Available online at www.sciencedirect.com

ScienceDirect

Procedia Structural Integrity 38 (2022) 182–191

Structural Integrity Procedia

www.elsevier.com/locate/procedia

FATIGUE DESIGN 2021, 9th Edition of the International Conference on Fatigue Design

Prediction of fatigue failure in small-scale butt-welded joints with explainable machine learning

Moritz Braun*, Leon Kellner, Sarah Schreiber, Sören Ehlers

Institute for Ship Structural Design and Analysis, Hamburg University of Technology, Am Schwarzenberg Campus 4(C), D-21073 Hamburg, Germany

Abstract

Butt-welded joints are common in many industries. The fatigue behavior of such joints depends on numerous factors, e.g. load level, local weld geometry, or parent material strength. To make things worse, these factors often interact, yet mutual influence can hardly be quantified by multivariate studies, i.e. varying one factor at a time out of many factors, due to the large number of required tests and the statistical nature of weld geometry. Consequently, fatigue assessment of such joints often deviates significantly between prediction and experimental result.

Thus, alternative methods are desirable in order to take various influencing factors into account. To this end, machine learning techniques were used to predict failure locations and number of cycles to failure of fatigue tests performed on small-scale butt-welded joint specimens. In addition to accurate predictions, an understanding of importance and mutual influence of the factors is desired, e.g. a ranking of the most important factors; however, capturing the influence of several possibly interacting factors usually requires complex nonlinear machine learning models. We used gradient boosted trees. Since these are black box models, the SHapley Additive exPlanations (SHAP) framework was used to explain the predictions, i.e. identify influential features and their interactions. Lastly, the model explanations are linked back to domain knowledge.

Keywords: Fatigue life prediction, Welded joints, Fatigue strength, Machine learning models, explainable AI, gradient boosted trees, SHAP

* Corresponding author. Tel.: +49-40-42878-6091; fax: +49-40-42731-4469.
E-mail address: moritz.br@tuhh.de

10.1016/j.prostr.2022.03.019

1. Introduction

Fatigue strength of welded joints is influenced by numerous factors. These are related to the magnitude of loading, geometric features of the welded connection, material properties, and environmental effects. To add to this complexity, these factors often interact but this interaction is difficult to quantify. The reason is that experimental investigations of all influential factors, i.e. varying one factor at a time out of many factors, due to the large number of required tests and the statistical nature of weld geometry. Consequently, fatigue assessment, e.g. lifetime prediction, often deviates significantly between model and experiment.

At the same time, with the progress in measuring technologies and condition monitoring, the amount and types of available data increases, e.g. 3D scans of weld geometry. Recently, it was found that a weld geometry measurement every 0.8 mm is required to accurately describe the geometry parameters along seam welds (Renken et al. 2021). Finite element simulations can make use of such data, e.g. through exactly replicating weld geometry. Yet replicating long seam welds is computationally too demanding to be feasible for a high number of analyses.

In short, there is a lack of multivariate studies and a high number of influential factors, which furthermore only increases through new measuring technologies. Consequently, FEM or classic statistical tools struggle to tackle this complexity. Alternatively, machine learning (ML) techniques can quickly process multivariate data, e.g. during in-production or weld quality assessments; however, such techniques need to be verified in case studies first. This study presents an application of ML techniques to analyze a large number of fatigue tests performed on small-scale butt-welded joint specimens. Additionally, the SHapley Additive exPlanations (SHAP) framework was applied to assess the mutual influence of the various influencing factors and to rank these factors by impact. Lastly, the explanations of the ML model results are linked back to structural mechanics domain knowledge.

Nomenclature

ACC	Accuracy
FAT	Fatigue design class
GMAW, FCAW, SAW	Gas-metal, flux-cored, and submerged arc welding
MAE	Mean absolute error
MCC	Matthews correlation coefficient
ML	Machine learning
RMSE	Root-mean-square error
SHAP	SHapley Additive exPlanations
WR, WT	Weld root and weld toe
XAI	eXplainable AI
t, t_1, t_2	Thickness, left, and right thickness
b	Width of specimen
k	Inverse slope exponent of a stress-life curve
$\sigma_{YS,Left}$, $\sigma_{YS,Right}$	Left and right yield strength
e	Axial misalignment
φ	Angular misalignment
N_f	Number of cycles to failure
F_a, F_{max}	Force amplitude and maximum force
σ_a, σ_{max}, $\Delta\sigma_{nom}$	Stress amplitude, maximum stress, and nominal stress range
R	Stress ratio
T	Temperature
W_T, W_B	Width top and bottom side
H_T, H_B	Height top and bottom side
r_{TL}, r_{TR}, r_{BL}, r_{BR}	Weld toe radius: top left, top right, bottom left, bottom right
α_{TL}, α_{TR}, α_{BL}, α_{BR}	Notch angle: top left, top right, bottom left, bottom right
u_{TL}, u_{TR}, u_{BL}, u_{BR}	Undercut depth: top left, top right, bottom left, bottom right

2. Methods

2.1. Machine learning models

The features we used are expected to exhibit various nonlinear interactions. Furthermore, the data contains missing or unusable values. This can make the use of classic statistical tools difficult, whereas machine learning (ML) models are particularly useful under such circumstances (Larranaga et al. 2006). Hence, as an alternative to classic statistical tools, ML approaches are used to predict fracture location and lifetime of welded specimens.

