Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)
Edited by Mário P. Brito, Terje Aven, Piero Baraldi, Marko Čepin and Enrico Zio

doi: 10.3850/978-981-18-8071-1_P065-cd

Explainable artificial intelligence for understanding the ageing classes of reinforced concrete bridge components

Francesca Marsili

Chair of Material and Building Preservation, Helmut-Schmidt University / University of the Federal Armed Forces Hamburg, Germany. E-mail: francesca.marsili@hsu-hh.de

Filippo Landi

Department of Civil and Industrial Engineering, University of Pisa, Italy. E-mail: filippo.landi@ing.unipi.it

Sylvia Keßler

Chair of Material and Building Preservation, Helmut-Schmidt University / University of the Federal Armed Forces Hamburg, Germany. E-mail: sylvia.kessler@hsu-hh.de

This article proposes an approach to the identification and interpretation of homogeneous ageing classes for reinforced concrete bridge components. The approach is articulated into three phases: in the first phase, homogeneous ageing classes are identified by considering the results of the visual inspections and the time sequence of condition states of the bridge components, applying a cluster analysis based on the k-means algorithm; in the second phase, the ageing class is predicted by means of a random forest algorithm, considering features of the bridge and of the components; in the third stage, the prediction is explained by applying a SHAP analysis. The results reveal that the prediction of the ageing class is influenced by the year of construction of the bridge and therefore of the component. This result opens up to a multiplicity of interpretations, which are considered in the article. The dependence of the ageing class on other variables is also discussed.

Keywords: remaining useful life prediction, cluster analysis, k-means, random forest, SHAP analysis, explainable artificial intelligence

1. Motivation of the research

In order to extend the useful life of existing structures such as reinforced concrete (r.c.) bridges, it is necessary to accurately estimate their remaining useful life and efficiently plan maintenance interventions. To this end, a valuable source of data are digital maintenance management systems (MMSs), where the results of visual inspections performed periodically on r.c. bridges are collected. However, analysis of such digital databases requires consideration of a number of issues that typically characterize these datasets.

First of all, it should be considered that building materials, construction techniques, environmental actions, and traffic loads acting on r.c. bridges vary over the decades. This makes it difficult to transfer past experience to more recently built structures. At the same time, results of visual inspections have been collected for some decades in digital databases, so while the (short) life of recent objects is well documented, information relating only to the last part of the life of older objects is available.

Secondly, more and more studies focus on predicting the remaining useful life while considering the evolution of the condition and the characteristics of the object, exploiting digital databases of MMSs and data-based artificial intelligence (AI) techniques. An example of such studies is Huang (2010). However, models that rely on data often have a black box character, and the results are often difficult to understand and interpret. This represents a limit to the use of these techniques by infrastructure managers, for whom it is important to comprehend the origin of the results.

The problem of the black box nature of data-driven AI models and the difficult explanation of their results are common to every field of their application. For this reason, in recent years, research has focused on developing approaches that support the application of these techniques, capable of making the results more understandable and facilitating their interpretation and explanation. Such approaches aimed at understanding and interpreting the results of machine learning models are referred to by the term Explainable AI (XAI) (Confalonieri et al., 2021). However, being a recent field of research, the implementation of XAI in infrastructure and maintenance management is still at the embryonic stage.

A more transparent approach that is already being pursued in the case of roads is to identify road segments having similar characteristics (called ”homogeneous groups,” HGs) and similar condition evolution. Cluster algorithms are also employed for this purpose (Mathavan et al., 2015). Asset managers identify the factors that determine the evolution of deterioration for each segment and develop a deterioration model for each HG.

Recent studies have shown that HGs can be identified for building materials with respect to their mechanical properties as well as bridge components with respect to their condition evolution, by analysing digital MMSs with cluster algorithms (Croce et al., 2018, 2020; Marsili et al., 2023): such approaches, while very promising, have not yet been duly explored. In addition, it is important to emphasized that although it is appropriate to identify groups of objects or components that degrade at similar rates, the ability to recognize such groups based on the characteristics of the bridge and its environment must be analyzed a posteriori and not assumed a priori.

The aim of this article is to develop a concept for the identification and understanding of homogeneous ageing classes of r.c. bridge components. The framework is developed considering r.c. bridges in Switzerland and the KUBA-DB database, which collects the results of periodical visual inspections on the objects.

