sustainability

MDPI

Article

Modeling the Cause-and-Effect Relationships between the Causes of Damage and External Indicators of RC Elements Using ML Tools

Roman Trach [1,2,*], Galyna Ryzhakova [3], Yuliia Trach [1,2], Andrii Shpakov [3] and Volodymyr Tyvoniuk [1,2]

[1] Institute of Civil Engineering, Warsaw University of Life Sciences, 02-776 Warsaw, Poland; yuliia_trach@sggw.edu.pl (Y.T.); volodymyr_tyvonyuk@sggw.edu.pl (V.T.)
[2] Institute of Civil Engineering and Architecture, National University of Water and Environmental Engineering, 33028 Rivne, Ukraine
[3] Department of Management in Construction, Kyiv National University of Construction and Architecture, 03037 Kyiv, Ukraine
* Correspondence: roman_trach@sggw.edu.pl

Abstract: Reinforced concrete (RC) structures are used in a wide range of applications, including high-rise buildings, nuclear power plants, oil and gas platforms, bridges, and other infrastructure. However, over time, RC structures can be subject to deterioration and damage, particularly from exposure to weather and environmental conditions, heavy traffic loads, and other factors. Regular inspections, diagnosing the condition, maintenance, and repair can help to mitigate the effects of degradation and extend the life of the structure. The task of this study was to determine the possible causes of the defects of the RC elements based on the identification of external indicators using the ML tools. This study created and compared the performance of four ML models, namely, Support Vector Regression (SVR), decision trees (DTs), random forest (RF), and Artificial Neural Networks (ANNs). The first comparison showed a rather low performance of all models, with a slight advantage of the ANN model. Later, six ANN models were optimized to obtain a higher level of performance. The next step of this study was the training, validation, and testing of ANN models. Analysis of MAPE and R2 metrics showed that the ANN model with an Adaptative Moment (ADAM) loss function and sigmoid activation had the best results (MAPE 3.38%; R^2 0.969). The novelty of the study consisted of the development of the ML model, which is based on the use of ANNs, and allows for the establishment of cause-and-effect relationships in the diagnosis of the technical condition of the RC elements. The advantage of using ANN to solve this problem is the possibility to obtain a forecast in the form of continuous values. Moreover, the model can be used further without retraining, and it can make predictions on datasets it has not yet "seen". The practical use of such a model will allow for the diagnosis of some causes of defects during a visual inspection of structures.

Keywords: external indicators; causes of defects; cause-and-effect relationships; artificial neural network; machine learning; reinforced concrete

Citation: Trach, R.; Ryzhakova, G.; Trach, Y.; Shpakov, A.; Tyvoniuk, V. Modeling the Cause-and-Effect Relationships between the Causes of Damage and External Indicators of RC Elements Using ML Tools. *Sustainability* **2023**, *15*, 5250. https://doi.org/10.3390/su15065250

Academic Editors: Yi Bao, Yazhou (Tim) Xie, Shan Li and Wenwei Wang

Received: 18 February 2023
Revised: 8 March 2023
Accepted: 15 March 2023
Published: 15 March 2023

1. Introduction

Reinforced concrete structures are the most common type of construction that utilizes a combination of concrete and reinforcement materials, typically steel, to provide increased strength and durability. According to the report "Design and Control of Concrete Mixtures" by the Portland Cement Association, concrete is the most widely used construction material in the world, with approximately 7.5 billion m^3 of concrete being produced each year [1]. Reinforced concrete comprises a significant portion of the production. Steel reinforcement is embedded within the concrete, creating a composite material that is strong in both compression and tension.

Reinforced concrete structures are used in a wide range of applications, including high-rise buildings, nuclear power plants, offshore oil and gas platforms, bridges, and other infrastructure. According to a report by Global Market Insights, the global RC market was valued at over 45 billion USD in 2020 and is expected to grow at a compound annual growth rate (CAGR) of around 5% from 2021 to 2027 [2]. According to the Ukrainian State Road Agency, Ukravtodor, there are over 47,000 bridges in Ukraine, with RC bridges consisting of approximately 57% of the total [3]. RC is a common material used in bridge construction in Ukraine and around the world due to its strength, durability, and relatively low cost. However, over time, RC bridges can be subject to deterioration and damage, particularly from exposure to weather and environmental conditions, heavy traffic loads, and other factors [4].

The degradation of RC structures refers to the deterioration or loss of the structural integrity of these structures over time due to a variety of factors. Some of the common causes of degradation in RC structures include corrosion of the reinforcing steel, chemical attack, physical damage, fire damage, environmental factors. The degradation of RC structures can be a serious concern, as it can compromise the safety and functionality of the structure. Regular inspections, diagnosing the condition, maintenance, and repair can help to mitigate the effects of degradation and extend the life of a structure.

There are several methods for diagnosing the condition of RC structures. Figure 1 shows some of the most used methods. The specific method or combination of methods used for diagnosing an RC structure will depend on factors such as the size and complexity of the structure, the severity of any damage, and the goals of the diagnostic process.