As ML models, gradient boosted trees were used since the data is tabular and tree-based algorithms perform well on this type and size of data (Klambauer et al. 2017; Lundberg et al. 2020). The XGBoost implementation (Chen and Guestrin 2016) was employed due to good performance in preliminary analyses and its integration with the explainability framework SHAP. The applied hyperparameters loosely follow the suggestions given in (Friedman 2002; Hastie et al. 2009): 500 trees, a learning rate of 0.07, a maximum number of leaves of 8, and a subsample size of 50 %. The same hyperparameters were used in both ML models.

Following standard ML procedure, the data was split into training (80%) and test data (20%). To evaluate the generalizability of the ML models, four-fold cross-validation was used on the training data (which, for this purpose, is again split into fold training and test data) (Refaeilzadeh et al. 2009). The test data remains unseen by the algorithm until final validation. Lastly, training was stopped if the model did not improve over five rounds of training. This is termed early-stopping and prevents overfitting, i.e. the detection of patterns in the noise of the data.

2.2. Explainability

Data-based models are often criticized for the lack of understanding and knowledge generated by using them, e.g. (Schmidt et al. 2019). Such a lack of understanding generally limits the usefulness of models. This is especially true if the user is interested in knowledge discovery. Hence, in addition to the accuracy of predictions, understanding why a model has made a prediction has become a key challenge in ML.

As a result, various tools under the umbrella term of explainable AI (XAI) have been developed. These tools can reveal which features drive model predictions, both for single observations as well as globally, i.e. for the complete data set. We used the SHAP framework which is based on game-theoretic approaches and Shapley values. By contrast to other XAI methods, SHAP has a solid theoretical foundation and is not based on heuristics (Molnar 2020).

The SHAP values are the difference between the expected value E, which is the average ML model output over a data set, and the actual prediction by the ML model. They can be interpreted as the impact of a feature on the prediction for one observation. As an example, take a binary classification task where $E = 0.5$ whereas the model predicts 0.9 for an observation. All SHAP values attributed to this observation would add up to the difference of 0.4. For tabular data with i features and k rows, the SHAP values are a $i \times k$ –matrix with the units of the prediction space. In the example, the k SHAP values in the row corresponding to the i-th observation would add up to 0.4. The information stored in this matrix is then used for further analyses. See also Appendix A.

3. Fatigue test data of small-scale butt-welded joints

In total, the data of 556 fatigue tests has been collected from previous studies of the authors (Braun et al. 2020b; Braun et al. 2021) and other unpublished projects. After initial data cleaning, e.g. removing missing values and run-outs, about 420 samples remain. To limit influencing factors, all specimens were cut from welded steel plates of 1000 mm × 500 mm (made in shops in flat 1G (PA) position) into 40-mm-wide and 500-mm-long stripes. The thickness varied within the range of 10 to 20 mm.

The quality of welded joints often depends on the chosen welding procedure. Some welding processes, such as flux-cored arc welding (FCAW), are known to produce weld transitions with favorable fatigue properties due to large weld toe radii. Thus, different welding techniques (GMAW, FCAW, SAW) were applied to create joints with varying weld quality. Other important characteristics of the test series include different weld shapes (I-, V-, Y-, and X-groves), parent material strengths in the range S235 to S690, small variations in left and right plate thickness, dissimilar joints, welding with or without root face, and on temporary root backing.

To include the effect of geometrical variations and misalignment in the investigation (see Fig. 1), all specimens were analyzed prior to testing. For the measurement of axial and angular misalignment a test setup with dial gauges was used, see (Braun et al. 2020a), and the local weld geometry was obtained by laser triangulation and post-processing by means of the curvature method, see (Schubnell et al. 2020; Renken et al. 2021). Mean geometry parameters—obtained along the welds using a slice density of 0.2 slices/mm—are used as input for the ML models.