The framework consists of three distinct phases, to which three different data analysis techniques correspond: 1) A phase based on unsupervised learning, in which, through a cluster analysis of the time sequences of the component condition, homogeneous aging classes are identified; 2) A phase based on supervised learning, in which a random forest algorithm is applied, the degradation class is predicted based on other characteristics of the bridge component, also related to the bridge itself and its location; 3) A phase in which a SHAP analysis is carried out, through which the results obtained by applying the random forest algorithm are explained and consequently the factors that influence the prediction of the class are better understood. The final aim of the article is to develop an approach to the assessment of the remaining useful life of existing infrastructure components that is understandable for the infrastructure managers themselves, and therefore easy to use.

The paper is articulated in the following parts: Section 2 describes qualitatively the methods at the basis of the proposed framework, Section 3 presents the case study at which the approach has been applied and Section 4 draws some conclusions.

2. Methods

2.1. *K-means clustering*

The k-means algorithm is an unsupervised learning approach finalized at clustering n observations into k groups in such a way that each observation belongs to the group with the nearest mean (Aggarwal and Reddy, 2014). The mean of the cluster is also referred to as cluster centroid and it represents a prototype of the cluster. The simplicity and efficiency of this algorithm have made it widespread and popular. However, certain factors may impact the performance of the algorithm, such as the initial choice of centroids, which is necessary to get the algorithm started, and the estimated number of clusters. The latter is not given a priori, which is why the algorithm is often supplemented by the calculation of a performance measure called the Silhoutte Width, based on which the optimal number of clusters can be identified. This requires repeating the analysis for an increasing number of clusters until the Silhoutte Width has reached its maximum value. The optimal number of clusters corresponds to the model with the highest Silhouette Width.

2.2. *Random Forest algorithm*

Random Forest (RF) is a supervised learning algorithm consisting of a number of independent decision trees (Breiman, 2001). A decision tree model partitions the given dataset into two groups based on a certain criterion until a given stopping condition is met. At the basis of the decision tree there are the so-called leaf nodes or leaves, which represent an output label.

A disadvantage of decision trees is overfitting, which will lead to poor generalization accuracy. One way to improve generalization accuracy is to build many individual trees by considering only a subset of the observations obtained by bootstrapping, and to average their predictions. The ensemble of the so-developed models is the RF. The RF model has greater predictive capacity than a single decision tree, however while the latter can be easily visualized, the former has a black box character. Nonetheless, it is possible to gain some insight on the complex model by calculating the importance of each variable. This is obtained by adding up the improvement in the objective function given in the splitting criterion over all internal nodes of a tree and across all trees in the forest, separately for each predictor variable.

RF models are characterized by a set of hyperparameters, which can be tuned to improve predictions. The most important parameters are the number of decision trees in the forest, the maximum number of features in each decision tree, the minimum number of samples required to be at a leaf node, the randomness of the bootstrapping of the samples used when building trees and the sampling of the features to consider when looking for the best split at each node.

2.3. *SHAP analysis*

SHapley Additive exPlanations (SHAP) values are an approach to explain machine learning black-box models borrowed from game theory Lundberg and Lee (2017). Shapley values are used to measure the contributions to a final outcome of each player of a coalition, while preserving the sum of contributions being equal to the final outcome. Given a data-based AI prediction model, SHAP values can be used to identify important predictors among the features characterizing the data set. The SHAP analysis consists in building a proxy model of the black box model through changes in the input and the assessment of the changes in the prediction. If the outcome significantly changes by changing the input value for a feature, the feature for that data point is an important predictor and it will have a high SHAP value.

The advantages of SHAP analysis are the following: 1) it provides a local interpretation, by assessing the contribution of each feature of the dataset to single prediction; 2) at the same time, it also gives a global interpretation, by showing whether the feature has a positive or negative impact on predictions; 3) it is model agnostic, which means that it can be used to explain a large variety of models including tree-based models and neural networks.

In conclusion, the SHAP values reveal interesting insights into how input variables influence the machine learning model's predictions, both locally, considering individual instances, and globally, considering the entire population, and are able to do so regardless of the model used.

