Received: 13 October 2023 | Revised: 19 February 2024 | Accepted: 7 April 2024

DOI: 10.1111/ffe.14315

ORIGINAL ARTICLE

FFEMS Fatigue & Fracture of Engineering Materials & Structures WILEY

Exploration of materials fatigue influence factors using interpretable machine learning

Christian Frie[1,2] | **Ali Riza Durmaz**[3] | **Chris Eberl**[2,3]

[1]CR, Robert Bosch GmbH, Renningen, Germany

[2]Laboratory for Micro and Materials Mechanics, University of Freiburg, Freiburg im Breisgau, Germany

[3]Fraunhofer Institute for Mechanics of Materials IWM, Freiburg im Breisgau, Germany

Correspondence
Christian Frie, CR, Robert Bosch GmbH, 71272 Renningen, Germany.
Email: cf300@students.uni-freiburg.de

Funding information
None reported.

Abstract

Data-driven fatigue strength predictions are gaining popularity. Nevertheless, many machine learning models lack trustworthiness due to their limited decision-making transparency which often hinders their practical application. In this investigation, we assess the expressiveness of the model-agnostic explainable AI method known as SHapley Additive exPlanations (SHAP) for data-driven fatigue strength prediction. Our study demonstrates that the SHAP feature sensitivity analysis underpins known physical relations from materials processing and fatigue theory. This even applies in view of the high-dimensional, cross-correlated fatigue feature space and despite data heterogeneity (different steels, component designs, and loads). For instance, SHAP indicates a fatigue strength increase with higher solid solution-strengthening element concentrations, such as chromium and nickel. SHAP identifies correlations rather than causality. Thus, data science and domain knowledge should be closely linked during the SHAP assessment. If this is satisfied, plausible causal relations can be inferred, and spurious ones arising from confounding variables can be discarded.

KEYWORDS

explainable AI methods, fatigue strength prediction, machine learning, material informatics

Highlights

- Fatigue strength prediction used an extended feature space (including specimen design).
- We showcase the benefit of SHapley Additive exPlanations (SHAP) interpretability over standard exploratory data analysis.
- The SHAP values extract cause–effect trends which are aligned with fatigue literature.
- An ideal future explainable AI (XAI) method meeting the material fatigue requirements is discussed.

1 | INTRODUCTION

Metal fatigue refers to the progressive degradation of metallic components caused by cyclic loading, leading to structural failure. Designing fatigue-resistant structural components requires the consideration of numerous influencing factors. Fatigue has an inherent multiscale characteristic where component geometry, applied load, surface quality, and microstructure characteristics jointly dictate the cyclic accumulation of plasticity at a microscopic scale, crack nucleation at the atomic scale, and different crack growth regimes at the microscopic, mesoscopic, and macroscopic scale. Capturing the underlying relations in constitutive equations to predict fatigue strength demands significant modeling and computational efforts[1–6]. Addressing this complexity comprehensively through conventional physics-motivated methods to date is challenging because of computational constraints and the lack of complete mechanistic knowledge. Data-driven approaches have emerged as a popular alternative for predicting fatigue strength, which marks a paradigm change in the fatigue modeling approach[7–13]. Many machine learning (ML) techniques can capture, autonomously extract, and represent intricate patterns in data that are difficult to identify through exploratory data analysis and encode in constitutive models. For instance, the universal approximation theorem[14] describes the representation power of artificial neural networks. By leveraging datasets and applying advanced ML techniques, the fatigue performance can be estimated efficiently and accurately without the extensive computational and modeling burden required by traditional physics-motivated methods. The rising popularity of data-driven approaches offers promising avenues for enhancing the design and optimization of components made from metal, ultimately leading to safer and more reliable engineering solutions.

However, data scarcity is a major concern in materials science and engineering. Moreover, model generalizability is considered another major limitation in applying ML techniques in various specialized domains. In natural science and engineering domains where the ambition for a holistic causal understanding is deeply ingrained, data-driven approaches have further been viewed critically due to nonphysical, black box function mapping, and the lack of traceability of the decision-making process.

Moreover, the available data in the domain of fatigue modeling are often particularly scarce, heterogeneous, and incomplete due to time and cost-intensive characterization efforts with different underlying testing standards. Cross-correlated features are common in materials science applications (e.g., quenching rate, hardness, and yield strength), which can be challenging in ML applications and especially in explainability analysis. Removing features based on such cross-correlations is nontrivial since fatigue prediction is an extreme value statistics problem where even fluctuations in weakly associated features can matter. Specifically, such features might contain information about the occurrence of critical defects in microstructures which dictate failure. Aside from these aspects, the unknown data quality compromises the trust in AI-models for fatigue prediction where reliability is of utmost importance. This drives the demand for explainable AI (XAI) methods in data-driven fatigue modeling.

