buildings

MDPI

Article

Customized AutoML: An Automated Machine Learning System for Predicting Severity of Construction Accidents

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

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

Abstract: Construction companies are under pressure to enhance their site safety condition, being constantly challenged by rapid technological advancements, growing public concern, and fierce competition. To enhance construction site safety, literature investigated Machine Learning (ML) approaches as risk assessment (RA) tools. However, their deployment requires knowledge for selecting, training, testing, and employing the most appropriate ML predictor. While different ML approaches are recommended by literature, their practicality at construction sites is constrained by the availability, knowledge, and experience of data scientists familiar with the construction sector. This study develops an automated ML system that automatically trains and evaluates different ML to select the most accurate ML-based construction accident severity predictors for the use of construction professionals with limited data science knowledge. A real-life accident dataset is evaluated through automated ML approaches: Auto-Sklearn, AutoKeras, and customized AutoML. The investigated AutoML approaches offer higher scalability, accuracy, and result-oriented severity insight due to their simple input requirements and automated procedures.

Keywords: construction accident; risk assessment; automated machine learning (AutoML); multi-level severity prediction

check for updates

Citation: Toğan, V.; Mostofi, F.; Ayözen, Y.E.; Behzat Tokdemir, O. Customized AutoML: An Automated Machine Learning System for Predicting Severity of Construction Accidents. *Buildings* **2022**, *12*, 1933. https://doi.org/10.3390/buildings12111933

Academic Editor: M. Amin Hariri-Ardebili

Received: 19 September 2022
Accepted: 3 November 2022
Published: 9 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 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 (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Despite all the technological and regulatory improvements, the construction field is still associated with a high level of risk [1–4]. Growing public concern about safety demands the construction sector to constantly evaluate and revise its safety culture to survive in today's competitive environment. Managing construction risks is vital for construction companies in achieving the expected project outcome for their different stakeholders [5]. Here, the risk is the likelihood of occurrence of a potential hazard item, while the hazard is the possible construction threat that can cause harm [6]. Additionally, the construction sector is constantly challenged by safety risks caused by adopting newly developed technologies, machinery, and equipment. This challenge mandates placing a prompt risk assessment (RA) system that evaluates the risky construction activities to prioritize their severity and plan the appropriate countermeasures against high-risk events. A detailed RA for all construction activities across the various occupations is vital for promoting excellence in safety culture and delivering injury-free projects. It is also equally essential to ensure the responsiveness of the placed RA procedure. Any changes in the construction plan and schedule or arrangement of the construction site affect the construction safety [7], which in turn necessitates revising the RA documents. Therefore, there is a need to implement a detailed RA procedure that promptly evaluates the potential construction risk within the highly complicated [8] and dynamic nature of construction sites.

To this end, various proactive training programs, advanced protective tools and equipment, and technologies have been introduced. Using past accident records from

construction sites helps decision-makers understand the accident insights for the timely implementation of appropriate countermeasures. Additionally, past accident records are vital for making a good prediction, which by itself provides early insight into a safety problem before its occurrence [9]. To improve the accuracy of guesswork within the RA procedure, construction safety literature has investigated different machine learning (ML) algorithms [10,11] for supporting site safety management through different tasks such as construction risk control [7,9,12–14], preventing the fall from height [15–17], and mitigating the transportation safety risks [18–20]. In addition, different ML approaches have been specifically explored to support construction professionals in proactively managing construction safety risks. In this regard, Artificial Neural Networks (ANNs) [7,20,21], Support Vector Machines (SVMs) [18,21,22], and Decision Trees (DTs) [16,23,24] have been widely adopted for proactive safety measurement against transportation and construction accidents. Other ML techniques used for the prevention of transportation and construction accidents are K-nearest Neighbors (KNN) [25], Gradient Boosting (GB) [19,26], and Random Forest (RF) [12,20,26]. For example, Goh and Chua [7] investigated the application of ANN for managing occupational safety and health management system. Their results demonstrated the ability of the ANN model to predict the critical elements that influence the occurrence of severe construction accidents. Ayhan and Tokdemir [14] also demonstrated the ability of ANN to predict the severity outcome of construction accidents while highlighting its high computational cost when trained over a high dimensional dataset. In a recent study, Zhu et al. [27] adopted a more holistic ML approach while evaluating the different algorithms over real-life accident datasets to predict the severity outcome of construction accidents. The algorithms are first pre-processed, which is followed by accident prediction and evaluation phases. It first processes and balances the input accident records and trains Logistic Regression (LR), DT, SVM, Naive Bayes (NB), KNN, RF, multilayer perceptron (MLP), and auto ML classifiers for severity prediction [27]. Then, the evaluation phase determines the critical risk factors and accident assessment rules using person correlation, RF, Principal Component Analysis (PCA), and DT. The severity of construction risks was classified into high and low severity (HS and LS) accidents with an F1 score of 84%. Likewise, a wide range of ML algorithms was employed to monitor, identify, evaluate, and prioritize the risk associated with different construction activities to assess the possible prevention [7,12,15,16,18–20,28–30].

Overall, past studies utilized the real-life construction dataset to demonstrate different ML approaches' effectiveness in predicting the RA procedure [7,12,19,20,25,26]. However, the research on proactive risk assessment using ML techniques cannot be assumed to be adaptable to all kinds of occupational safety datasets. As projects change, so do accident records' attributes, dimensions, and characteristics. The ML models suited to predict accident severity in one study may well collapse under another dataset. The predictive ability of an ML technique is generally limited to datasets with similar dimensions and characteristics. In addition, construction accident records are inconsistent across different projects and throughout the existing body of knowledge case studies. Even under a consistent accident record-keeping system, the increasing data size necessitates changing the existing ML predictor.

Construction safety literature recommends a wide range of conventional (i.e., LR, ANN, KNN, SVM, NB, and DT), hybrid, and ensemble (RF and GB) ML models to enhance the safety level of the construction site and supports decision-making within the RA procedure. While these studies have demonstrated the effectiveness of different ML techniques for risk prevention, they have not been generalized for accident records with other types of attributes. Among the ML methods, no single classifier works best for all real-life datasets [27,31], and most lack the resilience required for real-life datasets. On the one hand, efficient ML deployment demands data science knowledge. On the other hand, a data scientist cannot achieve a safety-based insight as a construction professional. There exists a knowledge gap for the effective employment of the most appropriate ML model

as a decision-support component within the RA procedure for proactive management of construction safety risks.

In response to this demand, automated ML systems are designated for modeling, training, evaluating, and selecting the most accurate ML predictor within a pipeline of algorithms. An automated severity classifier integrates the data management, model selection, severity prediction, and evaluation stages. This results in an abstraction that allows the system to find the fine-tuned architecture (for DNN-based approaches) and hyperparameters for tuning the model for the best prediction outcome. So, when the system is trained, it can classify the severity outcome for a construction hazard list with different safety features. Therefore, in this study, three automated ML methods are to be assessed, named AutoKeras and Auto-Sklearn [32], and a customized AutoML, to facilitate the deployment of the most appropriate ML approach based on the available dataset. The purpose of applying different ML algorithms is to make an ML pipeline for the severity classification of various accident records. Such a system will select the best severity classifier by fine-tuning existing supervised ML algorithms with their best hyperparameters. Hence, this study uses KNN, LR, NB, SVM, DT, RF, GB, and AdaBoost (AB)classifiers within their best optimizations to predict the severity level of construction activities. Accordingly, the four staged research method applied in this study is shown in Figure 1.

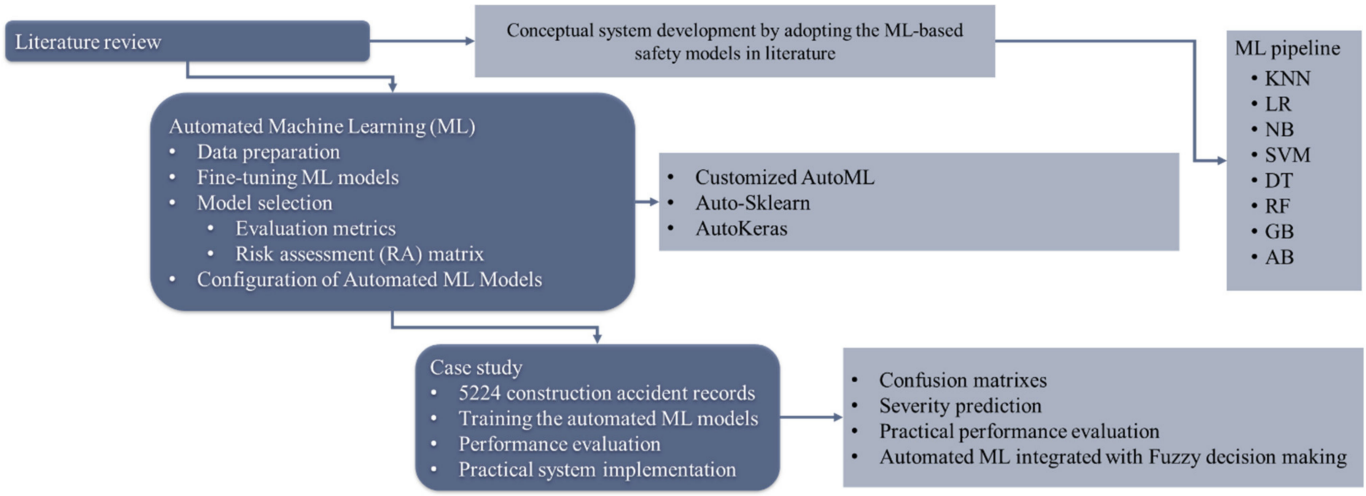


Figure 1. Outline of this study.

The customized automated ML methods are expected to increase the severity prediction accuracy by selecting the best ML classifier. The objective of the investigated automated ML approaches is to bridge the knowledge gap that prevents the effective deployment of ML-based severity predictors by construction professionals. Accordingly, the novelty of this study mainly lies in its first-time adoption of a customizable Automated ML model for predicting the severity outcome of different construction activities alongside the first-time illustration of the practical application of the Auto-Sklearn and AutoKeras safety models.