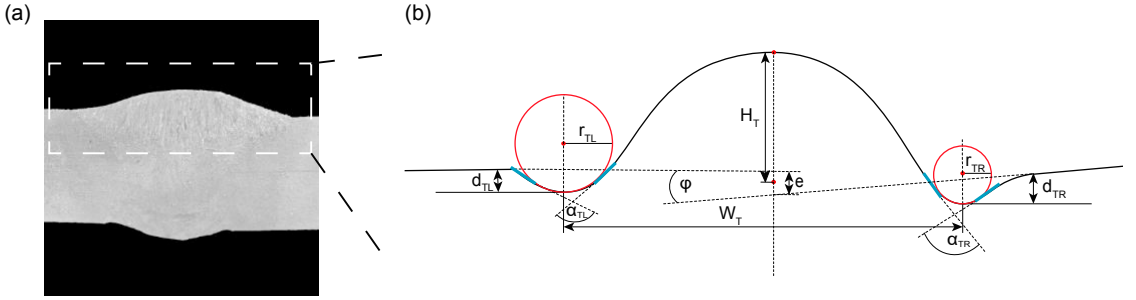

Fig. 1. Micrograph of specimen (a) with schematic presentation of the top side geometry of butt-welded joints (b), adopted from (Braun 2021a)

All tests have been performed on Schenk horizontal resonance machines at a test frequency of around 30 Hz and a varying stress ratio R. Only by performing all tests on the same test machines under identical conditions, it can be prevented that the results are flawed by external factors. Besides the specimen's misalignment and local weld geometry, environmental effects such as temperature affect fatigue strength (Braun et al. 2020c). Thus, data obtained from tests at sub-zero temperatures has been included in the data base. Finally, to be consistent with failure locations, all test specimens have been oriented in the test machine as they were oriented during welding, i.e. weld roots from temporary root backing are located on the bottom side.

The fatigue test results of all specimens are presented in Fig. 2 including corresponding design curves based on failure location (i.e. weld toe (WT) or weld root (WR)). We can see a large scatter of test results due to different testing conditions (level of misalignment and stress ratio) as well differences in local weld geometry. Typically, failure locations in butt-welded joints depend on both types of influencing factors.

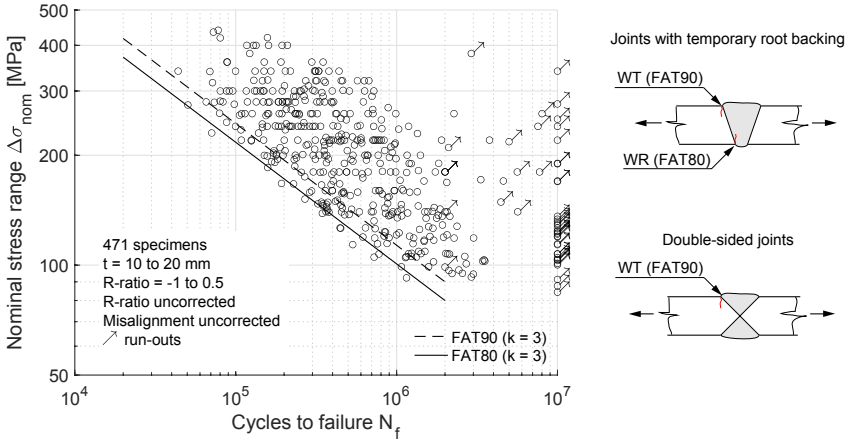

Fig. 2. Fatigue test results of small-scale butt-welded joints with corresponding fatigue design curves from (Hobbacher 2016)

4. Results

4.1. Performance of machine learning models

The model outputs and corresponding explanations are meaningful only if the performance is good. The fracture location classifier had an accuracy of 76 % and a Matthews correlation coefficient of 0.68. Though the overall

accuracy could be better, the classifier performs well on predicting *all* classes. This is reflected in the confusion matrix in Fig. 3 which gives the fraction of predicted- versus true classes for the test data. For instance, if a sample truly fractured top left (TL), this was correctly predicted as TL for 67% of the test data (the corresponding cell is in the first row, first column of the confusion matrix), whereas the incorrect predictions were 20% top right (TR), and equally 6.7% bottom left (BL) and bottom right (BR) (cells in first row and second, third, and fourth column). In most cases, the classifier correctly predicted whether a fracture will develop on the top or bottom side, whereas the combination of bottom and top as well as left or right appears to be less predictable.