1.1 | Related work

Several publications have addressed data-based fatigue modeling and fatigue strength predictions[7–13], differing in the applied AI methods. He et al[10] and Xiong et al[13] used symbolic regression to derive an analytical expression of the fatigue limit to impose a white-box model that can be physically interpreted. Unfortunately, symbolic regression, an evolutionary algorithm, is limited to a few explanatory variables, and the complexity of the derived analytical expression increases significantly due to complex feature interactions.[15] Weichert et al[12] used a Gaussian process regression (GPR) with a modified kernel function, which considers the relation of hardness and fatigue strength for determining the fatigue strength of stainless steels. GPR is a powerful ML method but similar to symbolic regression, limited to a small number of explanatory variables due to the curse of dimension.[16] Thomas et al[17] use the integrated gradients (IG) method on graph convolutional networks to derive a feature importance analysis for local fatigue damage prediction in a specific material. Kolyshkin et al[11] sequentially add features to train a random forest (RF) model to determine the most important features for fatigue strength prediction. The associated decrease in root-mean-squared error (RMSE) was utilized as a scoring function to include specific features. Agrawal and Choudhary[7] determine the most important features of an RF model with the decrease of impurity, a model internal metric, among the different trees in an ensemble. Still, neither the methods proposed by Agrawal and Choudhary,[7] nor Kolyshkin et al[11] can explain how a particular feature contributes to the fatigue strength prediction, as the applied methods only indicate a feature importance ranking. Determining the importance of a feature in fatigue prediction is not enough for thorough model validation and gaining insights into learned relations. Instead, partial dependence plots are required that show how a particular feature contributes to the overall prediction and how its contribution varies depending on other features.

Thus, the development of advanced XAI methods has gained increased importance in recent years.[18] XAI methods can be separated into local and global interpretability methods.[19] Local methods explain how the model's output is derived for a *particular prediction*. Global interpretability approaches seek to understand the overall structure behind the black box model to validate the global model behavior. Depending on the application, both method categories require domain knowledge for validation. Two of the most popular XAI methods for supervised ML models are LIME (Local Interpretable Model-agnostic Explanations)[20] and SHAP (SHapley Additive exPlanations).[21]

In this work, SHAP has been implemented for two reasons: First, it provides local and global interpretation analysis. Secondly, in contrast to LIME, SHAP has a solid theoretical foundation. Specifically, SHAP is based on the Shapley values, initially developed in the context of game theory, where the Shapley values provide a fair way to determine the contribution of each player to distribute a profit.[22] One downside of the Shapely values is the computational effort, especially for large datasets or deep learning models. For most applications, however, since materials science data are usually scarce to date, this might not be of significant concern. Several SHAP extensions have been developed to reduce the computational effort.[23] One disadvantage in the context of fatigue life prediction comes from the underlying assumption that features are independent. Therefore, the applicability of SHAP to identify feature contributions (e.g., chemical composition, heat treatment parameters, loading conditions) on fatigue lifetime will be investigated in detail. The importance of complying with the feature independence assumption will be investigated in this work. Our contributions are as follows:

- We analyze the performance of different feature combinations for fatigue strength prediction using a RF model and derive suggestions for future data collections to improve the predictive quality further.
- We apply the popular XAI, SHAP, to the RF model to investigate and discuss the data-based results from a physical/engineering perspective on multiple scales.
- We identify and discuss assumptions and limitations of current XAI methods in data-driven fatigue strength applications.

This study's outcomes contribute significantly to adopting XAI methods in material informatics and particularly fatigue strength prediction. The research emphasizes the importance of transparent AI models in addressing complex multiscale engineering challenges by providing valuable insights for engineers and data scientists. The integration of XAI methods empowers engineers to optimize material selection and design, enhancing the reliability and durability of critical structural components in a wide range of engineering applications.

2 | FUNDAMENTALS AND METHODOLOGY

The methodology section is split into four subsections. The first subsection describes the database and its composition (2.1). The subsection *Material properties and fatigue behavior* (2.2) introduces the fatigue strength influencing factors in steel components, which we seek to analyze with the SHAP values in Section 3. Based on this, the following subsection 2.3 explains the selection of the relevant features. The last subsection 2.4 covers the applied *ML* techniques and SHAP methodology.