The main objective of using AutoML approaches is to enable their use over the different datasets, with minimum setup requirements and limited ML-based knowledge requirements for the end-user(s), i.e., safety experts, construction managers, site supervisors, etc. For this, the developed customized AutoML system must have a flexible configuration that allows the selection of the most accurate ML-based construction accident severity predictor that works over different accident datasets. The developed automated ML approach allows construction professionals to dynamically train and evaluate available accident datasets over different ML approaches to select the most accurate ML-based severity predictor. Therefore, after the initiation of the proposed automated ML system by a data science expert, the construction professionals can constantly use the system over different construction accident datasets, having different numbers and types of accident features and

attributes for evaluating the severity outcome of the construction accidents. The automated ML system is expected to be responsive in situations in which the construction accident dataset constantly changes throughout the project and among different projects. To demonstrate the system implementation, its training and evaluation over a real-life construction accident dataset are presented through a case study. In addition, a simple graphical user interface was created to illustrate the user-based input selection as well as to identify the most accurate ML-based severity predictor using the developed system. The literature [9,33] improved the severity prediction of ML by integrating the fuzzy set theory. Thus, this study adopted fuzzy theory to improve the severity prediction of construction accidents. The fuzzy theory is a decision-support methodology proposed by Zadeh [34] that improves decision making when precise information is unavailable. Thus, to improve ML prediction, this study also integrates expert judgment with ML prediction within the fuzzy decision-making space.

2. Methodology

2.1. Configuration of Customized Automated Machine Learning (ML) Model

The outline of the developed automated ML system, including the three main stages, is shown in Figure 2. The analysis is performed with NumPy [35] library package for managing matrices and arrays and the Pandas [36] library for data manipulation and analysis, which were developed in a Python programming environment. The adopted ML algorithms are utilized with the widely used Python libraries TensorFlow [37] and Scikit-Learn [38,39].

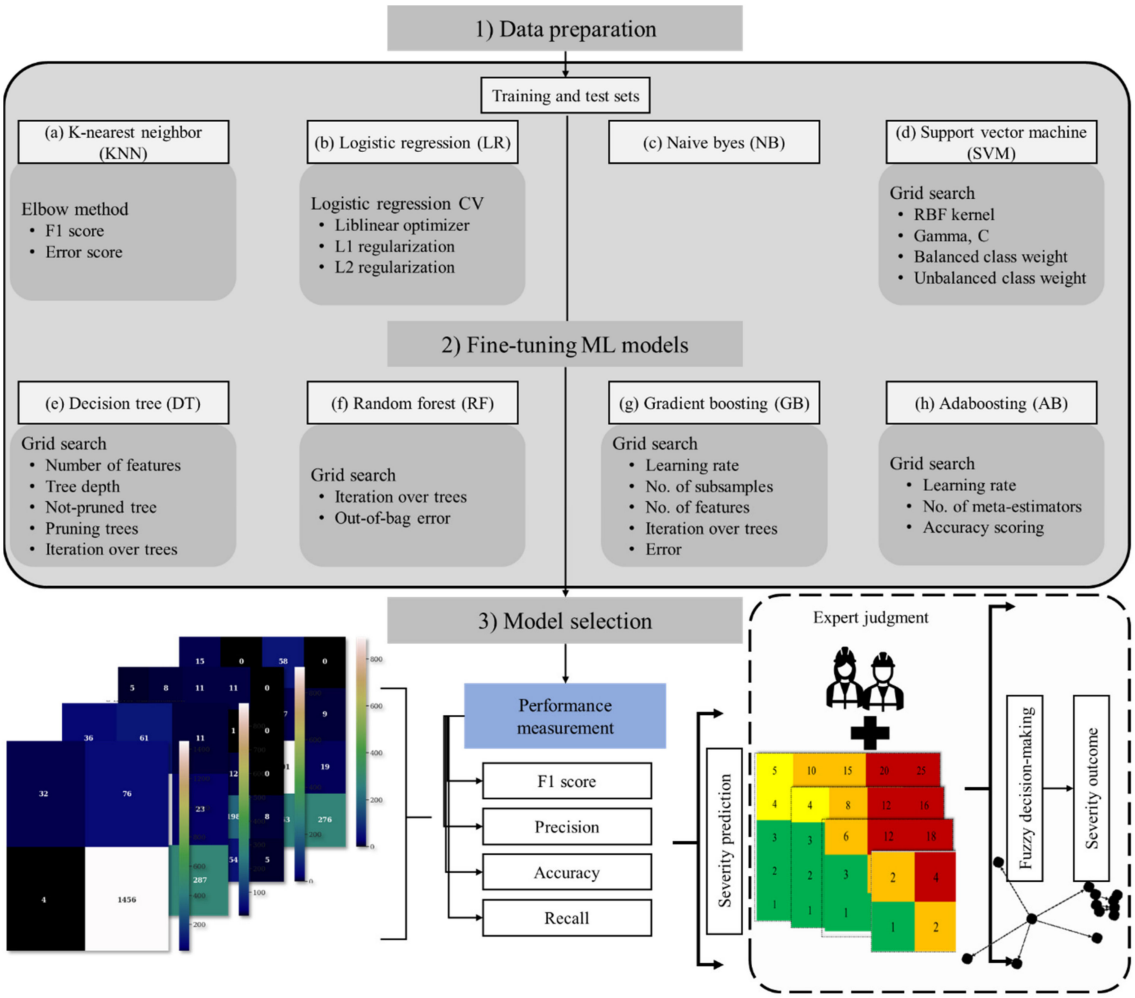


Figure 2. Outline of customized AutoML.

2.1.1. Stage 1: Data Preparation

The first stage involves automating the data preparation. At this stage, the system generates insights about the input accident dataset by performing statistical analysis and generating attribute frequency plots. Additionally, the non-numerical attributes are encoded into the numerical format.

2.1.2. Stage 2: Fine-Tuning ML Models

Subsequently, a pipeline of successful ML methods from the literature is created, which includes optimization, training, and evaluation for selecting the most accurate classifier for predicting the accident severity outcome. This stage includes the configuration of the eight ML approaches KNN, LR, NB, SVM, DT, RF, GB, and AB. In addition, a grid search is employed to fine-tune some ML approaches to search for the conventional values of different hyperparameters. This allows for model configurations with optimized parameters that allow for the most accurate severity prediction. Because it is not computationally efficient to search all possible values for a given hyperparameter, the grid search is configured with commonly used values for the given ML approach. However, after the first initiation of automated ML, construction professionals can repeatedly use the system without the need to adjust these hyperparameters. In addition, each of the eight ML models was specifically configured, considering the most influential hyperparameters that affect the model prediction accuracy based on conventional ML training approaches. Thus, it should be noted that the initial configuration of the ML approaches requires the knowledge of the data science expert.

(a) This work uses the KNN algorithm with a predefined number of neighbors for its simplicity, speed, and accuracy. The model adopts the elbow method and evaluates the KNN performance for 1–25 neighbors, automatically selecting the best number of neighbors. Thus, by measuring the accuracy and F1 scores, the models are fine-tuned with their best number of neighbors.

(b) LR is another ML classifier employed here to determine the probable accident severity outcome. It uses a binary classifier for two-level severity prediction and a "one-vs-rest" for higher levels. Three LR configurations are used here to fine-tune the training procedure and ensure the model's numerical stability: liblinear optimizer, liblinear optimizer with l1 regularization, and liblinear optimizer with l2 regularization.

(c) NB is another probabilistic method employed in the framework. This study train NB while considering the feature attributes as Gaussian features. NB is used for its simplicity and stability.

(d) SVM also handles higher dimensional and unbalanced datasets due to its memory efficiency. To avoid overfitting, the Kernel function set, radial basis function (RBF), is used. RBF is related to the predefined C and gamma parameters. C is used to improve the generalization ability of the model by trading off the simplicity of the decision surface with the misclassification error, while gamma represents the level of influence of each training sample. Our model finds the optimum RBF kernel for each severity classifier using a grid search for the four C and gamma values (1, 5, 10, and 15 and 0.0001, 0.0005, 0.001, and 0.005, respectively). Here, commonly used C and gamma values are selected, responding to a wide range of the dataset. Additionally, different penalties are also assigned to the C value to ensure model ability over the unbalanced dataset.

(e) The customized automated systems also adopt various tree-based methods for their excellent performance with unbalanced datasets due to their hierarchical structure and simple decision-making rules inferred from the feature attributes. The proposed framework first trains the DT with no limits set on the maximum number of depths, features, or levels to obtain the best decision for both the training and test sets, both pruned and not-pruned (DT [P] and DT [NP]). Then, using a grid search, DT performance is measured through iteration between a feature's leaves and depth to their maximums.

(f) RF randomly selects multiple DTs for accident training subsets and fits the model, which is referred to as "bagging." To obtain the optimum number of DTs, the RF is trained here with 20, 50, 100, 150, 200, 300, 400, and 500 trees, and the out-of-bag error associated with each training fold is recorded for the two-, three-, four-, and five-level severity classifiers. Additionally, the RF with extra trees is defined for iterations between 500, 600, 700, 800, 900, and 1000 DTs to enable the system for prediction over large record sizes.

(g) GB also uses DTs for the different severity classification levels. Unlike RF, GB builds a single tree at each stage while combining the decision-making results throughout the DT generation. Then (as with RF), the errors of 20, 50, 100, 150, 200, 300, 400, and 500 DTs are calculated, and the best GB is obtained using grid research. To further improve the fine-tuning duration of GB, the 0.1, 0.01, and 0.001 learning rates are used here, with 1 and 0.5 subsamples for (up to) two, three, and four features.