3. Application

3.1. *The data*

The Swiss Federal Roads Office (ASTRA) is the operator of Switzerland's road infrastructure, which comprises a wide range of structures including r.c. bridges. A very important activity in the management of bridges and their maintenance is the visual inspection carried out at regular intervals, during which the condition of the structure and its components is assessed and which thus provides the basic information for planning maintenance work. A distinction is made between three different types of inspection: primary, intermediate and special. During primary and intermediate inspections, the focus is on identifying damage processes and the information gathered during the inspection is recorded in the KUBA-DB database. A damage group consists of a set of homogeneous damage processes within the same segment of the structural element, which similarly affect the functionality of the component. The severity with which the damage afflicts the functionality of the element is assessed by means of an index with values between 1 and 5. 1 means that no damage is present, while 5 means that the damage has reached a critical state such that the functionality of the structural element is compromised. In a subsequent step, the overall condition of the structural member is evaluated, considering the damage groups and their combined effect on the safety and functionality of the component. The judgments relating to each structural component are finally aggregated at the level of the object, so that an index expressing its overall condition can be associated with it. The database collects not only the results of visual inspections performed on a regular basis, but also inventory data relating to the objects, such as the year of construction of the bridge, the bridge type, the segment of road to which the object belongs, the altitude of the bridge site, the size of the components. Further information that is collected in the database concerns the execution of maintenance interventions, such as the type of intervention and cost. However, while the condition and inventory data is complete, the latter data is only partial.

In the KUBA-DB database, the bridge structure is divided into many components. If each component is considered individually, data on sojourn time in each condition class may be scarce. Consequently, those components that have a similar structural behavior have been grouped together, so that the data set to be analyzed is more consistent. A distinction is mainly made between the superstructure and substructure of the bridge: in the case of the superstructure, elements such as deck slab, plate-girder and cross-girder are grouped together; in the case of the substructure, it is possible to group components such as wing-walls and abutment.

3.2. *The procedure*

First of all, the k-means algorithm is applied to identify groups of components having a similar deterioration development, which can be referred to as ageing classes. A characteristic of the condition database is that it is unbalanced: the results of visual inspections show that in most cases the components of the bridges are in good condition, assigning a condition index (CI) of 1 and 2. Conversely, results of inspections which indicate an unsatisfactory condition of the component (for which the component is assigned a CI of 3 and 4) are much less frequent. The CI 5 indicating an alarming condition is very rarely assigned. Accordingly, a cluster analysis is performed using the k-mean algorithm, taking into account the age of the component to which it is assigned index 2 and index 3. Considering the components to which also index 4 has been assigned results in an extremely limited data set, so a cluster analysis can hardly be conducted. By performing the cluster analysis and calculating the Silhouette Width for different cluster models characterized by an increasing number of clusters, it can be concluded that models characterized by two and three clusters have the highest Silhoutte Width, whose values are around 0.5. However, it is slightly higher for the model characterized by two clusters. Since the difference in the performance parameter is negligible, and in view of the application of stochastic models to predict the time at which a critical condition is reached, it is still possible to choose the model characterized by three clusters, which can lead to a more accurate prediction of the remaining useful life. The three groups correspond to homogeneous aging classes and can be referred to as ”fragile,” ”normal,” and ”robust,” depending on the rate of degradation of the components that make up each group, which can be fast, normal, or slow.

Once the homogeneous ageing classes have been identified, the RF algorithm is applied to the data: the dependent variable is the ageing class of the component, while the independent variables are the characteristics of the component, such as the year of construction of the bridge, the bridge typology, the road section where the bridge is located, the altitude of the bridge location, and the size of the component. Categorical features such as bridge typology and road section have been encoded using the label encoding approach. This approach is preferred when performing a SHAP analysis, as the SHAP value of the categorical variable is easier to calculate. When fitting the RF model to the data, model hyper-parameters such as the number of trees in the RF, the maximum number of features considered to divide a node and the maximum number of levels in each decision tree are also tuned. The predictive ability of the RF is then expressed through a confusion matrix and a classification report.

In the next step, a global SHAP analysis is performed. SHAP values quantify the importance for a model of each feature in order to make correct predictions. These values can be used to confirm domain experience, or conversely to generate new hypotheses and to validate new theories. The purpose of the SHAP analysis is to provide a global interpretation of the machine learning model and understand the main drivers of predictions across the population. This is accomplished by aggregating the SHAP values for individual instances across the entire population. These values can be displayed using multiple plots, such as bar plot, beeswarm plot and dependance plot.

The bar plot allows to examine the mean absolute SHAP value for each feature across all data. This quantifies, on average, the entity (positive or negative) of the contribution of each feature to the expected aging class. Features having higher mean absolute SHAP values have a greater influence. Compared to more traditional feature importance measures, SHAP values have the advantage of being more rigorous from a theoretical point of view and, in some cases, they can be expressed in more intuitive units of measure.