2.1 | Database

The dataset utilized for training data-driven approaches comprises a combination of Bosch's internal datasets and publicly available open-access datasets. These primary sources of data are as follows:

- National Institute for Materials Science (NIMS)[24]
- Datenbank und Auswertesystem Betriebsfestigkeit (DaBef)[25] (Database for fatigue strength)
- Robert Bosch internal database[26]

The gathered database encompasses 1250 S-N curve experiments conducted on roughly 30,000 specimens unevenly distributed among 58 different types of steel, all tested under varying conditions. The database consists of carbon, CrMo, CrNiMo, Mn, carburized, stainless, CrNi, Cr, and spring steels, where NIMS is the primary data source, providing around 90% of database entries. All steels experienced only a single tempering step. Each S-N curve experiment is extensively detailed within this tabulated database with up to 70 distinct attributes, including information on chemical composition, processing methods, forming techniques, heat treatment specifics, finishing details, static test results, applied loads, and geometric parameters during testing.

2.2 | Material properties and fatigue behavior

Several factors influence the material fatigue strength of a specimen or component. The influencing factors can be grouped into the material's chemical composition,

process parameters, loading conditions, the component's geometry, and environmental conditions. As this work focuses on steel materials, the process parameters are reduced to heat treatment steps, surface finishing, and surface treatment. Aside from the material's inherent characteristics, the applied load and component geometry are essential for reliable component design. The environmental conditions, such as temperature and corrosion, significantly affect the fatigue properties in the components application[27–31] but are neglected in this study as they are not represented in the database.

2.2.1 | Influence of tensile strength and hardness

The chemical composition and heat treatment determine the material's microstructure and, thus, the macroscopic mechanical properties of metallic materials. Higher tensile strength is typically accompanied by higher fatigue strength. The necessity of the steel's purity increases with rising tensile strength to avoid nonmetallic inclusions that can cause volume failure. Thus, the correlation is linear until approximately 1200 MPa tensile strength, where the predominant fatigue damage mechanism changes from surface to volume-induced failure.[32] Different strengthening mechanisms exist to control the tensile strength of metallic materials, especially solid-solution strengthening, Hall–Petch strengthening, and strengthening through martensitic transformation are essential mechanisms for increasing the tensile strength.[33] These strengthening mechanisms are typically induced by alloying and thermomechanical treatments with nuanced process control. The Hollomon–Jaffe parameter[34]

$$H_p = \frac{273.15 + T_{temp}}{1000} \cdot (Const + log_{10}(T_{time})) \qquad (1)$$

is a phenomenological model used in this work to relate the effect of a diffusion-controlled heat treatment step with the material's hardness. The Hollomon–Jaffe equation is based on the Arrhenius equation that describes the kinetics of diffusion processes within a heat treatment step.[34] Different time and temperature combinations can result in the same Hollomon–Jaffe parameter determining a certain hardness. Thus, the Hollomon–Jaffe parameter H_p determines the effect of tempering on the hardness as a function of tempering time and temperature, where T_{temp} is the tempering temperature in °C and T_{time} is the tempering time in minutes. *Const* is a dimensionless constant that depends on the steel type. Fritz[35] showed that the Hollomon–Jaffe parameter could also be applied to the austenitization heat treatment step by replacing the tempering temperature and time with the corresponding austenitization temperature and time since austenitization is also a diffusion-controlled heat treatment step[34,35]. Thus, this physics-motivated phenomenological equation can estimate the effects of the austenitization and tempering on the hardness through a single value for each process step.

2.2.2 | Influence of stress concentration

Any localized stress concentration can be the source of crack nucleation in a fatigue application. The stress concentration can be caused by notches in the component design or microstructural/process-induced characteristics, such as elastic and plastic incompatibilities at grain boundaries, increased surface roughness through slip bands, scratches, nonmetallic inclusions, and segregations[32,36–39].