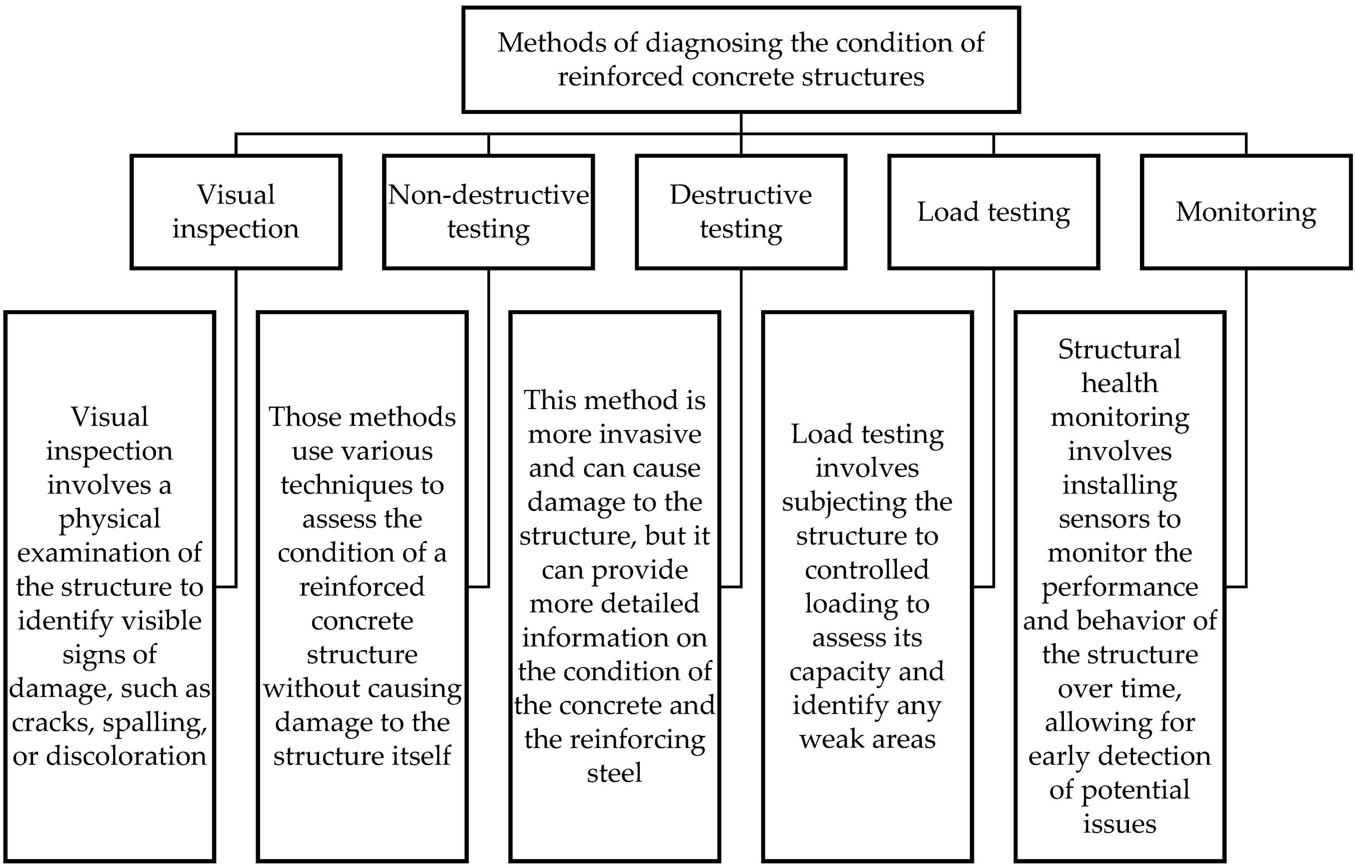


Figure 1. The most commonly used methods of diagnosing the condition of RC structures.

Visual inspection is a basic method of diagnosing the condition of RC structures that involves a physical examination of the structure to identify visible signs of damage or deterioration. During a visual inspection, a qualified inspector will closely examine the exterior and interior of the structure, looking for any indications of damage, such as cracks, spalling, or discoloration. On the basis of the received information, it is possible to investigate cause-and-effect relationships with external indicators of damage and the causes for their appearance.

Reinforced concrete structures can experience damage for a variety of reasons. To ensure the long-term performance of RC structures, it is important to consider the potential causes of damage and take appropriate measures during design, construction, and maintenance to prevent or mitigate them. Cause-and-effect relationships describe the connection between events or actions, where one event (the cause) leads to another event (the effect) [5]. Cause-and-effect relationships can be straightforward or complex, and can occur in a wide range of contexts, from physics and engineering to social sciences and economics. Identifying cause-and-effect relationships can be important for understanding the underlying mechanisms of a process, predicting outcomes, and making informed decisions. In some cases, cause-and-effect relationships may be simple and direct. For example, if a person touches a hot stove, the heat (cause) will result in a burn (effect). However, in many cases, cause-and-effect relationships are more complex and involve multiple variables or factors that interact with each other. It is important to note that correlation does not always indicate causation. Just because two events or variables are associated with each other does not necessarily mean that one causes the other. It is important to carefully consider the evidence and analyze the context to determine if there is a true cause-and-effect relationship.

The method of cause-and-effect relationships can help describe the relationship between external indicators of damage and the causes of damage to RC structures. By identifying the cause-and-effect relationship between these two factors, engineers can develop a better understanding of the underlying mechanisms that lead to the deterioration of RC structures. For example, if visual inspection reveals cracks on the surface of an RC structure, the cause-and-effect relationship can be explored to determine what factors may have contributed to the development of these cracks. Possible causes may include factors such as exposure to weather and environmental conditions, chemical exposure, design flaws, or construction defects. By identifying the cause-and-effect relationship between the presence of cracks and the underlying factors that led to their development, engineers can develop targeted strategies for repair and maintenance that address the root cause of the problem. Currently, there are different machine learning (ML) methods for establishing cause-and-effect relationships in engineering sciences [6]. Those methods involve using mathematical techniques to analyze data and identify correlations between variables, creating digital models of engineering systems and simulating the effects of various inputs or changes.

An analysis of the literature showed that the most often used methods to establish cause-and-effect relationships between the causes and external manifestations of damage to the RC structures were the methods based on fuzzy logic. Terenchuk et al. [7] used the modification of the Takagi–Sugeno–Kanga fuzzy network to model the evaluation system. The model can solve the classification problem under conditions of uncertainty and can set rules as a function of inputs. A model was also created based on the use of neural fuzzy networks to solve the problem of supporting decision-making regarding the assessment of the impact of repair and construction works on the technical condition of an object [8]. In addition, researchers use classical tools of fuzzy set theory to create a computer-assisted crack diagnosis system [9] and quantitative assessment of the condition of elements of RC structures [10].

Summarizing, in the long-term performance of RC structures, it is important to consider these potential causes of damage and take the appropriate measures to prevent or mitigate them [11]. Understanding cause-and-effect relationships is essential in engineering sciences for identifying problems, designing solutions, and improving the performance and reliability of structures [12]. Thus, the task of this study was to determine the possible causes of the defects of the RC elements based on the identification of external indicators using the ML tools. The novelty of the study consisted in the development of the ML model, which is based on the use of ANNs, and allows for the establishment of cause-and-effect relationships in the diagnosis of the technical condition of the RC elements. The advantage of using ANN to solve this problem is the possibility to obtain a forecast in the form of continuous values. Moreover, the model can be used further without retraining, and it can make predictions on datasets it has not yet "seen". The practical use of such a model will allow to diagnose some causes of defects during a visual inspection of structures.

The article has the following structure: Section 2 analyses cause-and-effect relationships between external indicators of damage and the causes of their appearance, as well as the possibility to use ML methods to detect defects in RC structures; Section 3 presents the process of datasets creation and the process of ML modeling; Section 4 presents the results of a productivity comparison of four ML models and uses an ANN to establish cause-and-effect relationships in the diagnosis of the technical condition of the RC structures. Section 5 presents the discussion; Section 6 presents conclusions of the study.

2. Background

2.1. Causes of Damage in Reinforced Concrete Structures

The literature presents various classifications of the causes of defects in RC structures. Woodson defined ten causes of distress and deterioration of RC: construction errors, design errors, accidental loadings, chemical reactions, erosion, shrinkage, freezing and thawing, settlement and movement, weathering, and temperature changes [13]. ENV 1504-9:1997 divided common causes into defects in concrete (mechanical, chemical, and physical causes) and reinforcement corrosion (carbonation, stray currents, and corrosive contaminants) [14]. A guide to concrete repair prepared for the Bureau of Reclamation of the United States Department of the Interior addresses 13 common causes of damage: faulty design, construction defects, excess of concrete mix water, alkali–aggregate reaction, sulfate deterioration, abrasion–erosion damage, deterioration caused by cyclic freezing and thawing, corrosion of reinforcing steel, cavitation damage, cracking, structural overloads, acid exposure, and multiple causes [15]. Delatte used another classification of concrete damage based on causes and mechanisms which were observed most frequently by a practicing civil engineer: static overloading, chemical attack, impact, fire, malicious damage, earthquakes [16]. Douglas and Ransom summarized the main sources of defect RCs and their percentage division in the United Kingdom: detachment 10%, biochemical decay 20%, dampness 50%, movement 20% [17]. The Circular defined eight causes of distress and deterioration of RC: plastic shrinkage, plastic settlement, thermal expansion and contraction, drying shrinkage, freezing and thawing, corrosion of reinforcement, alkali–aggregate sulfate attack and reaction material [18]. The Guide by ACI Committee "Causes, Evaluation, and Repair of Cracks in Concrete Structures" defined nine causes: thermal stresses, drying shrinkage, chemical reaction, corrosion of reinforcement, weathering, construction overloads, externally applied loads, design flaws, poor construction practices, errors in design and detailing [19].