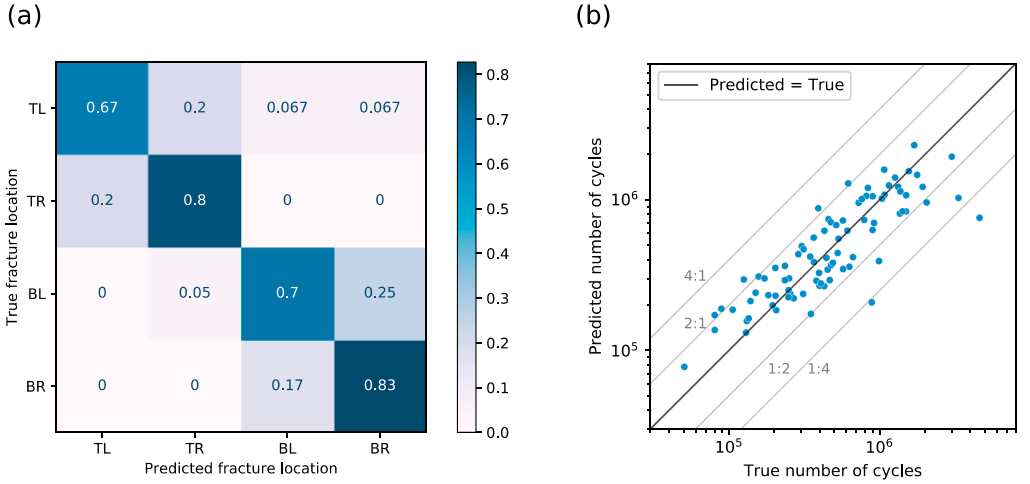

Fig. 3. (a) Confusion matrix indicating the true versus predicted fraction of fracture locations for the classifier (TL is top left, TR is top right, BL is bottom left, BR is bottom right); (b) Plot of predicted versus true number of cycles for the lifetime model.

For the lifetime prediction model, interpretation of average errors is less intuitive. As an alternative, in Fig. 3 all predictions of the test data are plotted over the true values. This indicates that the lifetime model performed well, albeit with a slight trend to underpredict for a high number of cycles, and overpredict for a lower number of cycles. For completeness, the errors on logarithmized targets are: $RMSE = 0.21$ and $MAE = 0.16$.

Lastly, during cross-validation, the classifier had an accuracy of 0.71 ± 0.07 and a MCC of 0.61 ± 0.09, the lifetime model had a RMSE of 0.2 ± 0.01 and a MAE of 0.15 ± 0.01 (mean $\pm$ standard deviation, $n = 4$). These performance values are similar to the above-described performance of the final models (i.e. trained on all training data) showing that the models generalize well. In all, both the fracture location classifier as well as the lifetime prediction model performed well and the results merited further analysis.

4.2. Global model insights

The ML models were not only created to achieve an accurate prediction of fracture location and lifetime, but to gain insight into the data and feature relations. That is to say, we are interested in which features have the most impact on the predictions and furthermore if these relations match expectation based on theory. All analyses in this section are based on the SHAP values. Recall that one SHAP value is the impact of a feature for one prediction, see Sect. 2.2 and Appendix A.

4.2.1. Fracture location

To begin with, the global impact of features is shown in Fig. 4 with a ranking of the average impact of the 10 most important features. The bars are the stacked impact on the prediction of the different classes or fracture locations, respectively. The distribution of the features' impact is relatively even between classes. Not surprisingly, the angular and axial misalignment have a high predictive power.

In Fig. 4(b) a set of beeswarm plots is exemplarily shown for the 'top left' class. Features are also sorted by global impact, that is, they have the same order as in Fig. 4(a). In addition to that, each dot is one sample, where its color

represents its feature value and its horizontal position shows the impact the feature has on the model output (for that sample). Dots pile up vertically to indicate density. This type of plot can reveal further effects, e.g. rare but high magnitude effects, or asymmetric impact of a feature. (Lundberg et al. 2018)

From a structural mechanics' perspective, a high (positive) angular misalignment increases secondary bending stresses on the top side of a specimen and thereby increases the likelihood of failure at either top left or top right weld toes. Similarly, axial misalignment leads to high secondary bending stresses on either left or right side of a weld depending on the direction of axial misalignment. Comparing the mean SHAP values, i.e. the mean impact of a feature, in Fig. 4(a), misalignment effects have a higher effect on failure location than local weld geometry. The slightly unsymmetrical distribution (Fig. 4(b)) for both misalignment types is expected to be related to effects of the local weld geometry and weld shape (i.e. V-type and Y-type vs. X-type grooves). Other important features are parameters of the top side weld geometry (weld toe radii and notch angles) but also on the bottom side such as the undercut depth u_{BL}.

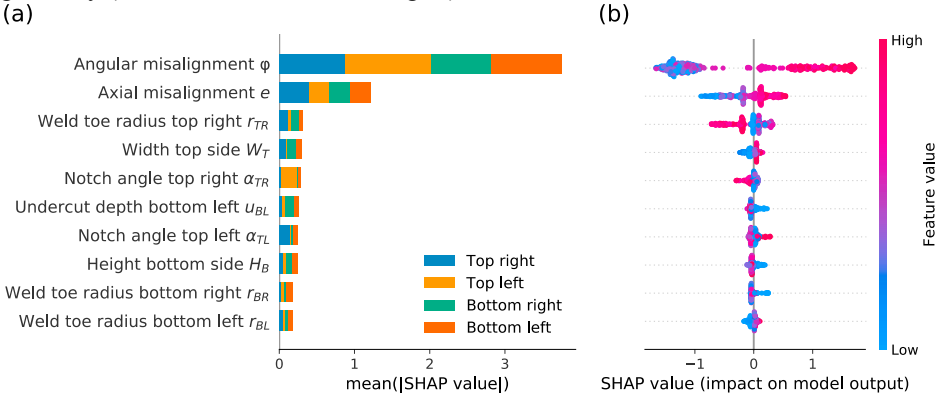

Fig. 4. (a) Average impact of features on fracture location for different possibly locations; (b) Set of beeswarm plots indicating SHAP value (horizontal axis) and feature value (colorbar) for every sample in the test data.