2.2.3 | Influence of specimen size

Gudehus and Zenner[28] show that smaller loaded volumes generally increase fatigue strength. We use the $V90_{Mises}$ parameter as a feature to determine the effect of the specimen size. $V90_{Mises}$ is the component volume in which 90% of the maximum Von-Mises stress is exceeded. Kloos[40] distinguishes four aspects of how the component size affects the fatigue properties. These are explained in the following. The first influencing effect is the technology factor, which describes all specimen size-dependent changes in the material condition through material manufacturing processes, such as heat treatments. The surface factor characterizes the limited hardening depth for case-hardened steels. The stress or geometry factor describes the material support of high-stress gradients due to stress reduction at notches through plastic deformation. Thus, fewer large areas have to withstand the maximum stress in smaller components than in the case of a flatter stress gradient or homogeneous loading, which results in higher fatigue strength for smaller components. Lastly, the statistical chance of encountering a nonmetallic inclusion critical to fatigue increases with higher loaded volumes, effectively decreasing the fatigue strength[28,40]. The effect of stress and geometry factors on fatigue strength becomes even more complex under different loading conditions.

2.2.4 | Influence of loading type and stress ratio

Bending applies an inhomogeneous loading where, from a macroscopic point of view, the load diminishes towards

the neutral plane, which inhibits crack growth and reduces the highly loaded volume. In contrast, axial loading exhibits a uniaxial homogeneous load, resulting in a lower fatigue strength comparatively[28,41]. Torsional loading is known to cause the lowest fatigue strength compared to the other loading types. This can be ascribed to the induced multiaxial shear stress state, which alters dislocation dynamics as well as crack initiation and growth mechanisms compared to normal stresses.[42] The relationship between the different loading types has been quantitatively investigated. According to Radaj and Vormwald,[41] the relation between shear and normal stress load for the fatigue limit for steel materials at a stress ratio $R_{ratio} = -1$ is given by

$$\tau_w = (0.55 - 0.6)\sigma_w. \tag{2}$$

This relation has been shown to approximate data by other sources[29,43,44]. Generally, tensile stresses have a negative influence, and compressive stresses positively affect fatigue strength.

2.2.5 | Influence of testing frequency

A variety of influencing factors and different frequency-induced secondary factors have been discussed in the literature, such as time-dependent effects of dislocation movement or diffusion mechanisms (corrosion, oxidation), temperature development at ultrahigh frequencies which can lead to thermal-activated dislocation movement, and testing conditions[28,45–51]. These factors are typically superimposed, and the influence of each factor varies depending on the material condition. As a result, universally valid conclusions about cause–effect relationships cannot be made. Furthermore, different testing conditions, such as the type of testing machines, the frequency, load control type, and the applied load signal, have been shown to influence the resulting fatigue strength significantly[28,29,49,52].

All the above factors influence the fatigue strength, and the effect can vary depending on the material, the manufacturing process, and the component design.

2.3 | Feature selection

In this work, various distinct features are chosen from the aforementioned database for fatigue strength $Sd_{50\%,2E7}$ prediction and interpretability analysis as listed in Table 1, based on the above-mentioned factors influencing fatigue lifetime. These features were chosen based on the availability in the database, selected features in other publications[7,11,53,54], and the generally known influencing factors on the fatigue strength[29,30,41]. For further analysis, three feature sets are composed of specific feature categories. Each category contains different information throughout the processing chain. For instance, *Feature Set 1* contains all features that are consistently available in the database. In contrast, *Feature Set 2* ignores the chemical composition and the heat treatment parameters. *Feature Set 3* includes the chemical composition and heat treatment parameters but leaves out the mechanical properties.

2.4 | ML

Breiman[55] proposed the RF that has become a popular ML method. It is a tree-based ensemble approach that is simple to train but can still model complex regression and classification tasks.[56] Borisov et al[57] and Grinsztajn et al[58] showed that ML methods, such as the RF, outperform deep learning methods for supervised regression learning of heterogeneous tabular data regarding the RMSE. Therefore, the RF is a simple, fast-to-train benchmark model for XAI methods.

SHAP trains a surrogate interpretable model, such as a linear regression, a logistic regression, or a decision tree model, to approximate the uninterpretable black box model locally. To achieve that goal, the original model input vector $\mathbf{x}_i$ is sampled to create a new input vector $\mathbf{x}_i'$ from which the black box model and the surrogate model make a prediction. A loss function assesses the surrogate model's prediction compared with the black box model. This loss function is optimized concerning the surrogate model parameters ϕ to approximate the original model locally. In its simplest form, a linear regression model g is used as a surrogate model to explain the prediction of a ML model f with model input $\mathbf{x}_i$,

$$g(\boldsymbol{z}') = \phi_0 + \sum_{j=1}^{M_{Feat}} \phi_j \boldsymbol{z}_j', \tag{3}$$