(h) Finally, AB is adopted for its meta estimation ability. AB first fits the model to the accident dataset and adjusts the weight of additional copies of AB classifiers through the response of misclassified instances. This allows the copied AB classifiers to focus exclusively on more difficult training instances. Here, the AB models are fine-tuned with 100, 150, and 200 estimators, using 0.01 and 0.001 learning rates for scoring accuracy.

Overall, applying the described fine-tuning to the developed ML pipeline increases the algorithms' generalization ability.

2.1.3. Stage 3: Model Selection

After configuring AutoML systems, this study evaluates the performance of automated ML approaches over a real-life accident dataset.

Evaluation Metrics

In this study, conventional accuracy, precision, recall, and F1 scores [40] are used to evaluate the classification performance of the severity classifiers. In this regard, the F1 score is a widely used approach for measuring ML accuracy [41] that comprises both precision and recall and, thus, compared to accuracy, is more appropriate for the unbalanced dataset. To this end, the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) of each severity prediction are recorded. Thus, accuracy is the sum of all TPs and TNs across all severity classes divided by the sum of all instances across all classes (Equation (1)):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

Accordingly, Equation (2) gives the F1 score for all two-, three-, four-, and five-level severity classes:

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

The precision and recall measurements are calculated for the two binary severity classes using Equations (3) and (4):

$$\text{Precision}_{\text{Binary}-\text{class}} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

$$\text{Recall}_{\text{Binary}-\text{class}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

As the proposed model is associated with both balanced and unbalanced datasets along with binary and multilevel classifications, multi-class precision is the sum of all TP classes over the total number of TPs and FPs in all severity classes (Equation (5)):

$$\text{Precision}_{\text{Multi-class}} = \frac{\sum_{i=2}^{\text{classes}} TP_i}{\sum_{i=2}^{\text{classes}} TP_i + FP_i} \quad (5)$$

Similarly, multi-class recall represents the total number of TPs in all severity classes over the total number of TPs and FNs in all classes (Equation (6)):

$$\text{Recall}_{\text{Multi-class}} = \frac{\sum_{i=2}^{\text{classes}} TP_i}{\sum_{i=2}^{\text{classes}} TP_i + FN_i} \quad (6)$$

The accuracy of each of the five accident severity classes—very low, low, medium, high, and very high severity (VLS, LS, MS, HS, and VHS)—are obtained to ensure that no misrepresentation results from the classification accuracy of the unbalanced dataset.

Risk Assessment (RA) Matrix

In the qualitative RA analysis, the RA matrix is used to evaluate the risk category of potential hazards based on their probability of occurrence and expected severity outcome or impact. The core idea of the RA matrix is to prioritize risk items to manage critical risk items beyond the acceptable range. Utilizing different sizes of RA matrices changes the upper extremities of acceptability, such as low probability and medium impact in a 3×3 RA matrix. However, defining the acceptability range and size of the RA matrix depends on different factors, including the risk tolerance of the company, its health and safety culture, and the availability of information. In addition, the complexity and uncertain nature of construction projects require dynamic hazard identification and RA, which directly affects short- and long-term decisions. Throughout the different construction stages, the RA procedure requires changing the acceptability range. For example, a 2×2 or 3×3 RA at the construction initiation and planning stages would be sufficient to prioritize the critical risk items. However, at the construction execution stage, more detailed RA can benefit from the detailed information of construction activity delivery and thus provide more categories of the construction risks using a 4×4 or 5×5 matrix. Correspondingly, the proposed automated ML allows multilevel severity prediction based on user preferences. Hence, the severity predictions of construction accidents in terms of accuracy, recall, precision, and F1 scores were evaluated to find the best ML classifier for the four (2×2, 3×3, 4×4, and 5×5) risk matrices. As it is shown in Figure 2, after classifying accident severity levels into the five classes (VLS, VS, MS, HS, and VHS), the customized AutoML system adopts the risk tolerance level for each of the four (2×2, 3×3, 4×4, and 5×5) RA matrices, showing different levels of risk categories.

2.2. Fuzzy Decision Making

Considering the importance of construction safety, where human life is at stake, RA requires 100% prediction accuracy. Unless the ML model achieves 100% accuracy, its prediction needs to be integrated with the expert's opinion. To further develop the applicability of the customized AutoML for the use of construction professionals, it is integrated with expert judgment within a fuzzy decision-making space. Correspondingly, this study integrated the severity predicted by the system with the expert judgment within the fuzzy decision-making space. This is to consider deviation from the expert's judgment and to address the uncertainty associated with using the ML prediction with an accuracy of less than 100%.

A fuzzy space accepts the collection of different severity classes in linguistic format or severity ranges (fuzzy sets), along with associated membership values between zero and one. These membership values are defined using different geometric shapes as membership functions, such as the S-curve, trapezoidal, and triangular [9,42]. The horizontal access of the fuzzy decision-making space represents the severity predicted by the ML, whereas the vertical axis corresponds to the associated decision within the geometry of the adopted membership value. Because construction management studies [9,43] preferred the

triangular fuzzy membership function, our study assigns a triangular space to each fuzzy set. Accordingly, to better illustrate the deployment of the developed customized AutoML at the construction site, the severity prediction of a single accident record is detailed. For this, the obtained severity predicted by the customized AutoML is adjusted within a triangular fuzzy membership. The required values of triangular membership functions were configured based on the study by Ayhan and Tokdemir [9] (Table 1).

Table 1. Fuzzy membership of severity classes predicted by AutoML.

Severity Classifier	3 × 3	4 × 4	5 × 5
VLS	–	(1, 2, 3)	(1, 2, 3)
LS	(1, 2, 4)	(2, 3, 4)	(2, 3, 4)
MS	(3, 4, 5)	(3, 4, 5)	(3, 4, 5)
HS	(4, 6, 6)	(4, 6, 6)	(4, 5, 6)
VHS	–	–	(5, 6, 6)

The predicted severity with the customized AutoML system is to be defined within the fuzzy membership space, as detailed in Table 1. Upon initiation of the fuzzy space of the model, the expert judgment is to be defined within the secondary fuzzy universe. Therefore, the incorporated expert judgment is defined within the triangular universe based on the Conoco-Philips pyramid (Table 2).

Table 2. Fuzzy membership of severity classes determined by expert judgment.

Expert Judgment	3 × 3	4 × 4	5 × 5
LS	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)
MS	(2, 3, 4)	(2, 3, 4)	(2, 3, 4)
HS	(3, 5, 5)	(3, 5, 5)	(3, 4, 5)
VHS	–	–	(4, 5, 5)

As shown in Table 2, the severity classes defined by expert judgment were defined within four levels (LS, MS, HS, and VHS). Here, the VLS judgment is excluded from the expert's decision-making space because it greatly influences the ML's decision in the HS and VHS cases. This ensures that the system prioritizes severe events identified by ML. Therefore, the high-severity decision spaces are integrated with the least severity membership functions within all the created fuzzy decision-making spaces. The system was developed by considering ML's prediction as the base decision that allows for reflecting the expert's judgment as the secondary decision about the accident severity outcome. Accordingly, the outcome of the accident severity prediction by ML and expert judgment within the defined fuzzy decision-making spaces were defined based on the severity levels defined by AutoML predictions (Table 1). Therefore, the resulting severity predictions are defined within three-, four-, and five-level severity classes using different risk matrix sizes.

2.3. Configuration of Auto-Sklearn and AutoKeras Models

Besides the developed customized AutoML, this study investigates the performance of existing AutoML system as Auto-Sklearn [32] and AutoKeras [44]. Auto-Sklearn is an open-source automated ML that uses Bayesian Optimization for searching and selecting the best ML classifier from a pipeline of 15 classifier algorithms. Additionally, before model training, 14 feature pre-processing algorithms, along with the necessary scaling, encoding, and handling of missing parameters, are performed. Auto-Sklearn obtains good generalization by stacking ML models within ensemble structures. Moreover, using past knowledge (meta-learning) speed up the ML selection for a new dataset. Likewise, AutoKeras is another open-source and AutoML-based system that runs parallelly over CPU and GPU to find the best DNN configuration using Neural architecture search (NAS). Bayesian optimization is used to guide the search for efficient neural network design. In addition, the performance of the customized AutoML system, Auto-Sklearn, and AutoKeras

approaches are to be investigated for predicting accident severity impact as a necessary procedure of RA.

3. Case Study

3.1. Data Description

The study utilizes 5,224 accident records obtained from 73 construction projects to demonstrate the applicability of the proposed system. A detailed data description can be obtained in a study by Ayhan and Tokdemir [9]. The descriptions of the 12-input data with the adopted project characteristic (PC) attributes are listed in Table 3.

Table 3. Description of the construction accident dataset.