The classification of the causes of defects by the stages of the lifecycle of RC structures is as follows [20]. Design-related causes. These are defects that result from problems in the original design of the structure (inadequate reinforcement, improper detailing, or insufficient concrete cover). Construction-related causes. These are defects that result from issues during the construction of the structure (poor workmanship, inadequate curing, or improper compaction) and defects that result from issues with the materials used in the construction of the structure (poor quality concrete or reinforcement, or the use of incompatible materials). Operational causes. These are defects that result from the way the structure is used or maintained (overloading the structure, inadequate maintenance, or exposure to chemical agents) and defects that result from the exposure of the structure to the environment (corrosion due to exposure to saltwater, freeze–thaw damage due to exposure to freezing and thawing cycles, or erosion due to exposure to water).

By identifying the specific cause or causes of defects in RC structures, engineers can develop targeted strategies for repair and maintenance that address the underlying problem.

It is important to diagnose the causes of defects as accurately as possible to prevent further deterioration and ensure the safety and longevity of the structure.

2.2. The Cause-and-Effect Relationships between External Indicators of Damage and the Causes for Their Appearance

The external indications of damage to RC may include cracks in the concrete surface, rust stains or corrosion on the reinforcing steel, spalling or flaking of the concrete surface, deformation or displacement of the concrete structure, and water infiltration or leakage. These provide valuable information about the potential causes of damage. For example, cracks on the surface of concrete may be an indication of excessive loading, exposure to extreme temperatures, or the corrosion of reinforcing steel. Spalling or flaking of the concrete surface may be caused by freeze–thaw cycles or exposure to chemicals. Concrete spalling or chipping on the surface of a bridge deck may be an indication of the corrosion of reinforcing steel caused by exposure to de-icing salts or other chemicals. Horizontal or vertical cracks on the walls of a building may be an indication of excessive loading or the settling of the foundation, or of thermal expansion and contraction. Cracks on the surface of a concrete slab may be an indication of excessive loading, shrinkage caused by drying, or differential settlement. Discoloration or staining of the concrete surface may be an indication of water infiltration, chemical exposure, or environmental pollution. Bulging or bowing of a retaining wall may be an indication of an inadequate design or construction, or of excessive lateral loading.

The relationships between external indicators and the causes of damage in RCs have already come to the attention of researchers. It is necessary to include the most important study guide by the ACI Committee, “Guide for Making a Condition Survey of Concrete in Service”; moreover, in it, a very large database was collected on the external indications of damage to concrete, illustrated by photographs [21]. Their purpose is to attempt to standardize the reporting of the condition of the concrete in a structure. According to the Guide, there are several types of cracks that can appear on the surface of concrete. Some of the most common types of cracks include:

- Plastic shrinkage cracks. These are thin, shallow cracks that usually appear within the first few hours or days after the concrete has been poured.
- Drying shrinkage cracks. These cracks usually appear a few weeks or months after the concrete has been poured. They are caused by the gradual drying and shrinking of the concrete as it loses moisture over time.
- Settlement cracks. These cracks are caused by uneven settling or subsidence of the ground beneath the concrete. They can appear as diagonal cracks that taper off at the surface.
- Overload cracks. These cracks are caused by excessive weight or load on the concrete, which can cause it to crack and deform.
- Structural cracks. These cracks are caused by defects or weaknesses in the concrete structure itself, such as inadequate reinforcement or poor construction practices. They can appear as wide, deep cracks that may be accompanied by deformation or the displacement of the concrete.

In Ukraine, the National Standard “Guidelines regarding the inspection of building objects to determine and assess their technical condition” have been developed and are used. This Guidelines establishes dependencies between defects and damage and the possible causes of their occurrence; corrosion of fittings (four external indicators), overloading of structures (five external indicators), and mechanical impact (two external indicators) are considered as causes of degradation [22]. The Portland Cement Association report “Concrete Information: Types and Causes of Concrete Deterioration” is also worth noting. The authors take turns considering ten causes of deterioration (corrosion of embedded metals, freeze–thaw, chemical attack, alkali–aggregate reactivity, abrasion/erosion, fire/heat, restraint to volume changes, overload and impact, loss of support, and surface defects) and describe the external manifestations of each of them [23]. As mentioned earlier, the authors defined ten causes of distress and deterioration of RCs, and seven symptoms are associated with them, namely, construction faults, cracking disintegration, distortion/movement, erosion, joint failures, seepage, and spalling [13]. Jain and Bhattacharjee developed their own classification of causes and defects: corrosion (three external indicators), alkali aggregate reactions (two external indicators), freeze–thaw attack (five external indicators), sulfate attack (two external indicators), acid attack (two external indicators), fatigue (one external indicator) [10]. Barkavi and Natarajan used the opposite approach to the previous study, that is, they developed a classification of "defects-causes". As external manifestations, the authors considered various types of cracks (D-cracking, random cracks, diagonal cracks, crazing, longitudinal cracks, transverse cracks, cracks at joint) and conducted a correlation with the causes [24]. Masi, Digrisolo, and Santarsiero [25] analyzed the inter- and intra-variability of the concrete strength of existing buildings using a very large database comprised of approximately 1600 core tests extracted from RC buildings.

It is important to note that the external indications of damage may not always provide a complete picture of causes, and that further investigation may be required to fully assess the condition of RC structures. In some cases, nondestructive testing techniques such as ultrasonic testing or a ground-penetrating radar may be required to fully assess the condition of RC structures and identify the causes of damage. It should be noted that the analysis of the literature in the direction of establishing correlation between the causes and external indicators of defects showed that there is a need for a systematization of knowledge.

2.3. Using ML Methods for the Detection of Defects in Reinforced Concrete Structures

Scientists have used various ML methods to detect defects in RC structures [26,27]. For example, regression analysis involves analyzing the correlation between the external manifestations and the defects of the structure using statistical techniques [28]. The Bayesian network is a graphical model that represents the probabilistic relationships between variables. It can be used to identify the most likely causes of the defects based on the observed external manifestations [29]. Neural networks are a type of ML algorithm that can be used to analyze large datasets and identify patterns that are difficult for humans to discern. They can be trained to identify the most likely causes of the defects based on the observed external manifestations [30,31]. Fuzzy logic is a mathematical framework for dealing with uncertainty and imprecision. It can be used to model the causal relationships between the external manifestations and the defects of the structure, even when the relationships are not well-defined [5,8].

These methods can be used individually or in combination to establish causal relationships between external manifestations and defects of RC structures. However, it is important to note that no single method can provide a definitive answer. Experts' judgment and domain knowledge are still required to interpret the results.

The tools of ML can be used for the detection of defects in RC structures by analyzing data from various sources such as images, videos, sound waves, and sensors. One approach is to use convolutional neural networks (CNNs) to automatically identify and classify defects such as cracks, spalls, and corrosion [32]. The CNN can be trained on a large dataset of labeled images of concrete structures to learn the features that distinguish between normal and defective areas. The computer vision techniques can be used to analyze video data of the structure, detecting movements or changes in the structure that could indicate a defect. Sensors can be used to collect data on the structural health of the concrete and ML algorithms can be used to detect anomalies in the data that may indicate the presence of defects.

Overall, ML can be used in various ways to detect defects in concrete structures, and the choice of specific approach will depend on factors such as the type of defect, available data, and the desired level of accuracy. These approaches can be combined with various ML algorithms, such as support vector machines, random forest, decision trees, and neural networks to create models that can accurately detect and classify different types of defects in RC structures.

3. Methods

3.1. The Process of Machine Learning Modeling

Machine learning can be used to establish causal relationships between external manifestations and defects of RC structures by analyzing large datasets and identifying patterns and relationships between the variables [33]. This process includes several steps described below. Stages (a–c) are described in detail in Section 3.3, stage (d) in Section 3.2, and stages (e–g) in Section 4.

Data collection. The first step is to collect data on the external manifestations and defects of RC structures and causes of defects.

Data preprocessing. The data is then preprocessed to remove noise, missing values, and outliers, and to normalize the data.