where M_{Feat} is the number of features, $\boldsymbol{z}'$ is a sampled coalition subset used as surrogate model input with $\boldsymbol{z}' \in \{0,1\}^{M_{Feat}}$ and $\phi_j \in \mathbb{R}$. An entry of 0 in $\boldsymbol{z}'$ means that the feature is absent in $\mathbf{x}_i'$ and an entry of 1 corresponds to the present/original feature. However, the ML model f cannot handle the binary values in $\boldsymbol{z}'$ as they do not represent valid ML model inputs. Thus, it requires a mapping function $h_x(\boldsymbol{z}') : \{0,1\}^{M_{Feat}} \rightarrow \mathbb{R}^{M_{Feat}}$ that maps 1's in $\boldsymbol{z}'$ to the original feature value and takes the feature's mean value across the entire dataset for 0's.

TABLE 1 Feature descriptions for the selected Features and the feature sets utilized for further analysis.

Category	Description	Feature symbol	Data type	Feature set 1	Feature set 2	Feature set 3
Chemical composition	wt% of carbon	C	Numeric	✓		✓
	wt% of silicon	Si	Numeric	✓		✓
	wt% of manganese	Mn	Numeric	✓		✓
	wt% of phosphorus	P	Numeric	✓		✓
	wt% of sulfur	S	Numeric	✓		✓
	wt% of chromium	Cr	Numeric	✓		✓
	wt% of nickel	Ni	Numeric	✓		✓
	wt% of molybdenum	Mo	Numeric	✓		✓
Heat treatment	Normalizing temperature in °C	Ntemp	Numeric	✓		✓
	Carburization temperature in °C	Ctemp	Numeric	✓		✓
	Carburization time in *min*	Ctime	Numeric	✓		✓
	H_P for hardening (Equation (1)) in Kh	$H_{P,Hard}$	Numeric	✓		✓
	Cooling rate for through hardening °C/min	THQCr	Numeric	✓		✓
	Temperature of the cooling medium in °C	QmTemp	Numeric	✓		✓
	H_P for tempering (Equation (1)) in Kh	$H_{P,Temp}$	Numeric	✓		✓
	Cooling rate for tempering °C/min	TCr	Numeric	✓		✓
Mechanical properties	0.2% proof stress in MPa	Rp02	Numeric	✓	✓	
	Tensile strength in MPa	R_m	Numeric	✓	✓	
	Elongation at fracture in %	A5	Numeric	✓	✓	
	Vickers hardness HV20 on surface	HV20	Numeric	✓	✓	
Component design, load, and testing	90% of Maximum loaded volume according to Wächter et al[44] in mm^3	$V90_{Mises}$	Numeric	✓	✓	✓
	Stress concentration factor	K_t	Numeric	✓	✓	✓
	Loading type	L_{Type}	Categorical	✓	✓	✓
	Stress ratio	R_{ratio}	Numeric	✓	✓	✓
	Surface roughness in µm	Rz	Numeric	✓	✓	✓
	Testing frequency in Hz	f_{test}	Numeric	✓	✓	✓

In contrast to LIME, SHAP weighs the prediction for an instance $\boldsymbol{z}'$ with a weighting kernel of the corrupted/permutated instance $\mathbf{x}_i'$ with

$$\boldsymbol{\pi}_{\mathbf{x}'}(\boldsymbol{z}') = \frac{(M_{Feat}-1)}{(M_{Feat}\, choose |\boldsymbol{z}'|)|\boldsymbol{z}'|(M_{Feat}-|\boldsymbol{z}'|)}, \tag{4}$$

to relate how close the instance $\boldsymbol{z}'$ is to the original model input $\mathbf{x}_i$ since the entries of $\boldsymbol{z}'$ are 1's for the original feature values and 0's for absent feature values, respectively. The number of original features in $\boldsymbol{z}'$ is denoted by $|\boldsymbol{z}'|$. Lundberg and Lee[21] showed that the coefficients ϕ correspond to the SHAP values when utilizing this weight kernel while minimizing the loss function. The loss function $\mathrm{L}_{SHAP}(\mathrm{f},\mathrm{g},\boldsymbol{\pi}_{\mathbf{x}'})$ of the weighted linear model

$$\mathrm{L}_{SHAP}(\mathrm{f},\mathrm{g},\boldsymbol{\pi}_{\mathbf{x}'}) = \sum_{\boldsymbol{z}' \in M_{Data}} [\mathrm{f}(h_x(\boldsymbol{z}')) - \mathrm{g}(\boldsymbol{z}')]^2 \boldsymbol{\pi}_{\mathbf{x}'}(\boldsymbol{z}') \tag{5}$$