Variable	Attribute	Label	Identifier	Frequency
PC1	Occupation	1	Rough work crew	2140
		2	Mechanical assembly crew	1419
		3	Finishing work crew	649
		4	Repairman	412
		5	Others	221
		6	Construction equipment operator	141
		7	Engineer	140
		8	Administrative affairs	102
PC2	Day time	1	PM	2789
		2	AM	2435
PC3	Activity	1	Daily activities	906
		2	Re-bar/formwork installation	805
		3	Assembly works	787
		4	The usage of hand-tool/equipment	707
		5	Lifting operations	470
		6	Welding/hot works	456
		7	Working with chemicals/MEP works	348
		8	Finishing works	341
		9	Transportation/construction equipment/the usage of vehicle	133
		10	Concreting	92
		11	Repair/maintenance works	59
		12	Working at height	24
		13	Excavation works	24
		14	Field measurement works	21
		15	Testing works	14
		16	Working with chemical materials	12
		17	Mobilization on/off-site	9
		18	Landscaping works	7
		19	Cable-pipe assembly/ working with containments	6
		20	Geotechnical works	2
		21	Material drops	1
PC4	Risky behaviors	1	Inability to perceive external risks	3190
		2	Violation of Safe work policy	505
		3	Incorrect physical movement	438
		4	Incorrect/absence of safe work policy	425
		5	Tending to use a shortcut	329
		6	Incorrect usage of equipment, hand-tool	214
		7	Others	119

Table 3. *Cont.*

Variable	Attribute	Label	Identifier	Frequency
PC5	Hazardous cases	1	Nonconformance (NCRs) in the working environment	2652
		2	NCRs in safety protection measures	818
		3	Others	754
		4	Mechanic hazards/NCRs	382
		5	NCR in usage of hand-tool/equipment/construction equipment	347
		6	Natural hazards	176
		7	Radiation exposure	51
		8	Chemical hazards	26
		9	Fire, explosion	18
PC6	Human factors	1	Problems resulting from the incorrect management system	1493
		2	Problems resulting from the unbalanced workload	1187
		3	Insufficient skill and perception	1096
		4	Physical disability	441
		5	Faulty management system	357
		6	Problems related to education level	285
		7	Others	275
		8	Psychological disability	90
PC7	Workplace factors	1	Problems with the method of statement	1037
		2	Inadequate incident analysis systems	959
		3	Inadequate communication (general)	873
		4	The lack of a management system	691
		5	Inadequate maintenance/repairment mechanism	485
		6	Insufficient control, tracking	445
		7	Others	365
		8	Lack of Occupational Health and Safety (OHS) training	202
		9	Incorrect protection measures	107
		10	Incorrect recruitment procedures	60
PC8	Experience	1	1–3 month	1832
		2	3–6 month	1255
		3	1 month	1028
		4	6–12 month	882
		5	12–24 month	227
PC9	Age	1	(17.924, 25.6]	1902
		2	(25.6, 33.2]	1711
		3	(33.2, 40.8]	862
		4	(40.8, 48.4]	509
		5	(48.4, 56.0]	203
		6	(56.0, 63.6]	34
		7	(86.4, 94.0]	1
		8	(71.2, 78.8]	1
		9	(63.6, 71.2]	1
PC10	Year	1	(2015.8, 2016.4]	2628
		2	(2014.6, 2015.2]	1791
		3	(2016.4, 2017.0]	778
		4	(2013.997, 2014.6]	27
PC11	Month	1	(0.989, 3.75]	1419
		2	(3.75, 6.5]	1401
		3	(6.5, 9.25]	1279
		4	(9.25, 12.0]	1125
PC12	Day	1	(0.97, 8.5]	1494
		2	(8.5, 16.0]	1282
		3	(16.0, 23.5]	1264
		4	(23.5, 31.0]	1184

In addition, each of the accident features was pre-processed using typical data preparation methods, such as format changes and scaling, before loading into the ML models. For example, the attributes of days, months, years, and ages were categorized into four intervals. To ensure the responsiveness of the automated ML approaches for different classes, the severity outcome column was categorized into two, three, four, and five severity levels (Figure 3).

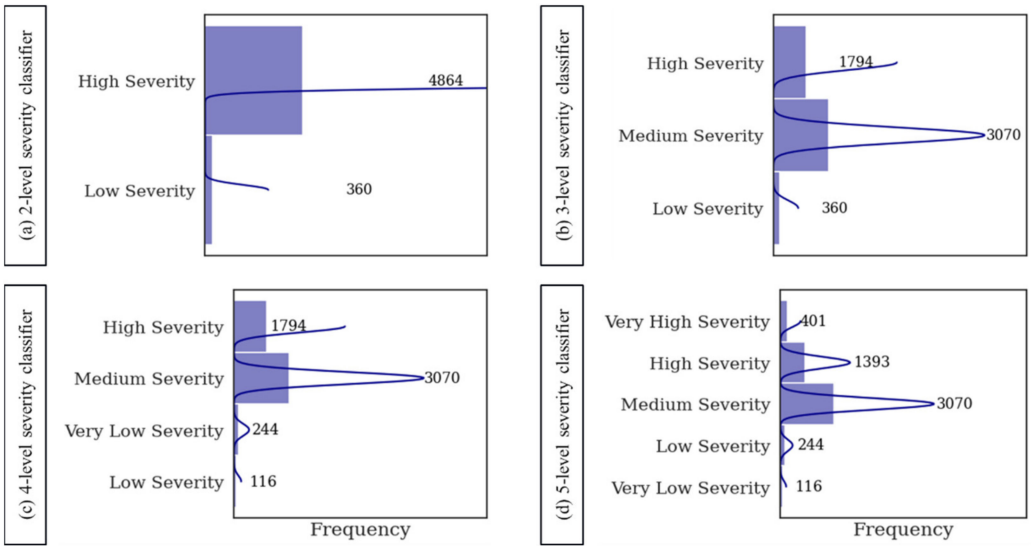


Figure 3. Classification of construction accident severity levels.

As it is shown in Figure 3, construction accidents are assigned within different severity classes based on the outcome risk. Accordingly, of the 5224 accidents considered here, 59% required first aid and 26% medical intervention, 15% of the minor accidents were caused by material or workday loss, 244 were near-miss accidents, and 4 involved fatalities. Furthermore, the system is to be utilized in two stages; first, training the system; second, predicting the severity impact of different hazard items using the trained system. These two stages are detailed using the described real-life construction accident dataset.

3.2. Training Machine Learning (ML) Models

Upon initiating the customized AutoML, the system is trained using the user input dataset (Figure 4). The system first accepts the entire input dataset and then uses 70% of the dataset for training the models while maintaining 30% of the input dataset to test the prediction performance of the ML approaches over the unseen dataset. However, for splitting the training and test datasets, instead of random splits, a stratified shuffle split is used. A stratified shuffle split is a strategy for returning the randomized folds by preserving the percentage of samples for each class without interfering with the input dataset. It should be noted that different training and test sets were used for the two, three, four, and five severity levels.

As it is shown in Figure 4, at the initial system setup, the directory path of the dataset along with a user interface that allows for minor adjustments (e.g., size of risk level matrix for the prediction output and unnecessary accident features) are defined by the data scientist expert. For example, the initial system configuration allows construction professionals to decide between training the automated ML system or using the previously trained models. In addition, the system allows adjusting the input feature columns, such as redundant columns. In the present example, the severity column is recorded using both categorical and numerical attributes. Similarly, the input dataset contains the accident/incident classification column, which is similar to the severity output column and thus needs to be removed from the input dataset. Thus, the input requirement of the customized AutoML system facilitates its implementation for the construction team. Afterward, the eight ML

models were applied to the 5224 construction accident records to predict the severity impact associated with different construction activities for the four (2×2, 3×3, 4×4, and 5×5) RA matrices. Here, the customized AutoML is automatically iterated and evaluated, whereby the fine-tuned hyperparameters are presented in Table 4.

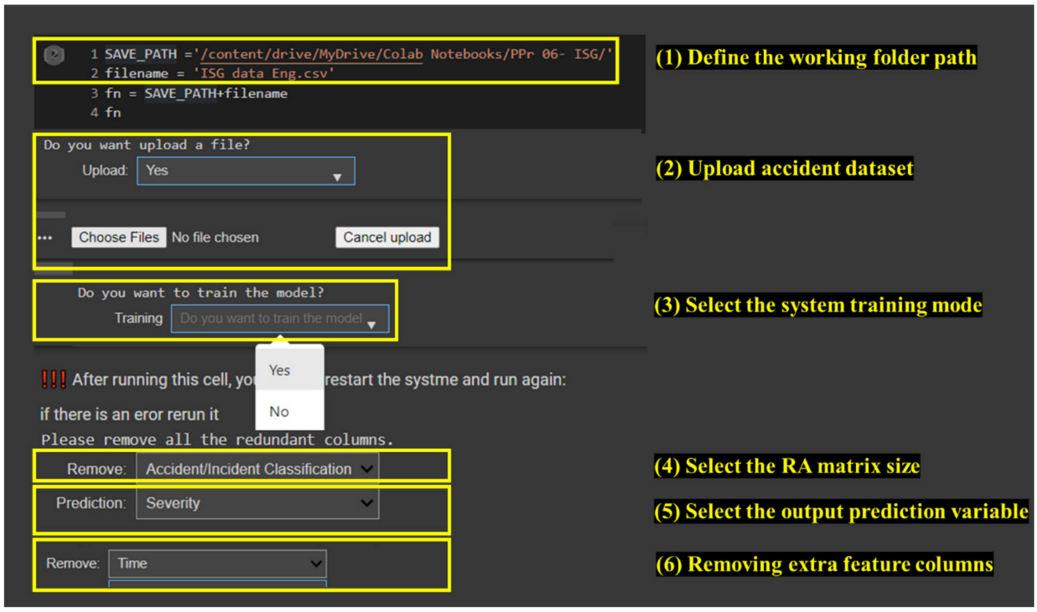


Figure 4. Sample script integrated user input space for model training.

Table 4. Automatically tuned hyperparameters.

Classifier	Hyperparameter	2×2	3×3	4×4	5×5
KNN	K	19	19	24	19
SVM	C	10	15	15	15
	Gamma	0.005	0.005	0.005	0.005
DT	Nodes	53	161	61	57
	Depth	5	7	5	5
RF	Trees	300	300	300	500
GB	Trees	150	100	70	80
	Learning Rate	0.1	0.1	0.1	0.1
AB	Learning Rate	0.1	0.1	0.1	0.1
	Estimators	100	100	100	100

Once each of the eight algorithms is fine-tuned, they are used for predicting the severity outcome of construction accidents. The classified severity impact of the identified hazards is to be used to assist the decision-making professionals at the construction site.

3.3. Severity Classification Results

As stated earlier, the customized AutoML is designed to select the ML that achieves the highest accuracy. Hence, Figure 5a shows the performance of eight ML algorithms within the proposed customized AutoML over the 1567 accidents in the test set. Additionally, to compare the prediction performance of the automated ML with those of Auto-Sklearn, and AutoKeras, their prediction accuracy is compared with the studied ML approaches.