Beeswarm plots reveal not only the relative importance of features, but also their actual relationships to the predicted outcome. Specifically, they show how the underlying values of each characteristic are related to the model's predictions. When the values of a feature have positive SHAP values with respect to a certain value of the dependent variable, it means that they contribute to its prediction. Even the distribution of the points can be informative, suggesting what is the magnitude of the impact on the prediction of different feature values.

Dependence plots are necessary to fully understand the relationship between the values of a feature and the predicted results of the model. The dependence plot reveals the relationship between SHAP values and feature values. In this plot, every instance of the dataset is represented by a point. The scatterplot represents the dispersion of the variable SHAP values versus the variable underlying the row values. SHAP values above the line $y = 0$ lead to prediction of the considered class label. Vertical dispersion in the dependence plot is due to interaction effects with other features. This means that an instance's SHAP value for a feature is not solely dependent on the value of that feature, but is also influenced by the values of the instance's other features. In many cases, interaction effects are not particularly important, but some applications show dramatic interactions between features.

In conclusion, the research illustrated in this paper consists of three steps: the first step is to apply a cluster algorithm finalized at identifying homogeneous aging classes for r.c. bridge components; the second step is to apply a RF model, considering the class identified with the cluster algorithm as the dependent variable and the features not included in the cluster analysis characterizing bridge components as independent variables; and finally, the SHAP analysis is performed, which yields a plethora of plots able to support the interpretation of the results of the RF and the task of features selection, as well as hypothesis generation and the development of an explanation for the bridge component ageing classes. This interpretation is important in the light of the use of the results of the cluster analysis in the context of a broader research, in which the application of stochastic models is foreseen through which it is possible to make estimates of the remaining useful life of the component.

3.3. *Results*

3.3.1. *Bridge substructure components*

Fig. 1 shows the result of the cluster analysis for the components belonging to the bridge substructure. According to this result, three classes could be identified. The "fragile", "normal" and "robust" classes contain 28, 96, 129 data points, respectively. The value of the silhouette width is 0.49: although the data structure is not particularly strong, it can be considered reasonable. In average, fragile bridge substructures reach CI 2

	Substructure				Superstructure			
	Precision	Recall	F1-score	Support	Precision	Recall	F1-score	Support
Robust	0.82	0.85	0.84	27	0.94	0.83	0.88	41
Normal	0.75	0.68	0.71	22	0.71	0.86	0.77	28
Fragile	0.81	0.87	0.84	15	0.67	0.57	0.62	7
Accuracy			0.80	64			0.82	76
Macro avg	0.79	0.80	0.80	64	0.77	0.75	0.76	76
Weighted avg	0.79	0.80	0.79	64	0.83	0.82	0.82	76

Table 1.: Classification report expressing the predicting performance of the trained random forest model.

after 14.5 years and CI 3 after 21.2 years; normal substructures reach CI 2 after 26.7 years and CI 3 after 37.2 years; robust substructures reach CI 2 after 40.2 years and CI 3 after 48.7 years.

The data set is divided into train and test, based on which the RF algorithm can be trained and its performance evaluated. Table 1 shows the resulting classification: an overall accuracy of 0.80 has been obtained. Fig. 2 represents the bar plot which shows the relative importance of each feature, according to the mean SHAP values. The most important feature is ”Year of construction”, followed by ”Road section”, while the feature “Bridge typology” has a very low importance and could be disregarded by the model. Fig. 3 represents the beeswarm plot showing not only the relative importance of each features but also the relationship between the values of the feature and the predicted outcome. This plot reveals that high values of the year of construction (thus corresponding to more recent bridges and bridge components) contribute to the prediction of the fragile class. By also considering the dependence plot (Fig. 4) it is possible to notice that low values (thus corresponding to older bridges and bridge components) contribute to the prediction of the robust class. A possible interpretation is that the results of the inspections related to old bridges belonging to the fragile class are not contained in the database, as data collection only began in the 1990s, when maybe these bridges had already been replaced by others or subject to major maintenance work. On the contrary, only the components of bridges built in a more recent era and belonging to the fragile class have already received a CI equal to 3, which therefore constitute this class only. The second most important feature is the roadway segment, but it has already a very low mean SHAP value. This result may indicate that the evolution of the deterioration might depend slightly on traffic loads and on some environmental conditions shared by objects belonging to the same road segment. However, to test this hypothesis, it would be necessary to collect more data and conduct a more in-depth analysis based on additional variables.