Feature selection. The next step is to select the relevant features that are most informative for establishing the causal relationships.

Model selection. A ML models are then selected based on the problem type, the data, and the available resources.

Model training. The selected models are trained on the data using supervised learning techniques.

Model evaluation. The trained models are then evaluated on a separate validation dataset to assess their performance and identify any potential issues such as overfitting or underfitting.

Model deployment. Once the model has been evaluated and validated, it can be deployed to make predictions on new data and establish causal relationships between external manifestations and defects of RC structures.

3.2. The Used Methods of Machine Learning

Support Vector Machine (SVM) is a supervised learning algorithm that is based on the concept of finding the hyperplane that best divides a dataset into different classes [34]. This type of ML algorithm is commonly used for classification and regression analysis.

The main objective of an SVM is to find a hyperplane that maximizes the margin between the different classes in the dataset. The margin is defined as the distance between the hyperplane and the closest data points from each class. The hyperplane that maximizes the margin is chosen as the best classifier for the dataset.

An SVM can handle both linearly separable and nonlinearly separable datasets by using kernel functions to transform the data into a higher-dimensional space where a linear hyperplane can be used to separate the data. SVMs have been used in various fields, including computer vision, text classification, bioinformatics, and finance, to name a few.

Since the regression problem was solved, the algorithm used was ML algorithm Support Vector Regression (SVR) [35]. As with the Support Vector Machine (SVM), for classification, SVR is based on the idea of finding the hyperplane that best separates the data points; however, in the case of regression, the goal is to predict a continuous output variable rather than a categorical one. SVR works by finding the hyperplane that minimizes error between the predicted output and the actual output of the training data, while also maximizing the margin, or distance, between the hyperplane and the data points. SVR uses a set of support vectors, which are the data points that are closest to the hyperplane, to construct the regression model.

Decision trees (DTs) are a type of supervised ML algorithm that is used for both classification and regression problems [36]. The method is based on a tree-like model that represents a set of decisions and their possible consequences. The tree consists of nodes that represent decisions, branches that represent possible outcomes, and leaves that represent the final outcomes. The algorithm works by recursively splitting the dataset into smaller subsets based on the features that are most informative for making decisions. At each node of the tree, the algorithm selects the feature that best separates the data based on a criterion such as information gain or Gini impurity. It then creates two or more child nodes representing possible outcomes of the decision based on the feature value.

The algorithm continues to split the data until it reaches a stopping criterion, such as a minimum number of samples per node or a maximum depth of the tree. The result is a tree that can be used to make predictions for new data by traversing the tree from the root node to a leaf node. Random forest is a supervised ML algorithm that is an extension of decision trees. It is used for both classification and regression tasks and is known for its high accuracy and robustness.

Random forest (RF) works by constructing a multitude of decision trees, where each tree is trained on a random subset of the training data and a random subset of the input features [37]. The trees are trained independently of each other and generate their predictions for new data. The final result is obtained by aggregating the predictions of all trees, such as by taking the majority vote for classification tasks or the mean value for regression tasks. This approach reduces the risk of overfitting and improves the generalization of the model, as the variance in the trees' predictions is reduced.

Artificial Neural Networks (ANNs) are a type of ML algorithm that is designed to mimic the structure and function of biological neural networks in the brain [38,39]. ANNs consist of three key components: input layer, hidden layers, and output layer. The input layer receives the data to be processed, which are then passed through the hidden layers, where the input is transformed using the connections between the nodes. Finally, the transformed data are passed through the output layer, which produces the desired output. Each node in the network performs a simple mathematical operation that involves taking the weighted sum of the inputs it receives and applying a nonlinear activation function to produce the node's output. The activation function introduces nonlinearity to the network, enabling it to model complex relationships between the inputs and outputs.

During training, the network learns to adjust the weights of the connections between nodes to minimize the difference between the predicted output and the true output for a given input. This process is typically performed using an optimization algorithm, such as gradient descent, that adjusts the weights in the direction that reduces the error. In summary, ANNs are powerful ML models that can learn to make predictions or perform other tasks by adjusting the weights of interconnected nodes in response to training data. They are widely used in various applications and have shown impressive performance on many tasks, such as predictive modeling, natural language processing, speech recognition, and image recognition. They have demonstrated remarkable success in various fields and have contributed to significant advances in artificial intelligence and ML.

The following metrics were used to evaluate models (the mean square error (MSE), the mean absolute error (MAE)) and compare models with each other (the mean absolute percentage error (MAPE), the coefficient of determination (R^2)). A comparative analysis of the performance of ML models was conducted using MAPE and R^2.

The MAPE indicator is determined using [40]:

$$\mathrm{MAPE} = \frac{100\%}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \left| \frac{\mathrm{y_i} - \mathrm{y_i'}}{\mathrm{y_i}} \right|. \tag{1}$$

where y_i'—the output calculated by the model, y_i—the target output.

R^2—a statistical measure used to predict future outcomes or test hypotheses based on other related information [41]:

$$\mathrm{R}^2 = 1 - \frac{\sum\left(\mathrm{y_i} - \mathrm{y_i'}\right)^2}{\sum\left(\mathrm{y_i} - \overline{\mathrm{y_i}}\right)^2}. \tag{2}$$

where $\overline{y_i}$—the mean of the target output data.

The mean square error (MSE) was used as a loss function [42]:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i' - y_i\right)^2. \tag{3}$$

The performance of the ANN was evaluated using the mean absolute error (MAE) quality metric [43]:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} y_i' - y_i. \tag{4}$$

3.3. The Creation of Datasets for Identification of Cause-and-Effect Relationships between External Indicators of Damage and the Causes of Damage to Reinforced Concrete Structures

Based on a review and analysis of the literature that contented cause-and-effect relationships between external indicators of damage and the causes for their appearance, the dependency database was created. Six causes of defects were chosen: corrosion of reinforcement (1), structural overloads (2), sulfate attack (3), alkali–aggregate reaction (4), freezing and thawing (5), and chloride attack (6), which are the output variables in the database. External indicators were matched to each of the causes of defects. Therefore, for example, the corrosion of reinforcement has 8 external indicators; structural overloads—8 external indicators; sulfate attack—6 external indicators, of which 3 indicators are the same as in the previous reasons; alkali–aggregate reaction—4 indicators, of which 3 are the same as in the previous reasons; freezing and thawing—6 indicators, of which 5 indicators are the same as in the previous reasons, chloride attack—6 indicators, of which 5 are the same as in the previous reasons.

The next step was to create a full dataset for training and testing the ML model. The output values (causes of defects) were assigned numbers from 1 to 6. The input variables are numerical values of the defects' indicator, obtained by "stepping" to increase their value. For example, the defect value x6 "Cracks at the joint, edges and opening" can be from 0.1 to 0.3 mm. A stepping increase in x6 value from 0.1 to 0.3 with a step of 0.01 mm allows the formation of 20 additional records in the dataset. When the defects were assessed by linguistic variables, for example, x7 "Stain", the verbal value was replaced by a numerical binary assessment (where 0—no stain; 1—stain is present). Accordingly, all linguistic and categorical input variables were transformed into numerical variables by encoding them into vectors of real numbers, which were input as variables for ML models.

Supplementary Materials shows a part of the set of input variables (x1 ... x23) and the output value (y) causes of defects. The dataset created in this way contained 409 records and can be described by the following relationship:

$$if\, x_1 = \alpha \wedge x_2 = \beta \wedge \ldots x_n = \Omega \; then \; y = [1;6] \tag{5}$$

where y—total numerical assessment of causes of defects, which can have a value in the range from 1 to 6, $x_1 \ldots x_n$—the type of the external defects indicator, $\alpha \ldots \Omega$—the numerical value of the external defects indicator.