Next, assessing the impact of features one by one can reveal additional patterns. Figures 5(a) and (b) are scatter plots, also termed partial dependence plots, where each dot is one sample. The vertical axis is the SHAP value (the impact of a feature for a prediction) and the horizontal axis is the feature value. Additionally, the dots' color indicates the value of a second feature (termed interaction feature).

Vertical color patterns reveal a possible interaction between both features, whereas horizontal color patterns show that the features behave proportionally, i.e. if one feature increases the other does too. Both figures present the impact of distinct features on the probability of fatigue crack initiation at the top left weld toe.

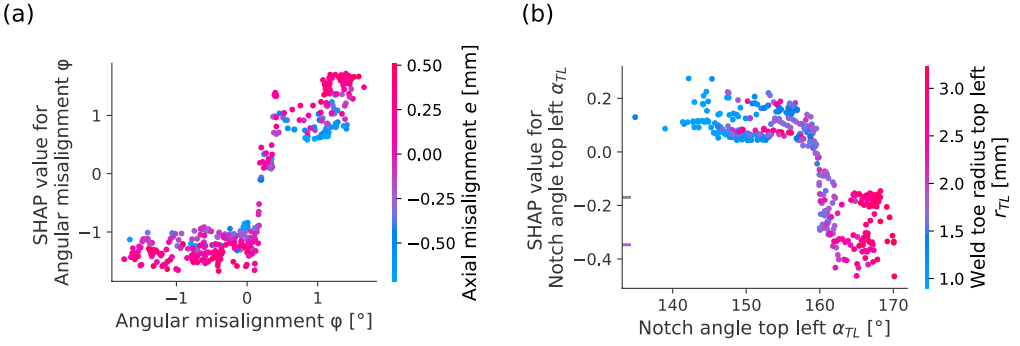

Fig. 5. SHAP dependence plots for (a) angular- (φ) and axial (e) misalignment; and (b) for top left notch angle (α_{TL}) and weld toe radius (r_{TL}).

Starting with Fig. 5(a), the mutual influence of both types of misalignments is visible. As mentioned, angular misalignment φ leads to tensile secondary bending stresses on either top- or bottom side of a specimen. Hence, a positive φ increases the probability (i.e. positive SHAP value) of fracture at the top left weld toe whereas a negative

φ decreases it. In combination with positive axial misalignment e, the probability is further increased or decreased depending on the sign of angular misalignment. In other words, the impact of φ is amplified or damped by e.

Fig. 5(b) presents two influential local weld geometry parameters with respect to failure at the top left weld toe. A small notch angle $\alpha_{TL} < 155°$ and weld toe radius $r_{TL} < 2$ mm clearly increase the probability of fracture at this location; however, the absolute SHAP value is significantly smaller than those obtained for the feature value angular misalignment (Fig. 5(a)). This is confirmed by the global feature ranking. Interestingly, a large notch angle α_{TL} and weld toe radius r_{TL} seem to have a stronger impact on fracture at other locations than small values have on this location. This can be seen from the unsymmetrical range of SHAP values from about 0.3 to -0.5 (Fig. 5(b)).

4.2.2. Lifetime

The SHAP values for the prediction of lifetime indicate a positive or negative effect of features on the number of cycles to failure. Similar to the previous assessment, features are ranked based on their average impact on model output. As expected, stress amplitude has by far the highest effect on the lifetime. The distribution of dots in Fig. 6(b) supports this conclusion. Again, angular misalignment has a high impact on the model prediction due to its significant impact on secondary bending stresses. Axial misalignment is expectedly less relevant than angular misalignment for small-scale specimens. For large-scale specimens, however, the opposite is often true, see (Lotsberg 2009).

The following load-related parameters (force amplitude, maximum force, and maximum stress) have all less significance than the first two; however, all three affect secondary bending stresses and are related to the stress amplitude. Thus, a high impact is expected. Regarding the geometrical features, height on the top side (H_T) is the most influential feature. This is surprising as fatigue strength is often associated with the local notch geometry at weld toes. Due to welds being fabricated with one or two weld caps, there is probably a correlation between weld geometry and local weld toe geometry. Additionally, H_T affects the nominal stress at weld roots for V-shaped groves (weld transitions on bottom side). This fact might explain the higher ranking of H_T compared to other geometrical features. The test temperature and material strength have less influence than the above-mentioned features. This is in line with expectations for butt-welded joints based on statistical assessments of fatigue tests performed at sub-zero temperatures (Braun 2021b).