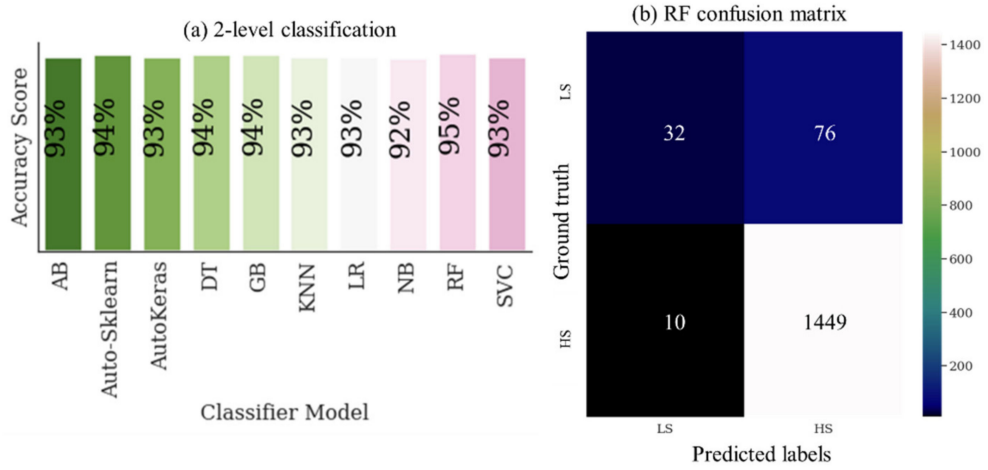


Figure 5. (**a**) 2-level accuracy performance. (**b**) Confusion matrix for RF.

Based on Figure 5, customized AutoML for two-level severity prediction achieved a prediction accuracy between 92% and 95 %. Here, the system used RF as the severity predictor, considering its higher prediction accuracy. The system also specifies the results of other evaluation metrics such as precision (94%), recall (95%), and F1 score (93%). Based on the results, AutoML outperformed Auto-Sklearn and AutoKeras in terms of prediction accuracy. Similarly, the three-level classification performance was evaluated, as shown in Figure 6.

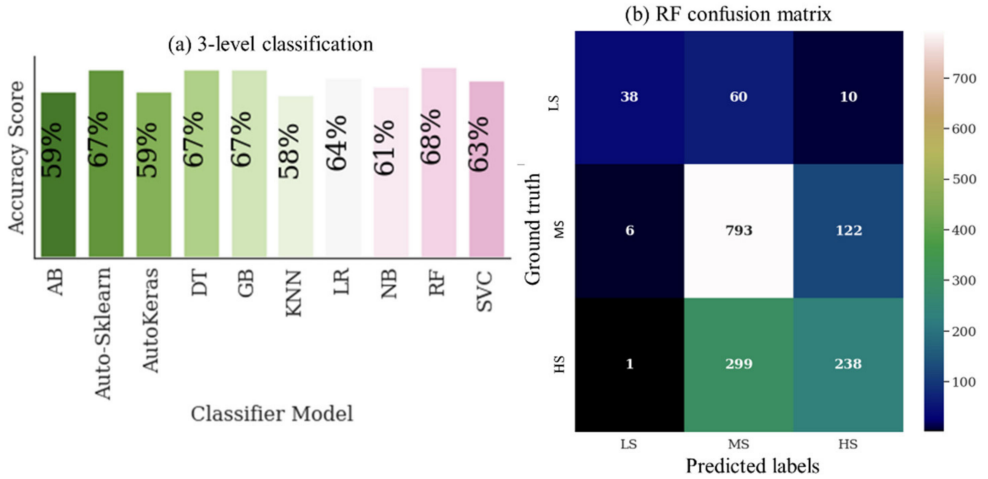


Figure 6. (**a**) Three-level accuracy performance. (**b**) Confusion matrix for RF.

The prediction accuracy of the three-level ML model was considerably lower than that of the two-level model. The ML approaches achieved a prediction accuracy between 58% and 68%, whereas RF outperformed the other predictors. Thus, the system selected the RF for severity prediction while noting its performance in terms of accuracy (68%), precision (68%), recall (68%), and F1 score accuracy (66%). Figure 7 shows the prediction performance of the adopted ML approaches for the four-level severity prediction of construction accidents.

In the four-level severity classification, the range of prediction accuracy remained between 58% and 68%, similar to the three-level classification. However, GB demonstrated improved prediction performance, while the prediction accuracy of RF decreased by 1%. Therefore, AutoML selected GB while recording its performance metrics as 68% accuracy, 67% precision, 68% recall, and 69% F1 score.

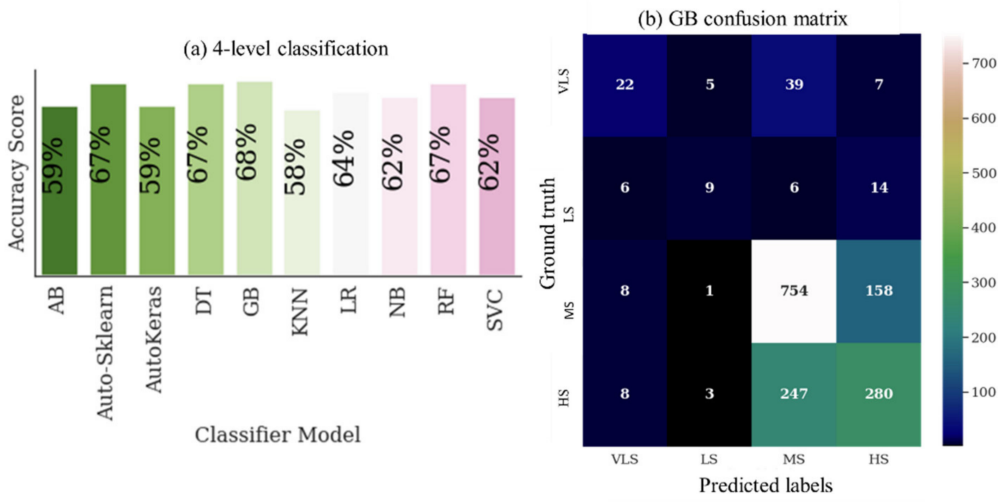


Figure 7. (**a**) Four-level accuracy performance. (**b**) Confusion matrix for GB.

Similarly, the accuracy performance of the proposed customized is evaluated for a five-level severity prediction (Figure 8).

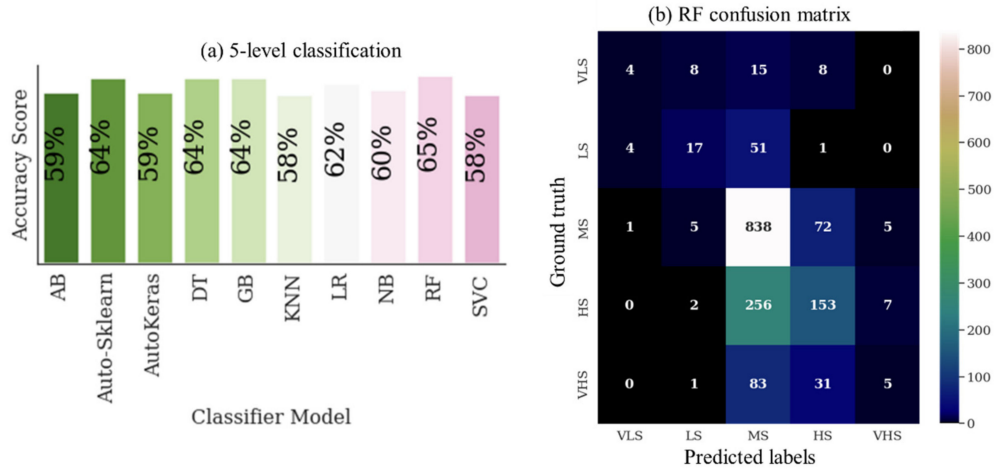


Figure 8. (**a**) Five-level accuracy performance. (**b**) Confusion matrix for RF.

The range of prediction accuracy of the five level-severity predictors was reduced between 58% and 65%. Here, RF resulted in the best prediction accuracy of 65% and was thus selected for the five-level severity prediction of construction accidents.

The analysis of the confusion matrix highlights the system's effectiveness in selecting the fine-tuned configuration of the most accurate severity predictor. Additionally, the automated ML is compared with AutoSklearn and AutoKeras in terms of their prediction accuracy. In the investigated AutoML systems, customized AutoML achieved either the same or higher accuracy than AutoSklearn and AutoKeras. Both customized AutoML and AutoSklearn displayed medium performance for underrepresented classes (LS and VLS) due to the considerably unbalanced classes of LS events, while AutoKeras was poorly performed.

After training the ML models, they are saved into the defined folder path and thus can be accessed upon request. In this regard, the proposed user input area for entering the characteristics of the specific event is illustrated in Figure 9. The customized AutoML requires loading accident data and defining the output prediction variable.

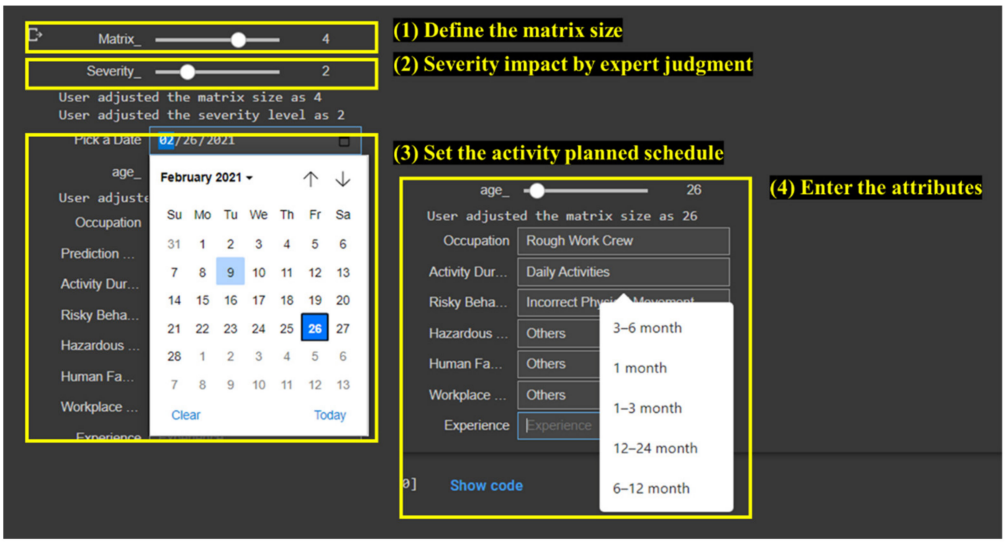


Figure 9. User input for requesting a severity prediction for a particular event outcome.