The creation of a high-quality dataset is a necessary and mandatory condition for obtaining the high productivity of predictive ML models. For this, the dataset must be prepared and preprocessed. After the collection stage, the data were normalized. Normalization is a data processing procedure where values are reduced to a certain specified range. Arithmetic mean and standard deviation were calculated for input data to normalize the data. Then, the arithmetic mean was subtracted from the input data, and the result was divided by the standard deviation. The dataset was divided into three datasets: % 60/20/20—training/validation/testing set.

4. Results

According to the information in the Methodology section, were created and trained four ML models: SVR, DT, RF, and ANN. Table 1 shows the effectiveness of the models based on the comparison of MAPE and R^2 metrics, and Figure 2 shows the scatter of the predicted values relative to the target values.

Table 1. Comparison of the ML models' effectiveness.

Models / Metrics	SVR	DT	RF	ANN
MAPE, %	21.34	41.5	31.67	20.21
R^2	0.5694	0.3485	0.4872	0.6416

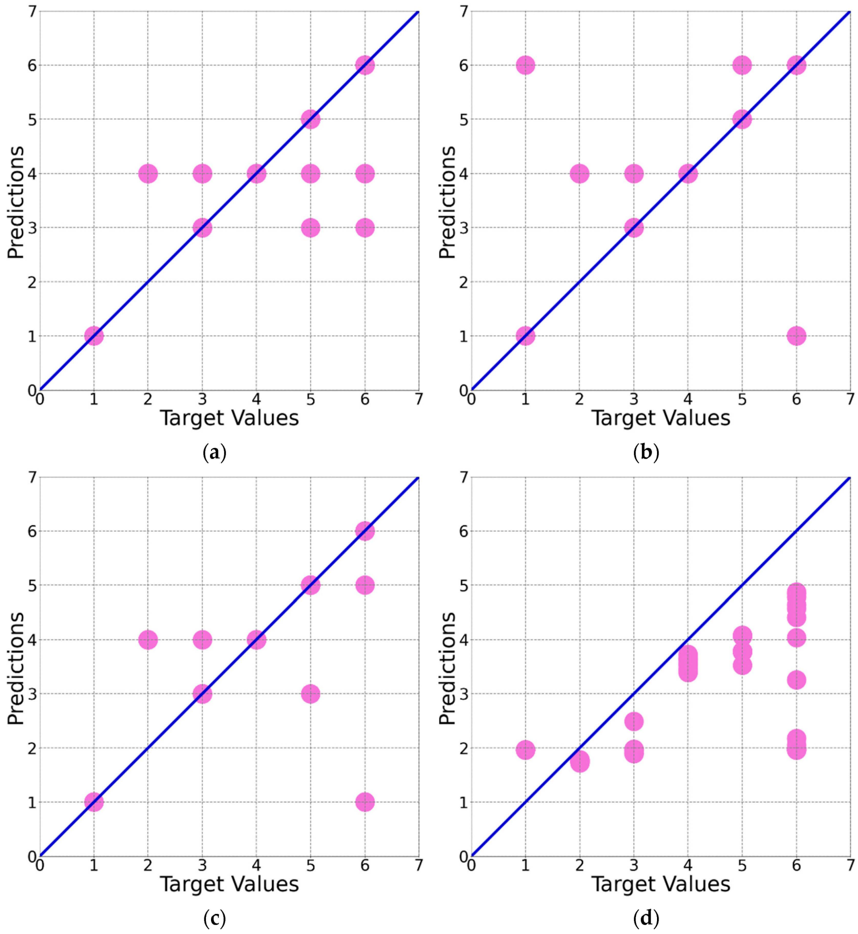


Figure 2. Scatter of predicted and target data values (**a**)—SVR model, (**b**)—DT model, (**c**)—RF model, (**d**)—ANN model.

The comparison showed a rather low performance of all models, while the ANN model had a slight advantage. In this regard, a decision was made to further optimize ANN models to obtain a higher level of productivity. Six ANN models were created, which had the same architecture but differed in activation functions and optimization algorithms. The network structure with an input layer that consisted of 23 neurons, one hidden layer with 64 neurons and an output layer with 1 neuron was established as optimal. Neural network models were checked with the activation functions: sigmoid, softmax, and ReLU.

The optimization algorithms SGD [44] and ADAM [45–47] were used.

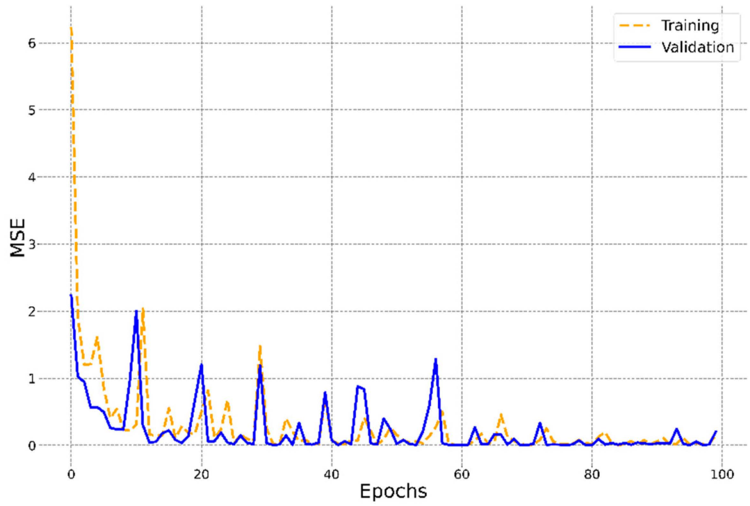
Table 2 shows the characteristics of the ANN models and a comparison of their performance. The performance of models testing was carried out for 100 epochs.

Table 2. Comparison of ANN model's performance.

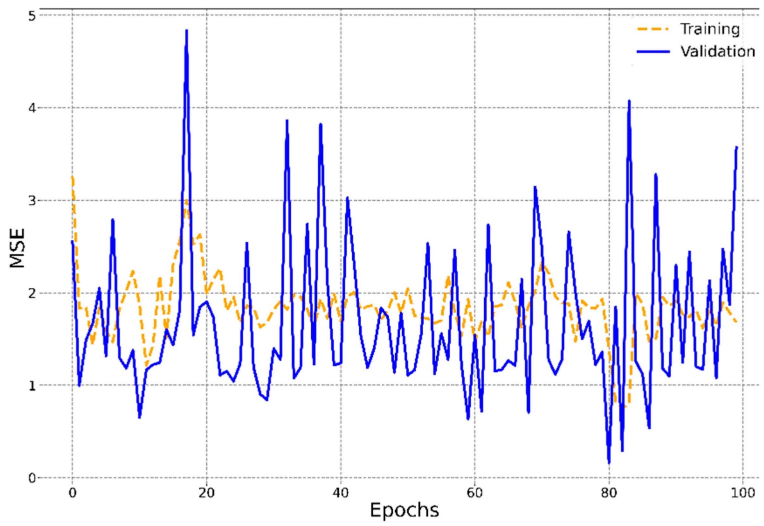
Models	ANN1	ANN2	ANN3	ANN4	ANN5	ANN6
activator	Sigmoid	ReLU	softmax	Sigmoid	ReLU	softmax
optimizer	SGD	SGD	SGD	ADAM	ADAM	ADAM
R^2	0.641	0.408	0.394	0.969	0.921	0.956
MAPE, %	20.12	19.14	19.86	3.38	8.09	3.97

It should be noted that the use of the ADAM optimization algorithm compared to the SGD algorithm showed a significant increase in productivity. The analysis of MAPE and R^2 metrics showed that the ANN4 model achieved the best performance results (lowest MAPE and highest R^2). This network used the activator sigmoid for neurons' hidden layers and the ADAM loss function optimizer. The best performance was achieved for 100 epochs: MAPE = 3.38%; R^2 = 0.969.