is minimized concerning the coefficients ϕ. The loss function is the sum of all sampled modified feature subsets $\boldsymbol{z}'$ in the dataset M_{Data}. The resulting coefficients ϕ_j after minimizing the loss function of the linear regression models $\mathrm{g}(\boldsymbol{z}')$ are the SHAP values. Thus, when explaining the scalar-valued prediction $\hat{\mathrm{y}}_i \in \mathbb{R}$ of a single dataset entry, vector $\mathbf{x}_i$ is now given by the local accuracy property

$$\hat{\mathrm{y}}_i = \mathrm{f}(\mathbf{x}_i) = \phi_0 + \sum_{j=1}^{M_{Feat}} \phi_j \tag{6}$$

of the SHAP values. A consequence of the chosen weight kernel in Equation (4) is that $\phi_0 = E[f(\mathbf{x})]$ represents the expected model output for all predictions of the given dataset and features. ϕ_0 can be interpreted as taking the mean of each feature to make a prediction. The prediction $\hat{y}_i$ of a single model input vector $\mathbf{x}_i$ of the black box model f becomes now explainable since the prediction was locally approximated by a linear regression model who's coefficients are interpretable. Therefore, the SHAP values describe the average contribution of a feature to the prediction among different feature coalitions. A global model interpretation can be achieved by aggregating the SHAP values of multiple single predictions to derive an overview of the average contribution of a feature value in different predictions.[21,22]

2.5 | Data preprocessing, ML training, and SHAP values estimation

Database entries were removed if any feature in Table 1 was missing, resulting in a remaining dataset size (M_{Data}) of 853 entries. Figure 1 illustrates the distribution of Vickers hardness at the surface. The database captures a wide range of low- to high-strength steels, with the NIMS database providing almost 90% of the dataset entries.

The raw experimental data were fitted using the maximum-likelihood approach, following the methodology described in Köder[59] on fatigue strength verification to determine the fatigue strength for a 50% failure probability at $2 * 10^7$ cycles, referred to as $Sd_{50\%,2E7}$. It is important to note that specimens enduring beyond $2 * 10^7$ cycles without failure are considered run-out specimens in all collected databases[24–26]. Consequently, the ultrahigh-cycle fatigue regime remains outside the scope of this analysis. The Hollomon–Jaffe parameter is calculated for the austenitization and tempering heat treatment step to aggregate the effect of temperature and time to single values. This simplifies the feature importance analysis for the heat treatment since the temperature and time cannot be analyzed independently as both characterize the microstructural changes. The *Const* in Equation (1) was chosen 20 and 30 for low and high alloyed steels, respectively.[34] Steels with an accumulated alloying element concentration greater than 5 wt% are considered as high alloyed steels while providing a concentration less than 5 wt% are labeled as low alloyed steels. This simplified binary classification is used for a proof of concept and has been used by other publications[60,61], as it requires experimental effort to reliably determine the *Const*.[62]

The dataset is stratified into 10 equal-sized partitions, where each partition includes the same portion of steel types. From the ten partitions, nine are used as a training dataset and one for testing the model's performance on an unseen testing dataset to predict the fatigue strength $Sd_{50\%,2E7}$. This procedure is repeated 10 times, where each partition constitutes the testing dataset once. This strategy is generally known as cross-validation. The features were directly processed without further normalization or standardization. The RF regressor from scikit-learn[63] was used with default hyperparameter values, which are listed in Appendix A (Table A1). Its performance on the testing datasets concerning the RMSE and the R^2-Score is calculated to evaluate the model's performance. Subsequently, the SHAP values for each feature for every test data instance are computed by minimizing the loss in Equation (5). The cross-validation procedure permits extracting a SHAP value for each instance and each feature in the dataset at testing time. This results in a SHAP value matrix with the same size as the entire database ($M_{Data} \times M_{Feat}$). The SHAP values were determined by the SHAP Python package.[64] The prediction of each instance can be calculated by summing up each SHAP value across the feature dimension, as described in Equation (6).