As displayed in Figure 9, the customized AutoML receives information about a particular event and predicts the severity outcome by calling the stored train models. Here, the system automatically employs the best severity predictor while allowing the user to cross-validate the prediction using the other trained models (Figure 10).

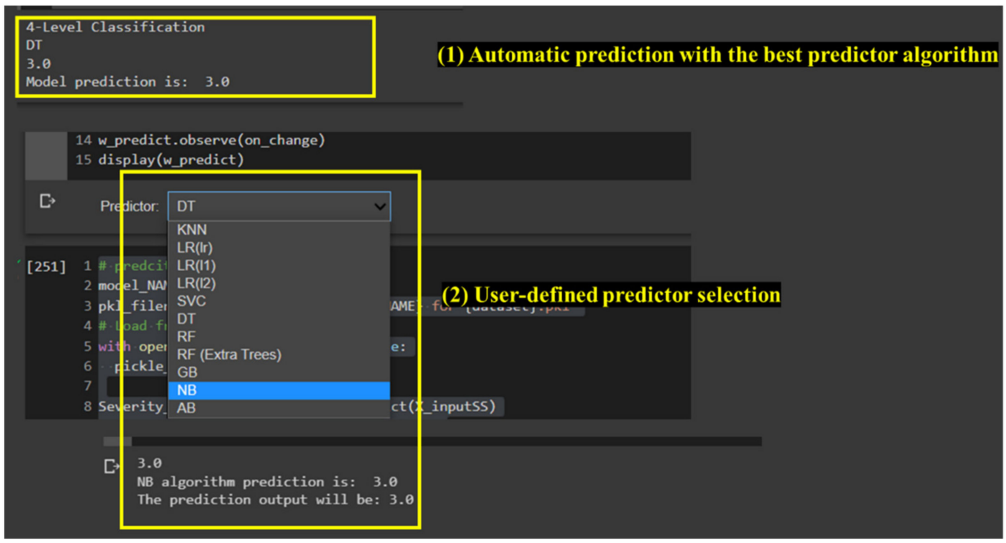


Figure 10. Severity prediction of newly entered accident dataset using a customized AutoML system.

To evaluate the computational efficiency of the different severity levels, the time required for each severity classifier was measured using the 1567 accidents in the test set (Figure 11).

As shown in Figure 11, the customized AutoML utilized less than 150 s, while Auto-Sklearn and AutoKeras used 175 s and 155 s, respectively. The computational time of AutoSklearn and AutoKeras are high due to controlling all the alternative classifiers. It should be noted that AutoSklearn is initiated by defining the user time as more than 120 s, while AutoKeras is initiated by specifying the number of iterations. Therefore, both can be configured for the different running times that, in turn, affect their selection accuracy. To have a comparable comparison, AutoSklearn is evaluated with 180 s as the running time limit, while each run time is also limited to a maximum of 30 s. Subsequently, for a five-class severity prediction, AutoSklearn iterated 15 ensemble classifiers within 184 s. The developed system for a three- and five-level classification consumed 207 and 224 s,

respectively. However, this includes all the model training, fine-tuning, optimization, confusion matrix calculation, and graph drawing times. Consequently, to further improve the customized AutoML, the fine-tuning procedure needs to be optimized.

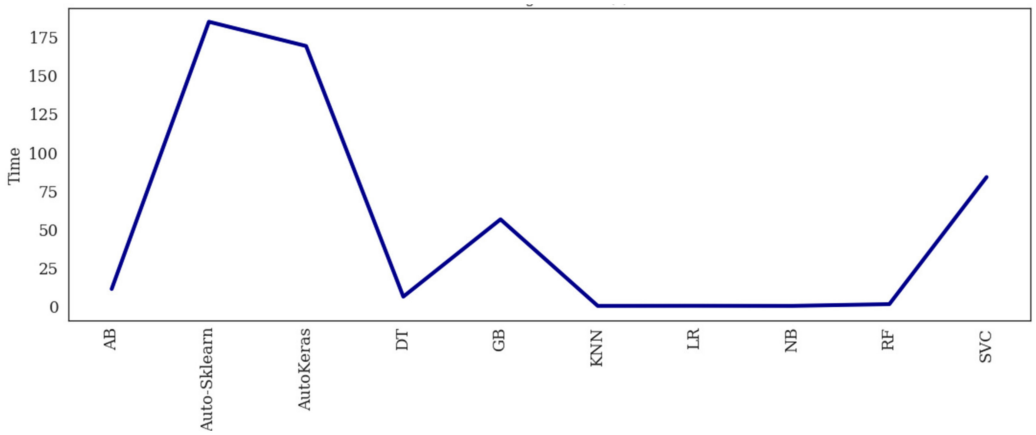


Figure 11. The computational time in the 5-level severity classification task.

4. Contribution of Automated Machine Learning (ML) System

4.1. Comparison with Literature

The existing safety literature evaluated different single and ensemble ML approaches over other safety datasets. However, a successful ML approach over a specific dataset (specific size, dimension, and feature) may not perform well over a different dataset. For example, KNN results in the best accuracy for the small-size dataset, while RF results are best for the larger dataset. Zhu et al. [27] achieved a high classification accuracy of 84.5% by automating data management and ML selection for 571 construction accident records. With this small dataset, their model proved equally efficient in classifying construction accidents into LS and HS severity levels. However, our study placed the investigated ML algorithms in the literature, including those investigated by Zhu et al. [27], within a single pipeline to develop a system for construction professionals with limited data science knowledge. This allowed the system to select the most accurate ML severity predictor, which removes the need for manual selection and fine-tuning of different ML approaches. In addition, the accuracy of the ML-based severity classifiers was evaluated for multilevel severity predictions appropriate for different sizes of the RA matrix.

Additionally, the results of the present study are compared with a study by Ayhan and Tokdemir [14] using a similar construction accident dataset. Ayhan and Tokdemir [14] improved ANN accuracy for severity prediction from 60% to 69% while using the latent class clustering analysis (LCCA) to control heterogeneity. The present study achieved better severity prediction over the same accident dataset using the customized AutoML. The developed system trained and evaluated the eight ML approaches and accordingly selected RF for two-level severity prediction whereby displayed 95% precision, recall, and classification accuracy with a 94% F1 score. Automating the successful ML approaches within a single pipeline guarantees the selection of the most accurate ML predictors from the customized pipeline.

4.2. Developing a Decision-Support RA Tool

The integration of the customized AutoML prediction and expert judgment in the fuzzy decision space is displayed in Table 5.

The resulting severity prediction within the developed fuzzy space is illustrated in Figure 12. Based on the predicted severity by the fuzzy system, all class levels categorized the hazard item as HS.

Table 5. Fuzzy membership of severity classes predicted by RMI.

Fuzzy Decision	3 × 3	4 × 4	5 × 5
VLS	–	(1, 7, 13)	(1, 7, 13)
LS	(1, 8, 16)	(8, 12, 16)	(8, 12, 16)
MS	(12, 16, 20)	(12, 16, 20)	(12, 16, 20)
HS	(11, 23, 35)	(11, 23, 35)	(11, 17, 25)
VHS	–	–	(15, 25, 35)

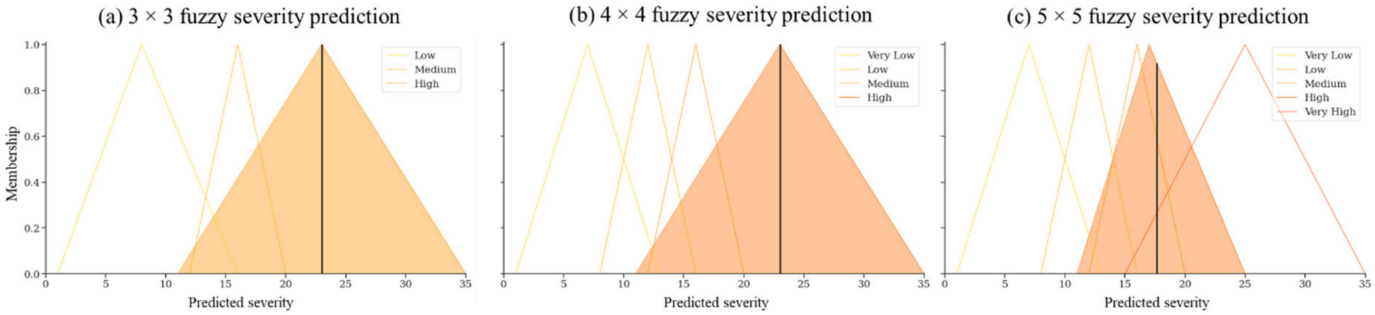


Figure 12. Sample fuzzy decision making, integrating customized AutoML with expert judgment.

In addition, while 3 × 3 and 4 × 4 level RA assigned the risk as a high priority, the 5 × 5 classification assigned it as a second priority risk. Moreover, the smaller fuzzy space in the 5 × 5 level RA matrix indicates the higher certainty achieved using the 5 × 5 RA matrix.