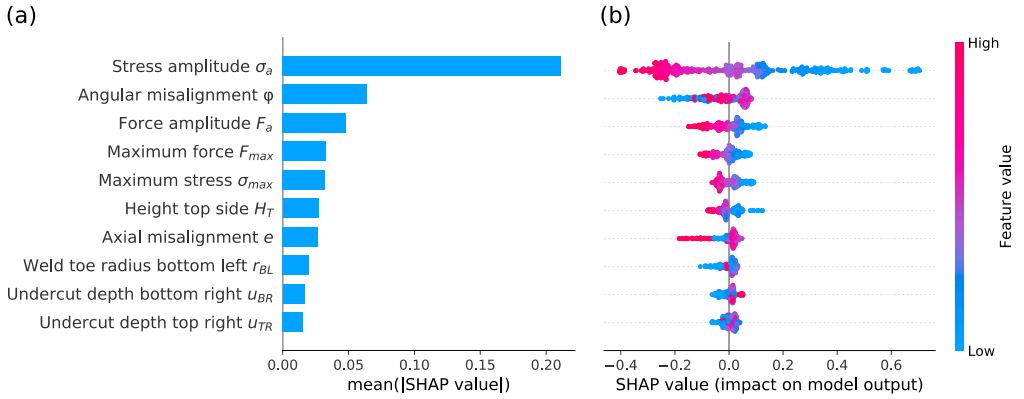

Fig. 6. (a) Average impact of features on lifetime; (b) Set of beeswarm plots indicating SHAP value (horizontal axis) and feature value (colorbar) for every sample in the test data.

Fig. 7 indicates the mutual influence of the two most important features on lifetime of the assessed butt-welded joint specimens. Four interesting aspects are apparent from this figure: Firstly, the distinct plateau of positive SHAP values is limited to $\varphi < \pm 0.5°$. Second, outside this interval and especially for $\varphi \geq 1°$ the SHAP values and thereby number of cycles to failure are significantly decreasing, which agrees well with limits on weld quality, see ISO 5817:2014. Third, within $\varphi < \pm 0.5°$ the effect of stress amplitude on SHAP value is smaller than outside. This is expected as secondary bending stresses due to misalignment are a function of φ and stress amplitude σ_a. Lastly, the negative effect of angular misalignment on SHAP values is more pronounced than vice versa and the distribution is unsymmetrical with respect to zero angular misalignment. This is an indication for other strong influencing factors,

e.g. related to unsymmetrical weld shapes (Y- and V-groove). Such effects can be assessed in detail considering the impact of individual features.

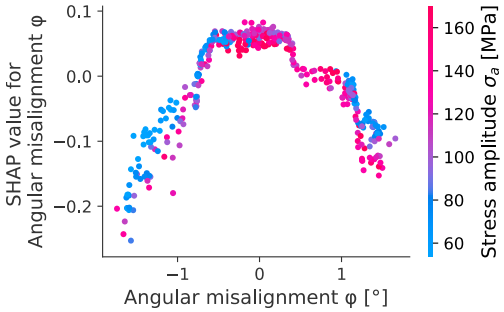

Fig. 7. SHAP dependence plot for angular misalignment (φ) and stress amplitude (σ_a).

5. Discussion

In all, the fatigue behavior of welded joints was successfully analyzed with tree-based machine learning (ML) models. The models were robust with respect to the input data as well as model parameters and, as shown with cross-validation, are expected to generalize well to new and previously unseen data. The model performance is good considering the relatively small database. Specifically, the ratio of samples vs. the number of features is probably suboptimal. Nonetheless, as the database is continuously extended, the model performance will likely increase in the future due to more available training data.

Generally, the results are in line with theoretical expectations. The added value of using ML models is a quantitative ranking of feature impact over a range of different tests. Moreover, a detailed assessment of the impact of single features and their interaction with secondary features is possible through SHAP dependence plots. These plots revealed some unexpected behavior such as the relatively high impact of the ‘height top side’ feature. With such analyses, multivariate studies are mimicked, though it should be kept in mind that these are still correlation-based models. Again, as more data becomes available in the future, similar discoveries are likely.

Lastly, ML techniques and investigations into mutual influencing factors on fatigue strength of welded joints help improving lifetime- and failure location prediction accuracy. Compared to finite element simulations, they offer a possibility to process large amounts of data, e.g. during in-production or weld quality assessments of seam welds. The main benefit is that once such methods are calibrated, they are practically effortless during application.