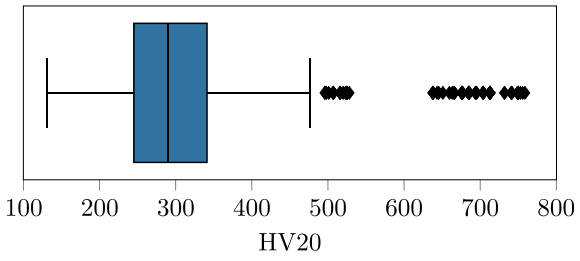

FIGURE 1 Distribution for Vickers hardness (HV20) at the surface for the steel materials in the database, including marks for the median, the 25%, and 75% quartiles. The whiskers are located at 1.5× the interquartile range. Database entries exceeding these boundaries are marked as diamonds. [Colour figure can be viewed at wileyonlinelibrary.com]

3 | RESULTS AND DISCUSSION

The results section is partitioned into four subsections. The first subsection *Performance Comparison of Three Feature Sets* (3.1) compares the RF performance for the three feature sets from Table 1. The second subsection *Feature Set Choice for Feature Importance Analysis* (3.2) discusses the choice of one of the three feature sets for further SHAP value analysis. The following subsections 3.3–3.5 address the capability of the SHAP values to identify physical and engineering-based mechanisms and characteristics on different material scales. Lastly, subsection 3.6 discusses the implication of an ideal XAI

method for data-driven fatigue applications considering the SHAP value results from subsections 3.3–3.5.

3.1 | Performance comparison of three feature sets

Table 2 shows the performance of the RF for the three feature sets from Table 1. The performance values for the RMSE and R^2-Scores are symbolized with the mean and standard deviation of the 10-fold cross-validation.

The performance for all three feature sets for the RMSE, as well as the R^2-Score, is comparable with the results of other publications predicting the fatigue strength[7,12,65]. However, this work utilizes an extended database, additionally considering component design features such as different load and specimen geometry parameters. These additional features are required in the component design process to reliably assess the design against fatigue failure, as in industrial applications, different component geometries and loads appear. Such modeling prospectively could allow for enhancing the data quantity by joining materials and component-level testing data for fatigue analyses. Despite this increase in complexity by additionally considering different loads and geometries, we still achieve high prediction fidelity. When all available features are considered in the RF training (*Feature Set 1*), the best RMSE and R^2-scores are achieved. Specifically, the RMSE reduces by 5 MPa, and the R^2-score increases marginally (see Table 2). Thus, considering chemical composition, heat treatment, and quasi-static mechanical properties together in *Feature Set 1*, even though a causal relation exists between these features, enhances the predictive quality significantly. Slightly worse predictive quality is attained using both reduced feature sets (*Feature Set 2* and *Feature Set 3*). Furthermore, those two models score similarly despite being trained with different features.

In principle, these results are plausible from a physical point of view since there is a strong linkage between the chemical composition and the heat treatment parameters with the mechanical properties. However, despite their cross-correlations, all features carry some unique information that should be preserved. The microstructure and, by extension, the component mechanical properties in *Feature Set 2* are an amalgamation of the whole manufacturing process history and thus encode corresponding information to some degree. In principle, this even entails processes that are not represented through the heat treatment features, for example, machining processes that affect residual mechanical stresses. Furthermore, the mechanical properties are not subject to error propagation through the whole process chain and are related (and correlated) to fatigue properties in a rather direct fashion[32,38]. In contrast, the long-range relations between fatigue properties and the nominal values of chemical composition/heat treatment are governed by intricate mechanisms that might be difficult to learn comprehensively with the given training signal and data distribution. While the quasi-static macromechanical material behavior is relevant to fatigue damage (e.g., plasticity onset), fatigue failure is a more complex and microstructure-sensitive phenomenon. Thus, solely considering the quasi-static mechanical properties is insufficient. In this regard, *Feature Set 3* provides a remedy through the chemical composition and heat treatment parameters that determine the microstructure. Thus, indirectly, information about the microstructural properties is provided to the model, which is critical to fatigue damage formation. We assume that the higher performance of *Feature Set 1* is ascribed to the fact that the model can exploit these feature interactions and complementary multiscale information. It is noted that solely considering the component design, load, and testing features is not sufficient for predicting the fatigue strength as these features do not hold any information about the tested material. Thus, supplementing at least one of the feature subsets chemical composition, heat treatment, and mechanical properties is crucial to cover the material-induced variance.

3.2 | Feature set choice for feature importance analysis

Combining the heat treatment parameters and the mechanical properties improves the model's performance. However, the SHAP values on this combined feature set can be deceiving and lead to inappropriate

TABLE 2 Performance comparison of the three feature sets represented with the mean prediction and standard deviation of the 10-fold cross-validation for the RMSE and R^2-score.