To illustrate the fuzzy decision about the severity impact on the RA matrix, the possible probability of the sample hazard items is evaluated (Figure 13). To evaluate the model's ability over different output sizes, the severity outcome of the construction accident is divided into two, three, four, and five severity levels. Within the described customized AutoML, the construction accident record is automatically pre-processed, modeled, finally fine-tuned, and evaluated to select the most accurate accident severity predictor.



Figure 13. Application of predicted multi-class severity within different RA matrices.

As outlined in Figure 13, the outcome of the developed Automated ML system can be interpreted within conventional RA matrices. Here, the size of the matrix depends on the construction stage where the developed system is utilized. At the initial construction stages, the construction work packages are not often detailed; thus, RA is performed with limited information. Thence, the developed automated system needs to be utilized over three-level severity classification (Figure 13a), benefiting from higher prediction accuracy at the expense of reducing the level of details. At this stage, the system prioritizes the probable risks with high severity to direct the countermeasures at the critical construction activities.

4.3. Practical Use of the Research

As the project progresses, the information about the construction tasks and associated details also increases. The increased knowledge about the details of each construction activity enhances the effectiveness of the planned risk control or risk mitigation strategies. Therefore, at the later stages of construction projects, in addition to probable risks, the constriction activities with the possibility of a medium and high severity outcome need to be addressed. At this stage, selecting between four- or five-level (Figure 13b,c) automated severity predictor depends on the amount of information available as well as the risk tolerance of a construction company. In this respect, the risk tolerance of a construction company is defined by its health and safety culture as well as its ability and available resources to implement the planned risk response strategies. Irrespective of a construction company's adopted risk tolerance level, it can benefit from the developed automated ML system to support the RA procedure at different stages of the construction procedure. To this end, Figure 14 suggests the process for employing the proposed automated ML system within the RA procedure.

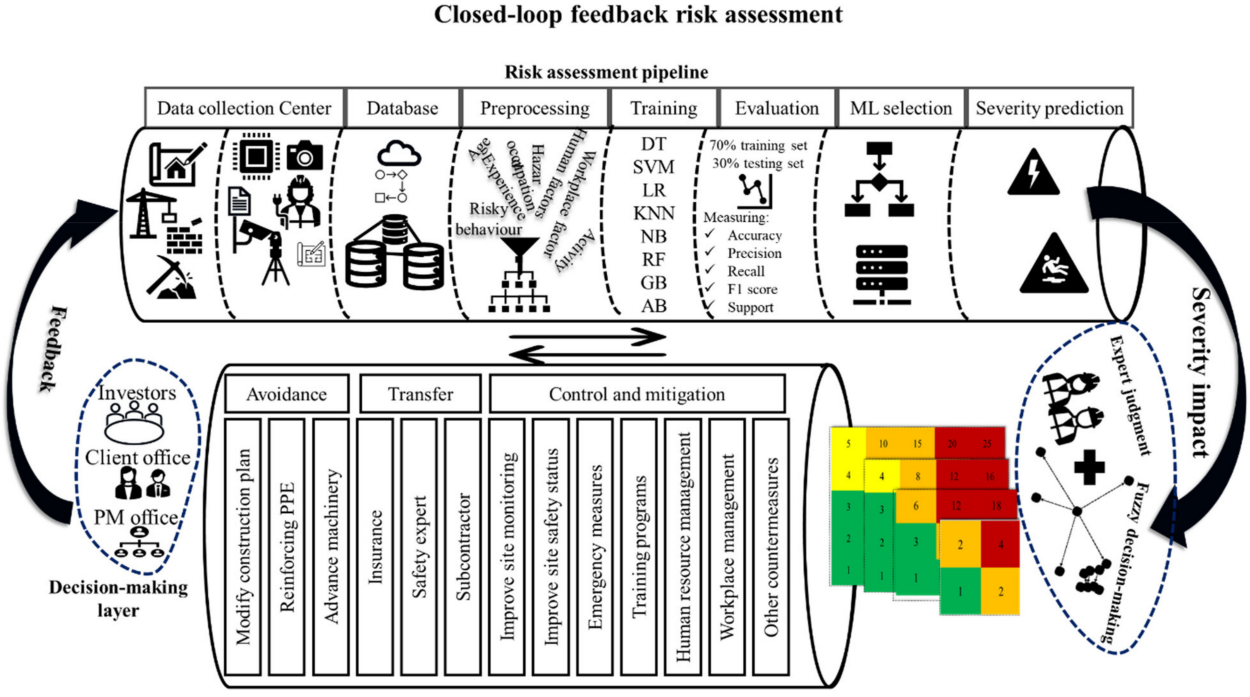


Figure 14. Developed automated ML system within a closed-loop feedback RA procedure.

As shown in Figure 14, integrating the developed automated ML system within the RA procedure allows construction professionals to dynamically monitor the construction site, collect the safety records, and predict the three-, four-, and five-level severity outcomes of different construction activities. The flexible multilevel severity classifications based on user preferences facilitate their utilization for different levels of risk matrices.

Changes in the accident record-keeping system do not necessitate system reconfiguration, as the system automatically finds the best ML appropriate for a given dataset. Therefore, safety experts can utilize the automated ML system and obtain the best prediction and insights upon an increase in the number of accident records or its dimension. This facilitates the constant utilization of construction severity predictors by construction professionals at different project stages, as the developed automated ML method ensures the selection of the most accurate accident severity prediction in response to a given construction accident dataset. In addition, the configuration of the system outcomes allows for the integration of the user decision with that of the automated ML system. The ability to integrate expert judgment within the system outcome is useful in situations where the scarcity of the accident dataset reduces the accuracy of ML. It supports construction professionals in benefiting from an adoptive RA procedure based on different decision-making criteria, including the health and safety culture of the company, its risk tolerance, and available financial, technological, and human resources in devising the adopted risk response strategy. Additionally, at the decision-making layer, the developed system enhances the visibility of the site safety condition with details about the reliability of the identified possible and probable critical construction activities. The attained informed guesswork assists the safety professionals in their short- and long-term decision-making for implementing required countermeasures. As a result, integrating the expert judgment with the predicted severity by the automated ML at one side, along with the decision about devising the adopted risk response strategies, structures a closed-loop feedback RA and control strategy based on the constant observations and measurement from the condition of the construction site within different stages of its lifecycle. Furthermore, employing the Automated ML approaches (AutoKeras, AutoSklearn, and customized AutoML) allows scaling the ML application dynamically across different projects. The scalability of the automated ML system is even more important for smaller construction projects (such as buildings), as they are often delivered by medium- to small-sized construction companies with a higher rate of occupational injury compared with larger companies [45,46].

Ultimately, the discussed advantages allow construction professionals to dynamically benefit from the most accurate severity prediction without the knowledge of training, fine-tuning, or evaluating different ML models. However, the developed automated ML requires the knowledge of a data scientist at system initiation. Compared with existing ML-based predictors, automated ML procedures allow site engineers and construction professionals to obtain the best prediction accuracy of different ML approaches with limited ML-related knowledge.

5. Conclusions

This study developed a novel automated ML system for predicting the safety outcomes of different construction activities that can be utilized by construction professionals with limited data science knowledge. Successful ML approaches in the literature are used within three automated ML approaches: AutoSklearn, AutoKeras, and customized AutoML. The core idea of adopting automated ML approaches is enabling their use over different datasets, with minimum setup requirements and limited ML-based knowledge requirements for the end-user(s), i.e., safety experts, construction managers, site supervisors, etc. The developed system uses the ML pipeline that automates the fine-tuning, training, and evaluation of different ML approaches for selecting the best severity predictor that is appropriate for a given construction dataset.

Overall, the developed customized AutoML provides higher scalability, accuracy, and result-oriented severity insight due to its simple input requirement and automated procedure. The customized AutoML system allows all the project team members to dynamically use the model for severity prediction of construction accidents throughout different project stages. Compared with AutoSklearn and AutoKeras, the customized AutoML achieved better accuracy. This study recognizes that predicting the severity outcome of construction accidents with higher accuracy enables the employment of appropriate proactive safety

responses, such as the application of realistic RA methodologies and improved on-time safety warnings.

Ultimately, this study intended to bridge the knowledge gap between the data scientist and construction professionals for deploying the suitable ML approach at construction sites, alongside increasing the accuracy and reliability of the RA procedure. The AutoML approaches benefit from the prediction accuracy of different ML methods, while their application is flexible over different datasets. Compared with developed AutoML systems (AutoSklearn and AutoKeras), customizing an automated ML system according to a specified task offers flexibility while still benefiting from the higher accuracy attained from the Automated ML selective process. In terms of computational efficiency, the customized AutoML can be further improved using advanced optimization strategies for fine-tuning the ML algorithms. AutoML approaches enable Safety professionals with limited ML experience to train high-quality severity predictors specific to their application. It also allows them to control the ML-based predictors over different scenarios and evaluate the severity outcome of different arrangements accordingly. Furthermore, the system allows for the integration of expert judgment with the severity prediction of AutoML to improve its prediction performance.

There are a few shortcomings of the developed automated ML severity predictor. Firstly, the automated severity classifiers are experimenting with highly unbalanced accident datasets. This reduces the prediction accuracy amongst the underrepresented severity classes. The class-balance methods (i.e., Oversampling and Undersampling approaches) could improve the prediction accuracy of the underrepresented severity classes. Secondly, the performance of some of the ML approaches within the customized ML pipeline could be influenced by the high dimensionality of the accident dataset: few accident records are being explained with many feature columns. Thus, the customized system may be further improved by accommodating automated dimensionality reduction approaches such as Principal Component Analysis (PCA). This may also improve the overall computational time of the automated ML approaches. Thirdly, the dataset collected from accidents within different construction projects, and thus explicitly experimenting with the developed automated ML system over larger and smaller construction projects such as mega projects, provide a more precise evaluation of the scalability of the developed automated ML system. Lastly, the developed automated ML system needs to be investigated over the dataset obtained from different geographies and using different data collection systems to better compare the robustness of the developed automated system with the existing ML approaches.