6. Conclusions

The obtained results agree well with expectations based on structural mechanics but highlight interactions that would usually require large number of simulations or time-consuming comparisons to test results. Consequently, machine learning techniques offer new perspectives on fatigue-relates problems and data comparisons. The following conclusions are drawn:

- With two machine learning models, a good prediction accuracy of failure locations and lifetime of small-scale butt-welded joints was achieved. The fracture location classifier had an accuracy of 76 % and a Matthews correlation coefficient of 0.68. The lifetime model performed well, with a slight trend to underpredict for a high number of cycles, and overpredict for a lower number of cycles. Both models generalized well.
- Explainable machine learning methods are capable of predicting the mutual influence of various influencing factors on failure locations and lifetime of welded joints. They also allow to rank these factors by impact, which can reveal an unexpectedly high- or low impact for specific features.
- The most influential parameters on failure location of small-scale butt-welded joints were macro-geometric axial and angular misalignment. For lifetime prediction, the most important features are load-related (stress and force amplitude, and corresponding maximum values) as well as angular misalignment. Among geometrical features,

weld height on the top side was surprisingly found to be the most influential feature followed by undercut depths on the bottom side. Both are expected to be related to the weld shapes and welding procedures (V- or Y-type groove and welding on temporary root backing).

- Finally, understanding the mutual influence of various features on fatigue strength of welded joints will improve the understanding of fatigue phenomena, and offer in-production fatigue assessment and accurate condition monitoring much quicker than e.g. with finite element simulations.

Acknowledgements

The authors would like to thank Shi Song and Finn Renken for performing the weld geometry measurements.

Appendix A. Calculation of SHAP values

To compute SHAP values, the following holds: f is the model, $f(\boldsymbol{x})$ a prediction for input vector $\boldsymbol{x}$. $f_x(S) = E[f(\boldsymbol{x})|\boldsymbol{x}_S]$ is the expected value of f conditioned on a subset S of the input features. Furthermore, the number of input features is M and N the set of all input features. Then

$$\phi_i = \sum_{S \subseteq N/\{i\}} \frac{|S|!(M-|S|-1)!}{M!} \left(f_x(S \cup \{i\}) - f_x(S)\right) \tag{1}$$

Now, ϕ_i is the difference between $E[f(\boldsymbol{x})|\boldsymbol{x}_1]$ and $E[f(\boldsymbol{x})]$. In general

$$\sum_{i=0}^{M} \phi_i = f(\boldsymbol{x}) \tag{2}$$

Appendix B. Performance metrics

The following metrics were used for evaluating performance. For the fracture location classifier, accuracy (ACC) and Matthews correlation coefficient (MCC) were used. These metrics are based on the number of true positive predictions (TP), true negatives (TN), false positives (FP), and false negatives (FN). The accuracy is simply the number of correct predictions divided by the number of all predictions.

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} = \frac{TP+TN}{\text{all predictions}} \tag{3}$$

and takes values $\in [0, 1]$.

Since the dataset is not balanced with respect to fracture location, i.e. the number of samples belonging to each class varies, the MCC was used additionally. Its value is only high if the classifier does well on the prediction of all classes. The following definition holds: with C the number of (output) classes, and S the number of samples, with two $S \times C$ matrices, X the predicted- and Y the true classes. $X_{ik} = 1$ if the sample i is predicted to be in class k and $X_{ik} = 0$ otherwise. $Y_{ik} = 1$ if i truly belongs to class k and 0 otherwise. Then the covariance between X and Y is

$$\text{cov}(X, Y) = \sum_{k=1}^{C} \frac{1}{C} \text{cov}(X_k, Y_k) \tag{4}$$

and the MCC is defined as

$$MCC = \frac{\text{cov(X,Y)}}{\sqrt{\text{cov(X,X)cov(Y,Y)}}} \tag{5}$$

For two classes (binary), this results in

$$MCC_{\text{binary}} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{6}$$

The MCC takes values $\in [-1, 1]$ where $+1$ is complete agreement, 0 means the prediction is no better than random and -1 signifies a complete disagreement. See also (Jurman et al. 2012).

For evaluating the performance of the lifetime prediction, two regression metrics were used: namely the root-mean-square-error (RMSE)

$$RMSE = \left[\frac{1}{S}\sum_{i=1}^{S}(y_i - x_i)^2\right]^{\frac{1}{2}} \tag{7}$$

with again the number of samples S, the true label y as well as the prediction x for sample i. Furthermore, the mean absolute error (MAE) is given due to its straightforward interpretation as absolute error.

$$MAE = \frac{1}{S}\sum_{i=1}^{S}|y_i - x_i| \tag{8}$$

References

Braun M. 2021a. Assessment of fatigue strength of welded steel joints at sub-zero temperatures based on the micro-structural support effect hypothesis. Technische Universität Hamburg.

Braun M. 2021b. The effect of sub-zero temperatures on fatigue strength of welded joints. International Institute of Welding IIW-Doc. XIII-2888-2021.

