managing Error in Construction Projects. *IEEE Trans. Eng. Manag.* **2022**, 1–8. [CrossRef]

36. Love, P.E.D.; Smith, J.; Ackermann, F.; Irani, Z.; Fang, W.; Luo, H.; Ding, L. Houston, We Have a Problem! Understanding the Tensions between Quality and Safety in Construction. *Prod. Plan. Control* **2019**, *30*, 1354–1365. [CrossRef]

37. Love, P.E.D.; Teo, P.; Morrison, J. Unearthing the Nature and Interplay of Quality and Safety in Construction Projects: An Empirical Study. *Saf. Sci.* **2018**, *103*, 270–279. [CrossRef]

38. Love, P.E.D.; Smith, J.; Ackermann, F.; Irani, Z.; Teo, P. The Costs of Rework: Insights from Construction and Opportunities for Learning. *Prod. Plan. Control* **2018**, *29*, 1082–1095. [CrossRef]

39. Hall, M.; Tomkins, C. A Cost of Quality Analysis of Building Project: Towards a Complete Methodology for Design and Build. *Constr. Manag. Econ.* **2001**, *19*, 727–740. [CrossRef]

40. Love, P.E.D.; Manual, P.; Li, H. Determining the Causal Structure of Rework Influences in Construction. *Constr. Manag. Econ.* **1999**, *17*, 505–517. [CrossRef]

41. Peter, E.; Love, D.; Heng, L.I. Quantifying the Causes and Costs of Rework in Construction. *Constr. Manag. Econ.* **2000**, *18*, 479–490. [CrossRef]

42. Davis, K.; Ledbetter, W.B.; Burati, J.L. Measuring Design and Construction Quality Costs. *J. Constr. Eng. Manag.* **1989**, *115*, 385–400. [CrossRef]

43. Zhao, Y.; Kok Foong, L. Predicting Electrical Power Output of Combined Cycle Power Plants Using a Novel Artificial Neural Network Optimized by Electrostatic Discharge Algorithm. *Measurement* **2022**, *198*, 111405. [CrossRef]

44. Nejati, F.; Tahoori, N.; Sharifian, M.A.; Ghafari, A.; Nehdi, M.L. Estimating Heating Load in Residential Buildings Using Multi-Verse Optimizer, Self-Organizing Self-Adaptive, and Vortex Search Neural-Evolutionary Techniques. *Buildings* **2022**, *12*, 1328. [CrossRef]

45. Chong, M.; Abraham, A.; Paprzycki, M. Traffic Accident Analysis Using Machine Learning Paradigms. *Informatica* **2005**, *29*, 89–98. [CrossRef]

46. Liang, Y.; Reyes, M.L.; Lee, J.D. Real-Time Detection of Driver Cognitive Distraction Using Support Vector Machines. *IEEE Trans. Intell. Transp. Syst.* **2007**, *8*, 340–350. [CrossRef]

47. Hacıefendioğlu, K.; Mostofi, F.; Toğan, V.; Başağa, H.B. CAM-K: A Novel Framework for Automated Estimating Pixel Area Using K-Means Algorithm Integrated with Deep Learning Based-CAM Visualization Techniques. *Neural Comput. Appl.* **2022**, *34*, 17741–17759. [CrossRef]

48. Bodendorf, F.; Merkl, P.; Franke, J. Intelligent Cost Estimation by Machine Learning in Supply Management: A Structured Literature Review. *Comput. Ind. Eng.* **2021**, *160*, 107601. [CrossRef]

49. Moayedi, H.; Mosavi, A. An Innovative Metaheuristic Strategy for Solar Energy Management through a Neural Networks Framework. *Energies* **2021**, *14*, 1196. [CrossRef]

50. Zhu, M.; Li, Y.; Wang, Y. Design and Experiment Verification of a Novel Analysis Framework for Recognition of Driver Injury Patterns: From a Multi-Class Classification Perspective. *Accid. Anal. Prev.* **2018**, *120*, 152–164. [CrossRef]

51. Meharie, M.G.; Mengesha, W.J.; Gariy, Z.A.; Mutuku, R.N.N. Application of Stacking Ensemble Machine Learning Algorithm in Predicting the Cost of Highway Construction Projects. *Eng. Constr. Archit. Manag.* **2021**; *ahead-of-print*. [CrossRef]

52. Matías, J.M.; Rivas, T.; Martín, J.E.; Taboada, J. A Machine Learning Methodology for the Analysis of Workplace Accidents. *Int. J. Comput. Math.* **2008**, *85*, 559–578. [CrossRef]

53. Mostofi, F.; Toğan, V.; Başağa, H.B. House Price Prediction: A Data-Centric Aspect Approach on Performance of Combined Principal Component Analysis with Deep Neural Network Model. *J. Constr. Eng. Manag. Innov.* **2021**, *4*, 106–116. [CrossRef]