Figures 3 and 4 show a comparison of the MSE and MAE results for the training and validation datasets for the best (ANN4) and worst (ANN1) ANN models.

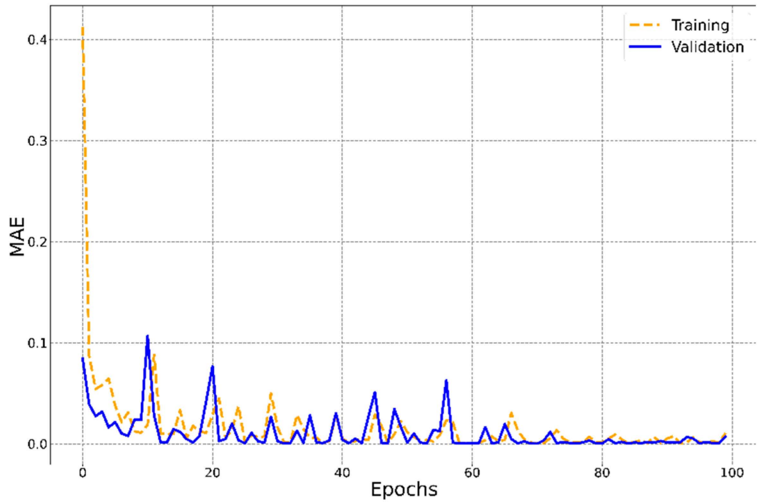


(**a**)

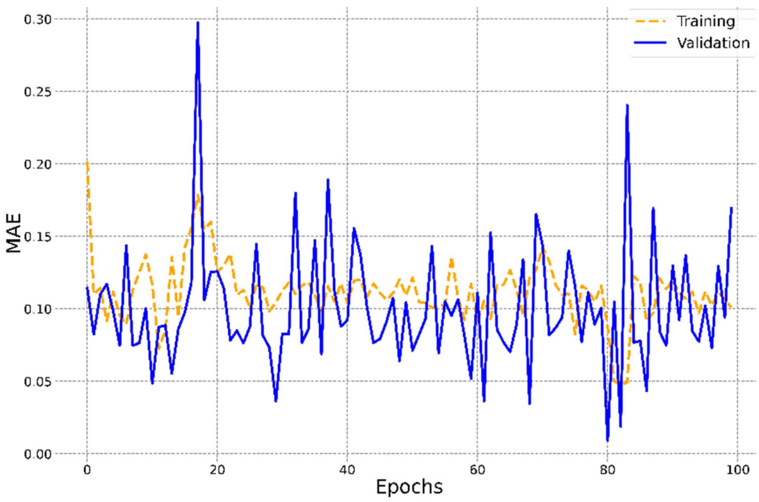


(**b**)

Figure 3. The results of MSE for training and validation datasets of ANN4 and ANN1: (**a**) MSE results ANN4; (**b**) MSE results ANN1.



(**a**)



(**b**)

Figure 4. The results of MAE for training and validation datasets of ANN4 and ANN1, (**a**)—MAE results ANN4, (**b**)—MAE results ANN1.

Figure 3a shows the comparison results of MSE for the training and validation datasets of model ANN4. The maximum MSE of training set was in the first epoch—6.238; validation set in the first epoch—2.230. The minimum MSE of training set was in the 96th epoch—0.015 and validation set in the 96th epoch—0.003. Figure 3b shows the comparison results of the MSE for the training and validation datasets of model ANN1. The maximum MSE of the training set was in the first epoch—3.384; validation set in the 42nd epoch—3.520. The minimum MSE of training set was in the 97th epoch—0.844 and validation set in the 96th epoch—0.184.

Figure 4a shows the comparison the results of MAE for the training and validation datasets of model ANN4. The maximum MAE of training set was in the 1st epoch—0.412; validation set in the 1st epoch—0.084. The minimum MAE of training set was in the 85th epoch—0.0003 and validation set in the 96th epoch—0.0003. Figure 4b shows the comparison the results of MAE for the training and validation datasets of model ANN1. The maximum MAE of training set was in the 1st epoch—0.201; validation set in the 17th epoch—0.293. The minimum MAE of training set was in the 82nd epoch—0.048 and validation set in the 80th epoch—0.185.

The model architecture, weights, and configuration after training and testing were saved in H5 and JSON files. This allows for further use of the model without a new training.

In addition, the model can obtain a forecast on datasets that it has not yet "seen". The next stage involved checking the effectiveness of the model and comparing its forecast with real data. For this purpose, real data of visual inspection of RC with defects were used. The dataset contained 12 records, two records for each cause of defects. The data on the detected defects for each of the causes were loaded into the ANN model and the values of the predicted causes (y') were calculated. Table 3 shows the obtained results of the target value (y), prediction (y'), and deviation in % between these values.

Table 3. Verification of the ANN model.

Data Set No.	Target Value, y	Prediction, y′	Deviation (Δy), %
1	1	0.998	0.169
2	1	0.998	0.184
3	2	1.988	0.622
4	2	1.999	0.073
5	3	3.003	0.114
6	3	2.870	4.349
7	4	4.041	1.019
8	4	4.191	4.783
9	5	3.534	29.316
10	5	5.880	17.592
11	6	5.933	1.109
12	6	5.933	1.109

The level of deviations between the target and forecast values was estimated as low, which confirms the quality of the forecast except for causes "5". At the "Target value" = 5, the deviations between the forecast and the target for both datasets were quite significant (29.316% and 17.592%). Most likely, such a deviation is due to outliers and anomalies in the input data at this value.

5. Discussion

This study has several limitations. The model predicts only six causes of damage to RC structures. The following reasons, which can be identified by at least three external manifestations, were chosen. With a smaller number of input variables, the effectiveness of the model will decrease, since certain sets of external manifestations may indicate different causes of damage. The model does not consider external indicators that are the consequences of obvious causes, such as fire, explosion, and mechanical damage due to an impact. All ML models, including ANNs, were used to solve regression problems. This approach, unlike classification, allowed us to obtain a forecast in the form of continuous values. In our opinion, this can be important because, for example, the obtained value of three and a half can indicate the simultaneous influence of reasons 3 and 4.

The possibility of using the proposed ANN model to predict the causes of RC defects was analyzed and compared in terms of the quality of its work with models from similar studies. It should be noted that the ML method, which was most often used to solve similar problems, was based on fuzzy logic. Terenchuk et al. [27] used a fuzzy knowledge base and one of the modifications of the Takagi-Sugeno-Kanga fuzzy network to model the evaluation system. The main criteria for choosing this modification were its ability to solve the classification problem under conditions of uncertainty and the ability to set rules as a function of inputs. To identify the initial values, the membership measures obtained using the clustering method were used. Pasko and Terenchuk [26] created a model of a system based on knowledge and used neural fuzzy networks to solve the problem of supporting decision-making regarding the assessment of the impact of repair and construction works on the technical condition of the object. As a result, a conceptual model of an expert system with a Cascade ARTMAP-integrated artificial neural network was proposed. Kim et al. [42] presented a computer-assisted crack diagnosis system based on fuzzy set theory for RC structures, which aids a non-expert in diagnosing the cause of cracks. The inputs to the system are linguistic variables concerning the crack symptoms and some numeric data about concrete. The system used input data and was based on built-in rules; moreover, the proposed system executed fuzzy inference to evaluate the crack causes. The authors claimed that the proposed system provided results similar to those obtained by experts. Jain and Bhattacharjee [18] applied fuzzy concepts for a quantitative assessment of the condition of the elements of RC structures. Input data on various manifestations of damage were obtained based on visual inspection. The rule base was created based on expert evaluation. Predictive data and condition assessment obtained from using the model can be used for maintenance and repair.