3.3.2. *Bridge superstructure components*

Fig. 5 shows the result of the cluster analysis for the components belonging to the bridge superstructure. Three classes could be identified also in this case. The ”fragile”, ”normal” and ”robust” classes contain 37, 85, 89 data points, respectively. The value of the silhouette width is 0.45, suggesting a slightly weaker data structure than the components belonging to the bridge substructure. In average, fragile bridge superstructures reach CI 2 after 22.3 years and CI 3 after 31.3 years; normal superstructures reach CI 2 after 33.1 years and CI 3 after 42.9 years; robust superstructures reach CI 2 after 42.6 years and CI 3 after 50.4 years.

The RF algorithm is trained and its performance evaluated: Table 1 shows the classification report resulting from the application of the RF, for which overall accuracy of 0.82 has been obtained. According to the bar plot showing the mean SHAP values (Fig. 6), the most important feature is again the year of construction, while the feature “Bridge typology” has again a very low importance and could be disregarded by the model. By considering the beeswarm plot in Fig. 7, higher values of the year of construction (thus corresponding to more recent bridges and bridge components) contribute to the prediction of the class ”fragile”, while lower values (thus corresponding to older bridges and bridge components) contribute to the prediction of the class ”robust” (as revealed by the dependence plot in Fig. 8), and intermediate values contribute to the prediction of the class ”robust”. These confirm the results that were found by analysing the bridge substructure. However, it should be notices that in the case of the bridge superstructure, the range of values for the year of construction is smaller and there are no components in the dataset of recent construction (after the year 1987). The second most important feature is ”Altitude”, however, looking at the relative SHAP plots, it is not possible to establish a clear correlation between its values and the aging family. The impact of altitude on infrastructure aging rate can have opposing effects, and these results were expected.

4. Conclusions

This article suggests a procedure to identify and interpret homogeneous ageing classes for reinforced concrete bridge components. In the approach three different methodologies are combined, namely a cluster analysis, a random forest algorithm and a SHAP analysis. The SHAP analysis allows to explain the results of the classification obtained applying the random forest algorithm and to highlight which variables influence the prediction of the homogeneous ageing classes the most. It emerges that the year of construction of the bridge and therefore of the component has a strong impact on the prediction. Other variables affect the prediction, but to a lesser extent. This allows us to formulate some hypotheses that can deepen the explanation of this result. It is possible to conclude that this analysis offers relevant insights for the use of ageing classes in the context of a broader research, aimed at improving the estimation of the remaining useful life of reinforced concrete bridges. Finally, the authors would like to remark that the application of data-based AI models for the prediction of ageing class or component sojourn time in a given condition class requires a particularly critical attitude.

Acknowledgement

This research study is funded by the FFG project ”ENDURE - Estimation of the remaining service life of bridges through the development and testing of hybrid models”, whose aim is to improve bridge life-cycle assessment analysing the infrastructure management databases of the D-A-CH countries.

References

Aggarwal, C. C. and C. K. Reddy (2014). *Data Clustering, Algorithms and Applications*. Taylor Francis Group.

Breiman, L. (2001). Random forests. *Machine Learning 45*, 5–32.

Confalonieri, R., L. Coba, B. Wagner, and T. R. Besold (2021). A historical perspective of explainable artificial intelligence. *WIREs Data Mining and Knowledge Discovery 11*(1), e1391.

Croce, P., P. Formichi, F. Landi, B. Puccini, V. Zotti, and F. Marsili (2020). Statistical parameters of steel rebars of reinforced concrete existing structures. *Proceedings of the 30th European Safety and Reliability Conference Statistical Parameters of Steel Rebars of Reinforced Concrete Existing Structures*, 4751–4757.

Croce, P., F. Marsili, F. Landi, F. Klawonn, and P. Formichi (2018). Evaluation of statistical parameters of concrete strength from secondary experimental test data. *Construction and Building Materials 163*, 343–359.

Huang, Y.-H. (2010). Artificial neural network model of bridge deterioration. *Journal of Performance of Constructed Facilities 24*(6), 597–602.

Lundberg, S. M. and S.-I. Lee (2017). A unified approach to interpreting model predictions. NIPS’17, Red Hook, NY, USA, pp. 4768–4777. Curran Associates Inc.

Marsili, F., F. Landi, and K. Sylvia (2023). Life-cycle of structures and infrastructure systems. In F. Biondini and D. M. Frangopol (Eds.), *Life-cycle assessment of R.C. bridge components based on cluster analysis and stochastic process*, pp. 2489–2496. Taylor Francis.

Mathavan, S., M. Rahman, and K. Kamal (2015). Use of a self-organizing map for crack detection in highly textured pavement images. *Journal of Infrastructure Systems 21*(3), 04014052.