Braun M, Kahl A, Willems T, Seidel M, Fischer C, Ehlers S. 2021. Guidance for Material Selection Based on Static and Dynamic Mechanical Properties at Sub-Zero Temperatures. *Journal of Offshore Mechanics and Arctic Engineering*. 143(4), 1-45. https://doi.org/10.1115/1.4049252

Braun M, Milaković A-S, Andresen-Paulsen G, Fricke W, Ehlers S. 2020a. A novel approach to consider misalignment effects in assessment of fatigue tests. *Ship Technology Research*. submitted for publication,

Braun M, Milaković A-S, Ehlers S, Kahl A, Willems T, Seidel M, Fischer C. 2020b. Sub-Zero Temperature Fatigue Strength of Butt-Welded Normal and High-Strength Steel Joints for Ships and Offshore Structures in Arctic Regions. ASME 2020 39th International Conference on Ocean, Offshore and Arctic Engineering; June 28-July 3; Fort Lauderdale, FL, USA.

Braun M, Milaković A-S, Renken F, Fricke W, Ehlers S. 2020c. Application of Local Approaches to the Assessment of Fatigue Test results obtained for Welded Joints at Sub-Zero Temperatures. *International Journal of Fatigue*. 138. https://doi.org/10.1016/j.ijfatigue.2020.105672

Chen T, Guestrin C. 2016. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16 22nd ACM SIGKDD International Conference*. San Francisco, CA, USA. New York, USA: ACM Press: Association for Computing Machinery.

EN ISO 5817:2014 Welding - Fusion welded joints in steel, nickel, titanium and their alloys (beam welding excluded) - Quality levels for imperfections. 2014. Brussels, BE:

Friedman JH. 2002. Stochastic gradient boosting. *Computational Statistics & Data Analysis*. 38(4), 367-378. https://doi.org/10.1016/s0167-9473(01)00065-2

Hastie T, Tibshirani R, Friedman J. 2009. *The Elements of Statistical Learning*. Springer Science & Business Media. (Springer Series in Statistics). https://doi.org/10.1007/978-0-387-84858-7

Hobbacher AF. 2016. *Recommendations for Fatigue Design of Welded Joints and Components*. 2nd ed. Springer International Publishing Switzerland. (IIW Collection). https://doi.org/10.1007/978-3-319-23757-2

Jurman G, Riccadonna S, Furlanello C. 2012. A comparison of MCC and CEN error measures in multi-class prediction. *PLoS One*. 7(8), e41882. https://doi.org/10.1371/journal.pone.0041882

Klambauer G, Unterthiner T, Mayr A, Hochreiter S. 2017. Self-normalizing neural networks. Curran Associates Inc.; Proceedings of the 31st International Conference on Neural Information Processing Systems; Long Beach, California, USA.

Larranaga P, Calvo B, Santana R, Bielza C, Galdiano J, Inza I, Lozano JA, Armananzas R, Santafe G, Perez A et al. 2006. Machine learning in bioinformatics. *Brief Bioinform*. 7(1), 86-112. https://doi.org/10.1093/bib/bbk007

Lotsberg I. 2009. Stress concentrations due to misalignment at butt welds in plated structures and at girth welds in tubulars. *International Journal of Fatigue*. 31(8-9), 1337-1345. https://doi.org/10.1016/j.ijfatigue.2009.03.005

Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee SI. 2020. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat Mach Intell*. 2(1), 56-67. https://doi.org/10.1038/s42256-019-0138-9

Lundberg SM, Erion GG, Lee S-I. 2018. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:180203888*.

Molnar C. 2020. *Interpretable machine learning*. Morrisville, United States: Lulu.com.

Refaeilzadeh P, Tang L, Liu H. 2009. Chapter Chapter 565, Cross-Validation. In: Liu L, ÖZsu MT, editors. *Encyclopedia of Database Systems*. Boston, MA: Springer US; p. 532-538.

Renken F, von Bock und Polach RUF, Schubnell J, Jung M, Oswald M, Rother K, Ehlers S, Braun M. 2021. An algorithm for statistical evaluation of weld toe geometries using laser triangulation. *International Journal of Fatigue*. 149. https://doi.org/10.1016/j.ijfatigue.2021.106293

Schmidt J, Marques MRG, Botti S, Marques MAL. 2019. Recent advances and applications of machine learning in solid-state materials science. *npj Computational Materials*. 5(1). https://doi.org/10.1038/s41524-019-0221-0

Schubnell J, Jung M, Le CH, Farajian M, Braun M, Ehlers S, Fricke W, Garcia M, Nussbaumer A, Baumgartner J. 2020. Influence of the optical measurement technique and evaluation approach on the determination of local weld geometry parameters for different weld types. *Welding in the World*. 64(2), 301-316. https://doi.org/10.1007/s40194-019-00830-0