In this study, a multilayer perceptron (MLP) ANN type was used. This type of ML tool is highly flexible, widely used, and can be applied to study the relationships between input and output data [43]. The development of the MLP structure was carried out using the trial and error method [39]. It was found that a model with one hidden layer of neurons is more effective than one with two and three. It was also determined that the best results (minimum value of MAPE and maximum value of R^2) are generated by models with 64 neurons in the hidden layer.

6. Conclusions

During the lifecycle of RC structures, it is important to consider potential causes of damage and take appropriate measures to prevent or mitigate their impact. Understanding cause-and-effect relationships is essential in engineering sciences for identifying problems, designing solutions, and improving the performance and reliability of structures. The object of the study was the analysis and identification of cause-and-effect relationships between the causes of damage to RC elements and their external indicators. The literature analysis showed that the establishment of this type of cause-and-effect relationship is possible using the methods and tools of ML. The aim of this study was to determine the possible causes of the defects of the RC elements based on the identification of external indicators using ML tools. This study created and compared the performance of four ML models, namely, Support Vector Regression (SVR), decision trees (DTs), random forest (RF), and Artificial Neural Networks (ANNs). The first comparison showed a rather low performance of all models, with a slight advantage of the ANN model. Later, six ANN models were optimized to obtain a higher level of performance. The next step of this study was the training, validation, and testing of ANN models. The analysis of MAPE and R^2 metrics showed that the ANN model with ADAM loss function and sigmoid activation had the best results (MAPE 3.38%; R^2 0.969). The novelty of the study is the development of the ML model, which is based on the use of ANNs, and allows for the establishment of cause-and-effect relationships in the diagnosis of the technical condition of the RC elements. The model can be used further without retraining, and it can make predictions on datasets it has not yet "seen". The practical use of such a model allows for the diagnosis of some causes of defects during a visual inspection of structures. However, the use of this model should not be without the visual inspections of skilled engineers as well as of permanent monitoring, especially for highly important structures. Further research can be aimed at increasing the database, increasing the level of model performance, and expanding the model by adding new causes of damage to RC structures.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/su15065250/s1, Table S1: Input variables that were used for the ANN models.

Author Contributions: Conceptualization, R.T., Y.T., G.R., A.S. and V.T.; methodology, Y.T. and R.T.; software, R.T. and V.T.; validation, Y.T., G.R. and A.S.; formal analysis, Y.T., R.T., G.R., A.S. and V.T.; investigation, Y.T. and R.T.; resources, Y.T. and R.T.; data curation, G.R., A.S., V.T. and Y.T.; writing—original draft preparation, A.S., G.R., V.T. and R.T.; writing—review and editing, G.R., Y.T. and A.S.; visualization, V.T. and R.T.; supervision, R.T.; funding acquisition R.T. and Y.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated and analyzed during the current study are available from the authors upon reasonable request.

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

References

1. Kosmatka, S.H.; Panarese, W.C.; Kerkhoff, B. *Design and Control of Concrete Mixtures*; Portland Cement Association: Skokie, IL, USA, 2002; Volume 5420.
2. Veloutsou, C.; Christodoulides, G.; Guzmán, F. Charting Research on International Luxury Marketing: Where Are We Now and Where Should We Go Next? *IMR* **2022**, *39*, 371–394. [CrossRef]
3. Moshynskyi, V.; Striletskyi, P.; Trach, R. Application of the Building Information Modelling (BIM) for Bridge Structures. *Acta Sci. Pol.-Archit. Bud.* **2022**, *20*, 3–9. [CrossRef]
4. Bodnar, L.; Koval, P.; Stepanov, S.; Panibratets, L. Operational state of bridges of Ukraine. *Avtošljachovyk Ukraïny* **2019**, *2*, 57–68. [CrossRef]
5. Nimawat, D.; Gidwani, B.D. Identification of Cause and Effect Relationships among Barriers of Industry 4.0 Using Decision-Making Trial and Evaluation Laboratory Method. *BIJ* **2021**, *28*, 2407–2431. [CrossRef]
6. Dzięcioł, J.; Sas, W. Perspective on the Application of Machine Learning Algorithms for Flow Parameter Estimation in Recycled Concrete Aggregate. *Materials* **2023**, *16*, 1500. [CrossRef]
7. Terenchuk, S.; Pashko, A.; Yeremenko, B.; Kartavykh, S.; Ershova, N. Modeling an Intelligent System for the Estimation of Technical State of Construction Structures. *EEJET* **2018**, *3*, 47–53. [CrossRef]
8. Pasko, R.; Terenchuk, S. The Use of Neuro-Fuzzy Models in Expert Support Systems for Forensic Building-Technical Expertise. *ScienceRise* **2020**, *2*, 10–18. [CrossRef]
9. Kim, Y.M.; Kim, C.K.; Hong, G.H. Fuzzy Set Based Crack Diagnosis System for Reinforced Concrete Structures. *Comput. Struct.* **2007**, *85*, 1828–1844. [CrossRef]
10. Jain, K.K.; Bhattacharjee, B. Application of Fuzzy Concepts to the Visual Assessment of Deteriorating Reinforced Concrete Structures. *J. Constr. Eng. Manag.* **2012**, *138*, 399–408. [CrossRef]
11. Kanaan, D.; Soliman, A.M.; Safhi, A.E.M. External Sulfate Attack of Ambient-Cured One-Part Alkali-Activated Self-Consolidating Concrete. *Sustainability* **2023**, *15*, 4127. [CrossRef]
12. Shen, Y.; Yu, Z.; Li, C.; Zhao, C.; Sun, Z. Automated Detection for Concrete Surface Cracks Based on Deeplabv3+ BDF. *Buildings* **2023**, *13*, 118. [CrossRef]
13. Woodson, R.D. *Concrete Structures: Protection, Repair and Rehabilitation*; Butterworth-Heinemann: Amsterdam, The Netherlands; Boston, MA, USA, 2009; ISBN 978-1-85617-549-4.
14. Yu, C.W.; Bull, J.W. (Eds.) *Durability of Materials and Structures in Building and Civil Engineering*; Whittles Publishing: Caithness, UK, 2006; ISBN 978-1-870325-58-5.
15. Torgal, F.P. (Ed.) *Eco-Efficient Concrete*; Woodhead Publishing in Materials; Woodhead Publishing: Cambridge, UK; Philadelphia, PA, USA, 2013; ISBN 978-0-85709-424-7.
16. Delatte, N. (Ed.) *Failure, Distress and Repair of Concrete Structures*; Woodhead Publishing in Materials; CRC Press: Boca Raton, FL, USA, 2009; ISBN 978-1-84569-703-7.
17. Douglas, J.; Ransom, W.H. *Understanding Building Failures*, 3rd ed.; Taylor & Francis: London, UK; New York, NY, USA, 2007; ISBN 978-0-415-37082-0.
18. Transportation Research Board Control of Cracking in Concrete, Transportation Research Circular 2006. Available online: https://www.trb.org/Publications/Blurbs/158019.aspx (accessed on 1 February 2023).
19. ACI Committee 224 Causes, Evaluation, and Repair of Cracks in Concrete Structures, American Concrete Institute., Farmington Hills, MI, USA, 2007. Available online: https://www.concrete.org/store/productdetail.aspx?ItemID=224107&Format=DOWNLOAD&Language=English&Units=US_AND_METRIC (accessed on 1 February 2023).
20. Trach, R.; Moshynskyi, V.; Chernyshev, D.; Borysyuk, O.; Trach, Y.; Striletskyi, P.; Tyvoniuk, V. Modeling the Quantitative Assessment of the Condition of Bridge Components Made of Reinforced Concrete Using ANN. *Sustainability* **2022**, *14*, 5779. [CrossRef]
21. ACI Committee 201 Guide for Making a Condition Survey of Concrete in Service, American Concrete Institute., Farmington Hills, MI, USA, 1997. Available online: http://dl.mycivil.ir/dozanani/ACI/ACI%20201.1R-92%20Guide%20for%20Making%20a%20Condition%20Survey%20of%20Concrete%20in%20Service_MyCivil.ir.pdf (accessed on 1 February 2023).
22. *DSTU-N B V. 1.2-18:2016*; Guidelines Regarding the Inspection of Building Objects to Determine and Assess Their Technical Condition; Ministry of Regional Development and Construction of Ukraine: Kyiv, Ukraine, 2016.
23. PCA. Concrete Information: Types and Causes of Concrete Deterioration., Portland Cement Association, Washington, DC, USA, 2002. Available online: https://openlibrary.org/books/OL17614660M/Types_and_causes_of_concrete_deterioration (accessed on 1 February 2023).