Author Contributions: Conceptualization, Formal Analysis, Writing—Original Draft Preparation, Writing—Review and Editing, V.T. and F.M.; Resources, Data Curation, Writing—Review and Editing, Y.E.A. and O.B.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Kang, Y.; Asce, M.; Siddiqui, S.; Suk, S.J.; Chi, S.; Kim, C.; Asce, A.M. Trends of Fall Accidents in the U.S. Construction Industry. *J. Constr. Eng. Manag.* **2017**, *143*, 04017043. [CrossRef]
2. Chiang, Y.-H.; Wong, F.K.-W.; Liang, S. Fatal Construction Accidents in Hong Kong. *J. Constr. Eng. Manag.* **2017**, *144*, 04017121. [CrossRef]
3. Guo, S.; Tang, B.; Liang, K.; Zhou, X.; Li, J. Comparative Analysis of the Patterns of Unsafe Behaviors in Accidents between Building Construction and Urban Railway Construction. *J. Constr. Eng. Manag.* **2021**, *147*, 04021027. [CrossRef]
4. Kim, T.; Chi, S. Accident Case Retrieval and Analyses: Using Natural Language Processing in the Construction Industry. *J. Constr. Eng. Manag.* **2019**, *145*, 04019004. [CrossRef]

5. Bahamid, R.A.; Doh, S.I.; Khoiry, M.A.; Kassem, M.A.; Al-Sharafi, M.A. The Current Risk Management Practices and Knowledge in the Construction Industry. *Buildings* **2022**, *12*, 1016. [CrossRef]
6. Colmenarejo, J.I.S.; Camprubí, F.M.; González-Gaya, C.; Sánchez-Lite, A. Power Plant Construction Projects Risk Assessment: A Proposed Method for Temporary Systems of Commissioning. *Buildings* **2022**, *12*, 1260. [CrossRef]
7. Goh, Y.M.; Chua, D. Neural Network Analysis of Construction Safety Management Systems: A Case Study in Singapore. *Constr. Manag. Econ.* **2013**, *31*, 460–470. [CrossRef]
8. Jahan, S.; Khan, K.I.A.; Thaheem, M.J.; Ullah, F.; Alqurashi, M.; Alsulami, B.T. Modeling Profitability-Influencing Risk Factors for Construction Projects: A System Dynamics Approach. *Buildings* **2022**, *12*, 701. [CrossRef]
9. Ayhan, B.U.; Tokdemir, O.B. Predicting the Outcome of Construction Incidents. *Saf. Sci.* **2019**, *113*, 91–104. [CrossRef]
10. Kononenko, I.; Kukar, M. Introduction. In *Machine Learning and Data Mining*, 1st ed.; Horwood Publishing: Chichester, UK, 2007; pp. 1–36. ISBN 978-1-904275-21-3.
11. Poh, C.Q.X.; Ubeynarayana, C.U.; Goh, Y.M. Safety Leading Indicators for Construction Sites: A Machine Learning Approach. *Autom. Constr.* **2018**, *93*, 375–386. [CrossRef]
12. Tixier, A.J.P.; Hallowell, M.R.; Rajagopalan, B.; Bowman, D. Application of Machine Learning to Construction Injury Prediction. *Autom. Constr.* **2016**, *69*, 102–114. [CrossRef]
13. Ayhan, B.U.; Tokdemir, O.B. Safety Assessment in Megaprojects Using Artificial Intelligence. *Saf. Sci.* **2019**, *118*, 273–287. [CrossRef]
14. Ayhan, B.U.; Tokdemir, O.B. Accident Analysis for Construction Safety Using Latent Class Clustering and Artificial Neural Networks. *J. Constr. Eng. Manag.* **2020**, *146*, 04019114. [CrossRef]
15. Piao, Y.; Xu, W.; Wang, T.-K.; Chen, J.-H. Dynamic Fall Risk Assessment Framework for Construction Workers Based on Dynamic Bayesian Network and Computer Vision. *J. Constr. Eng. Manag.* **2021**, *147*, 04021171. [CrossRef]
16. Mistikoglu, G.; Gerek, I.H.; Erdis, E.; Mumtaz Usmen, P.E.; Cakan, H.; Kazan, E.E. Decision Tree Analysis of Construction Fall Accidents Involving Roofers. *Expert Syst. Appl.* **2015**, *42*, 2256–2263. [CrossRef]
17. Goh, Y.M.; Binte Sa'adon, N.F. Cognitive Factors Influencing Safety Behavior at Height: A Multimethod Exploratory Study. *J. Constr. Eng. Manag.* **2015**, *141*, 04015003. [CrossRef]
18. Yongchang, M.; Chowdhury, M.; Sadek, A.; Jeihani, M. Real-Time Highway Traffic Condition Assessment Framework Using Vehicle–Infrastructure Integration (VII) With Artificial Intelligence (AI). *IEEE Trans. Intell. Transp. Syst.* **2009**, *10*, 615–627. [CrossRef]
19. 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]
20. 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]
21. Tango, F.; Botta, M. Real-Time Detection System of Driver Distraction Using Machine Learning. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 894–905. [CrossRef]
22. 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]
23. Sugumaran, V.; Ajith Kumar, R.; Gowda, B.H.L.; Sohn, C.H. Safety Analysis on a Vibrating Prismatic Body: A Data-Mining Approach. *Expert Syst. Appl.* **2009**, *36*, 6605–6612. [CrossRef]
24. Kwon, O.H.; Rhee, W.; Yoon, Y. Application of Classification Algorithms for Analysis of Road Safety Risk Factor Dependencies. *Accid. Anal. Prev.* **2015**, *75*, 1–15. [CrossRef] [PubMed]
25. Farid, A.; Abdel-Aty, M.; Lee, J. A New Approach for Calibrating Safety Performance Functions. *Accid. Anal. Prev.* **2018**, *119*, 188–194. [CrossRef] [PubMed]
26. Farid, A.; Abdel-Aty, M.; Lee, J. Comparative Analysis of Multiple Techniques for Developing and Transferring Safety Performance Functions. *Accid. Anal. Prev.* **2019**, *122*, 85–98. [CrossRef]
27. Zhu, R.; Hu, X.; Hou, J.; Li, X. Application of Machine Learning Techniques for Predicting the Consequences of Construction Accidents in China. *Process Saf. Environ. Prot.* **2021**, *145*, 293–302. [CrossRef]
28. Hegde, J.; Rokseth, B. Applications of Machine Learning Methods for Engineering Risk Assessment—A Review. *Saf. Sci.* **2020**, *122*, 104492. [CrossRef]
29. Oguz Erkal, E.D.; Hallowell, M.R.; Bhandari, S. Practical Assessment of Potential Predictors of Serious Injuries and Fatalities in Construction. *J. Constr. Eng. Manag.* **2021**, *147*, 04021129. [CrossRef]
30. Pan, Y.; Zhang, L. Roles of Artificial Intelligence in Construction Engineering and Management: A Critical Review and Future Trends. *Autom. Constr.* **2021**, *122*, 103517. [CrossRef]
31. Ayhan, M.; Dikmen, I.; Talat Birgonul, M. Predicting the Occurrence of Construction Disputes Using Machine Learning Techniques. *J. Constr. Eng. Manag.* **2021**, *147*, 04021022. [CrossRef]
32. Sobrecueva, L. *Automated Machine Learning with AutoKeras Deep Learning Made Accessible for Everyone with Just Few Lines of Coding*; Packt Publishing: Birmingham, UK, 2021; ISBN 9781800567641.
33. Ung, S.T.; Williams, V.; Bonsall, S.; Wang, J. Test Case Based Risk Predictions Using Artificial Neural Network. *J. Saf. Res.* **2006**, *37*, 245–260. [CrossRef] [PubMed]
34. Zadeh, L.A. Fuzzy Sets. *Inf. Control.* **1965**, *8*, 338–353. [CrossRef]

35. 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]
36. McKinney, W. Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference; SciPy, Austin, TX, USA, 28 June–3 July 2010; Volume 1, pp. 56–61.
37. Xu, T.; Jin, X.; Huang, P.; Zhou, Y.; Lu, S.; Jin, L.; Pasupathy, S. Early Detection of Configuration Errors to Reduce Failure Damage. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), Savannah, GA, USA, 2–4 November 2016; ISBN 978-1-931971-33-1.
38. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Müller, A.; Nothman, J.; Louppe, G.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2012**, *12*, 2825–2830.
39. Buitinck, L.; Louppe, G.; Blondel, M.; Pedregosa, F.; Mueller, A.; 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.
40. Chinchor, N.; Sundheim, B. Evaluation Metrics. 1993. Available online: https://aclanthology.org/M93-1007.pdf (accessed on 18 September 2022).
41. Chicco, D.; Jurman, G. The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation. *BMC Genom.* **2020**, *21*, 6. [CrossRef]
42. Arditi, D.; Tokdemir, O.B.; Suh, K. Effect of Learning on Line-of-Balance Scheduling. *Int. J. Proj. Manag.* **2001**, *19*, 265–277. [CrossRef]
43. Zeng, J.; An, M.; Smith, N.J. Application of a Fuzzy Based Decision Making Methodology to Construction Project Risk Assessment. *Int. J. Proj. Manag.* **2007**, *25*, 589–600. [CrossRef]
44. Jin, H.; Song, Q.; Hu, X. Auto-Keras: An Efficient Neural Architecture Search System. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 19–23 August 2018; pp. 1946–1956. [CrossRef]
45. Ozmec, M.N.; Karlsen, I.L.; Kines, P.; Andersen, L.P.S.; Nielsen, K.J. Negotiating Safety Practice in Small Construction Companies. *Saf. Sci.* **2015**, *71*, 275–281. [CrossRef]
46. McVittie, D.; Banikin, H.; Brocklebank, W. The Effects of Firm Size on Injury Frequency in Construction. *Saf. Sci.* **1997**, *27*, 19–23. [CrossRef]