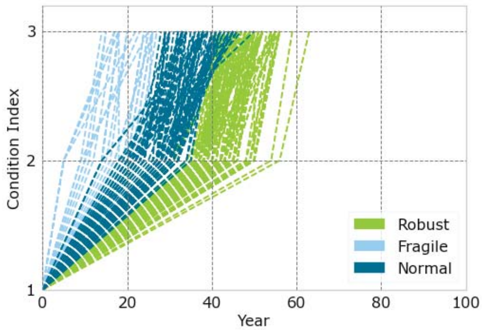

Fig. 1.: Results of the cluster analysis (bridge substructure).

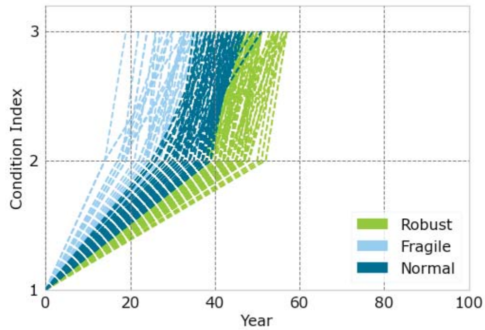

Fig. 5.: Results of the cluster analysis (bridge superstructure).

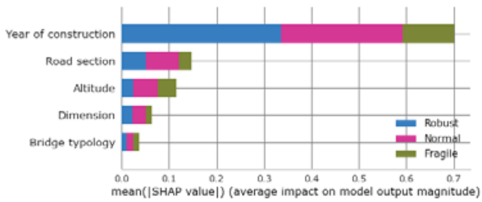

Fig. 2.: SHAP bar plot showing the overall relative importance of each features (bridge substructure).

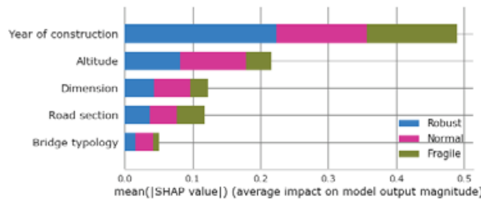

Fig. 6.: SHAP bar plot showing the overall relative importance of each features (bridge superstructure).

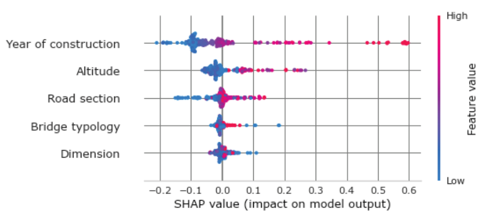

Fig. 3.: SHAP beeswarm plot with respect to the prediction of the class ”fragile” (bridge substructure).

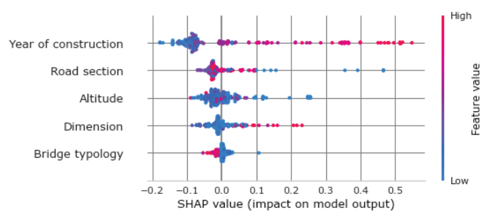

Fig. 7.: SHAP beeswarm plot with respect to the prediction of the class ”fragile” (bridge superstructure).

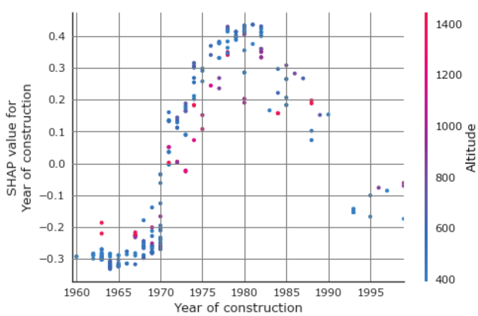

Fig. 4.: SHAP dependence plot of ”Year of construction” with respect to the prediction of class ”normal” (bridge substructure).

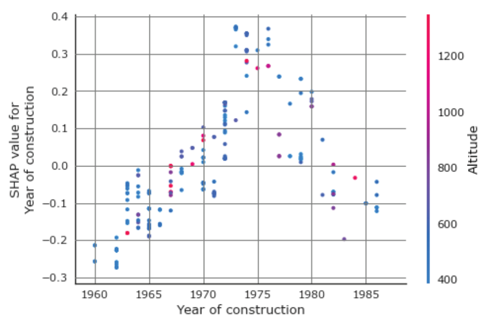

Fig. 8.: SHAP dependence plot of ”Year of construction” with respect to the prediction of class ”normal” (bridge superstructure).