54. Ding, C.; Wu, X.; Yu, G.; Wang, Y. A Gradient Boosting Logit Model to Investigate Driver's Stop-or-Run Behavior at Signalized Intersections Using High-Resolution Traffic Data. *Transp. Res. Part C Emerg. Technol.* **2016**, *72*, 225–238. [CrossRef]

55. Wang, Y.-R.; Yu, C.-Y.; Chan, H.-H. Predicting Construction Cost and Schedule Success Using Artificial Neural Networks Ensemble and Support Vector Machines Classification Models. *Int. J. Proj. Manag.* **2012**, *30*, 470–478. [CrossRef]

56. Farid, A.; Abdel-Aty, M.; Lee, J. A New Approach for Calibrating Safety Performance Functions. *Accid. Anal. Prev.* **2018**, *119*, 188–194. [CrossRef] [PubMed]

57. Jiang, Q. Estimation of Construction Project Building Cost by Back-Propagation Neural Network. *J. Eng. Des. Technol.* **2019**, *18*, 601–609. [CrossRef]

58. Bala, K.; Bustani, S.A.; Waziri, B.S. A Computer-Based Cost Prediction Model for Institutional Building Projects in Nigeria. *J. Eng. Des. Technol.* **2014**, *12*, 519–530. [CrossRef]

59. Peško, I.; Mučenski, V.; Šešlija, M.; Radović, N.; Vujkov, A.; Bibić, D.; Krklješ, M. Estimation of Costs and Durations of Construction of Urban Roads Using ANN and SVM. *Complexity* **2017**, *2017*, 2450370. [CrossRef]

60. Wu, P.; Liu, A.; Fu, J.; Ye, X.; Zhao, Y. Autonomous Surface Crack Identification of Concrete Structures Based on an Improved One-Stage Object Detection Algorithm. *Eng. Struct.* **2022**, *272*, 114962. [CrossRef]

61. Zheng, Y.; Gao, Y.; Lu, S.; Mosalam, K.M. Multistage Semisupervised Active Learning Framework for Crack Identification, Segmentation, and Measurement of Bridges. *Comput.-Aided Civ. Infrastruct. Eng.* **2022**, *37*, 1089–1108. [CrossRef]

62. Hacıefendioğlu, K.; Ayas, S.; Başağa, H.B.; Toğan, V.; Mostofi, F.; Can, A. Wood Construction Damage Detection and Localization Using Deep Convolutional Neural Network with Transfer Learning. *Eur. J. Wood Wood Prod.* **2022**, *80*, 791–804. [CrossRef]

63. Sholevar, N.; Golroo, A.; Esfahani, S.R. Machine Learning Techniques for Pavement Condition Evaluation. *Autom. Constr.* **2022**, *136*, 104190. [CrossRef]

64. Fan, C.-L. Defect Risk Assessment Using a Hybrid Machine Learning Method. *J. Constr. Eng. Manag.* **2020**, *146*, 04020102. [CrossRef]

65. Kim, E.; Ji, H.; Kim, J.; Park, E. Classifying Apartment Defect Repair Tasks in South Korea: A Machine Learning Approach. *J. Asian Archit. Build. Eng.* **2021**, *21*, 2503–2510. [CrossRef]

66. Doğan, N.B. Predicting the Cost Impacts of Construction Nonconformities Using CBR-AHP And CBR-GA Models. Master's Thesis, Middle East Technical University, Ankara, Turkey, 2021.

67. Saqlain, M.; Jargalsaikhan, B.; Lee, J.Y. A Voting Ensemble Classifier for Wafer Map Defect Patterns Identification in Semiconductor Manufacturing. *IEEE Trans. Semicond. Manuf.* **2019**, *32*, 171–182. [CrossRef]

68. Kansara, D.; Singh, R.; Sanghvi, D.; Kanani, P. Improving Accuracy of Real Estate Valuation Using Stacked Regression. *Int. J. Eng. Dev. Res.* **2018**, *6*, 571–577.

69. Neloy, A.A.; Haque, H.M.S.; Islam, M.M.U. Ensemble Learning Based Rental Apartment Price Prediction Model by Categorical Features Factoring. In Proceedings of the 2019 11th International Conference on Machine Learning and Computing—ICMLC '19, Zhuhai, China, 22–24 February 2019; ACM Press: New York, NY, USA, 2019; Volume Part F1481, pp. 350–356.

70. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array Programming with NumPy. *Nature* **2020**, *585*, 357–362. [CrossRef]

71. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Dubourg, V.; Thirion, B.; Grisel, O.; Brucher, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

72. Buitinck, L.; Louppe, G.; Blondel, M.; Pedregosa, F.; Müller, A.C.; Grisel, O.; Niculae, V.; Prettenhofer, P.; Gramfort, A.; Grobler, J.; et al. API Design for Machine Learning Software: Experiences from the Scikit-Learn Project. *arXiv* **2013**, arXiv:1309.0238. [CrossRef]

73. Zhou, P.; El-Gohary, N. Domain-Specific Hierarchical Text Classification for Supporting Automated Environmental Compliance Checking. *J. Comput. Civ. Eng.* **2016**, *30*, 513. [CrossRef]