24. Barkavi, T.; Natarajan, C. A Decision Support System for the Condition Assessment of Distressed Concrete Buildings. In Proceedings of the AEI 2017, Oklahoma City, OK, USA, 11–13 April 2017; American Society of Civil Engineers: Oklahoma City, OK, USA, 2017; pp. 880–891.
25. Masi, A.; Digrisolo, A.; Santarsiero, G. Analysis of a Large Database of Concrete Core Tests with Emphasis on Within-Structure Variability. *Materials* **2019**, *12*, 1985. [CrossRef]
26. Feng, M.Q.; Kim, D.K.; Yi, J.-H.; Chen, Y. Baseline Models for Bridge Performance Monitoring. *J. Eng. Mech.* **2004**, *130*, 562–569. [CrossRef]
27. Khoa, N.L.; Zhang, B.; Wang, Y.; Chen, F.; Mustapha, S. Robust Dimensionality Reduction and Damage Detection Approaches in Structural Health Monitoring. *Struct. Health Monit.* **2014**, *13*, 406–417. [CrossRef]
28. Ranković, V.; Grujović, N.; Divac, D.; Milivojević, N. Development of Support Vector Regression Identification Model for Prediction of Dam Structural Behaviour. *Struct. Saf.* **2014**, *48*, 33–39. [CrossRef]
29. Phan, V.-T.; Tran, V.-L.; Nguyen, V.-Q.; Nguyen, D.-D. Machine Learning Models for Predicting Shear Strength and Identifying Failure Modes of Rectangular RC Columns. *Buildings* **2022**, *12*, 1493. [CrossRef]
30. Moradi, N.; Tavana, M.H.; Habibi, M.R.; Amiri, M.; Moradi, M.J.; Farhangi, V. Predicting the Compressive Strength of Concrete Containing Binary Supplementary Cementitious Material Using Machine Learning Approach. *Materials* **2022**, *15*, 5336. [CrossRef]
31. Soyoz, S.; Feng, M.Q. Long-Term Monitoring and Identification of Bridge Structural Parameters. *Comput.-Aided Civ. Infrastruct. Eng.* **2009**, *24*, 82–92. [CrossRef]
32. Wu, C.-S.; Zhang, J.-Q.; Qi, L.-L.; Zhuo, D.-B. Defect Identification of Concrete Piles Based on Numerical Simulation and Convolutional Neural Network. *Buildings* **2022**, *12*, 664. [CrossRef]
33. Taffese, W.Z.; Sistonen, E. Machine Learning for Durability and Service-Life Assessment of Reinforced Concrete Structures: Recent Advances and Future Directions. *Autom. Constr.* **2017**, *77*, 1–14. [CrossRef]
34. He, Y.; Huang, Z.; Liu, D.; Zhang, L.; Liu, Y. A Novel Structural Damage Identification Method Using a Hybrid Deep Learning Framework. *Buildings* **2022**, *12*, 2130. [CrossRef]
35. Chen, H.; Li, X.; Wu, Y.; Zuo, L.; Lu, M.; Zhou, Y. Compressive Strength Prediction of High-Strength Concrete Using Long Short-Term Memory and Machine Learning Algorithms. *Buildings* **2022**, *12*, 302. [CrossRef]
36. Chotkan, S.; van der Meij, R.; Klerk, W.J.; Vardon, P.J.; Aguilar-López, J.P. A Data-Driven Method for Identifying Drought-Induced Crack-Prone Levees Based on Decision Trees. *Sustainability* **2022**, *14*, 6820. [CrossRef]
37. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
38. Onyelowe, K.C.; Ebid, A.M.; Riofrio, A.; Baykara, H.; Soleymani, A.; Mahdi, H.A.; Jahangir, H.; Ibe, K. Multi-Objective Prediction of the Mechanical Properties and Environmental Impact Appraisals of Self-Healing Concrete for Sustainable Structures. *Sustainability* **2022**, *14*, 9573. [CrossRef]
39. Sas, W.; Dzięcioł, J.; Głuchowski, A. Estimation of Recycled Concrete Aggregate's Water Permeability Coefficient as Earth Construction Material with the Application of an Analytical Method. *Materials* **2019**, *12*, 2920. [CrossRef]
40. Trach, Y.; Trach, R.; Kalenik, M.; Koda, E.; Podlasek, A. A Study of Dispersed, Thermally Activated Limestone from Ukraine for the Safe Liming of Water Using ANN Models. *Energies* **2021**, *14*, 8377. [CrossRef]
41. Trach, R.; Trach, Y.; Kiersnowska, A.; Markiewicz, A.; Lendo-Siwicka, M.; Rusakov, K. A Study of Assessment and Prediction of Water Quality Index Using Fuzzy Logic and ANN Models. *Sustainability* **2022**, *14*, 5656. [CrossRef]
42. Trach, Y.; Chernyshev, D.; Biedunkova, O.; Moshynskyi, V.; Trach, R.; Statnyk, I. Modeling of Water Quality in West Ukrainian Rivers Based on Fluctuating Asymmetry of the Fish Population. *Water* **2022**, *14*, 3511. [CrossRef]
43. Trach, R.; Trach, Y.; Lendo-Siwicka, M. Using ANN to Predict the Impact of Communication Factors on the Rework Cost in Construction Projects. *Energies* **2021**, *14*, 4376. [CrossRef]
44. Ruder, S. An Overview of Gradient Descent Optimization Algorithms. *arXiv* **2017**, arXiv:1609.04747.
45. Kowalski, J.; Połoński, M.; Lendo-Siwicka, M.; Trach, R.; Wrzesiński, G. Method of Assessing the Risk of Implementing Railway Investments in Terms of the Cost of Their Implementation. *Sustainability* **2021**, *13*, 3085. [CrossRef]
46. Tran-Ngoc, H.; Khatir, S.; De Roeck, G.; Bui-Tien, T.; Abdel Wahab, M. An Efficient Artificial Neural Network for Damage Detection in Bridges and Beam-like Structures by Improving Training Parameters Using Cuckoo Search Algorithm. *Eng. Struct.* **2019**, *199*, 109637. [CrossRef]
47. Li, R.; Zhang, X.; Dai, H.; Zhou, B.; Wang, Z. Interpretability Analysis of Heartbeat Classification Based on Heartbeat Activity's Global Sequence Features and BiLSTM-Attention Neural Network. *IEEE Access* **2019**, *7*, 109870–109883. [CrossRef]