	Feature set 1	Feature set 2	Feature set 3
RMSE in MPa	29.00 ± 4.29	34.26 ± 5.03	35.18 ± 4.75
R^2-score	0.97 ± 0.010	0.96 ± 0.013	0.96 ± 0.014

Abbreviation: RMSE, root-mean-squared error.

feature importance results. When analyzing the SHAP values for *Feature Set 1*, features related to the mechanical properties dominated the outcome due to their strong correlation with the target variable ruling out the feature importance analysis of other features. Under those circumstances, the effect of alloying elements and the heat treatment parameters are nonidentifiable, as will be further discussed in the heat treatment section (3.4). Moreover, the feature independence assumption from the SHAP value theory was significantly harmed due to the mechanical property's dependence on the heat treatment parameters, resulting in nonreliable feature importance. The layer-wise relevance propagation method (LRP)[66–68] or integrated-gradients (IG) method[69] could be alternative methods for understanding the model's behavior, as it does not rely on the feature independence assumption. However, both methods apply only to neural networks, which require more modeling effort and data science knowledge due to the increased choice of hyperparameters to receive a similar model performance.[57] While these methods can provide importance rankings and qualitative feature effects on target variables, they do not facilitate quantitative decomposition of feature contributions[66–69]. In this work, the feature importance analysis is conducted on *Feature Set 3* rather than *Feature Set 2*, as we want to explore whether models can deduce realistic relations under the challenging conditions of (i) sparse data population, (ii) mapping alloying, and processing characteristics directly to fatigue properties, that is, without considering mechanical property features, which hold information about the microstructural state and are strongly correlated to fatigue strength, as an intermediate support. *Feature Set 3* uses the chemical composition and the heat treatment parameters, which are also not independent as cross-correlations between these features exist. For instance, different chemical elements stabilize certain phases and can also significantly influence the starting temperature of the martensitic transformation, which in turn affects the volume fraction of martensite and, therefore, the steel's hardness and fatigue properties. Despite these dependencies, the regressor based solely on chemical composition and heat treatment parameters falls slightly short of full features (see Table 2), rendering it a feasible choice for interpretability analysis.

3.3 | Feature importance analysis for component design, load, and testing features

Figure 2 illustrates the feature partial dependence plots of different component designs and testing features. The ordinate marks the SHAP value in MPa of a given feature with its corresponding feature value on the abscissa. The loading type of the specific dataset entry colors the points. A single subplot illustrates the SHAP value distribution of all dataset entries for a particular feature. We opted for this since the model performance for *Feature Set 3* in Table 2 shows minor standard deviations for the RMSE and R^2-score. This indicates that the model performance is independent of the training and testing dataset sampling, and the RF achieves high predictive performance for all datasets. Thus, all SHAP values can be illustrated aggregated to investigate the learned relations and the impact of a specific feature on the output globally. Higher SHAP values indicate a rise in fatigue strength, while lower values are associated with reduced fatigue strength. Typically, when feature importance studies are concerned, relative feature importance is considered where features are ranked against each other. One potential ranking criterion is the range between a feature's minimum and maximum SHAP value (SHAP range). This criterion is used to arrange the subplots in Figure 2 according to their relative feature importance. We first describe and study all subplots in Figure 2 qualitatively, followed by quantitative analysis.

Figure 2A illustrates the influence of the specimen size ($V90_{Mises}$) on the fatigue strength. The transition area for the change of SHAP values is sharp and could be approximated with a horizontal and vertical line with an angle of 90°. This curve progression stands out among all SHAP distribution plots. The shape could be associated with the concept of representative volumes. The likelihood of encountering a fatigue-critical defect increases if the highly loaded volumes exceed a specific size threshold. The volume threshold could be characteristic of engineering alloys and degrees of purity typically encountered. In fatigue loading, the sensitivity to microstructural extremes is especially pronounced, which could cause this sharp shape. Due to the specimen size range in the dataset, we exclude the possibility of a mechanistic change, such as surface-dominated dislocation dynamics commonly observed in thin films. Typically, standardized specimens are tested, reducing the database variability. Expanding the database with more entries of different specimen sizes could be beneficial to clarify the effect of different $V90_{Mises}$ on the fatigue strength.

The SHAP values indicate rising fatigue strength with increasing frequency f_{test} in Figure 2B. This trend has been observed for many materials and summarized in a survey by Hong et al.[70] The increase in fatigue strength can be attributed to the inability of dislocation motion to follow the high strain rates imposed during high-frequency loading for most steel components despite existing contrary effects, such as thermal-induced dislocation movement[45,48]. Additionally, the frequency

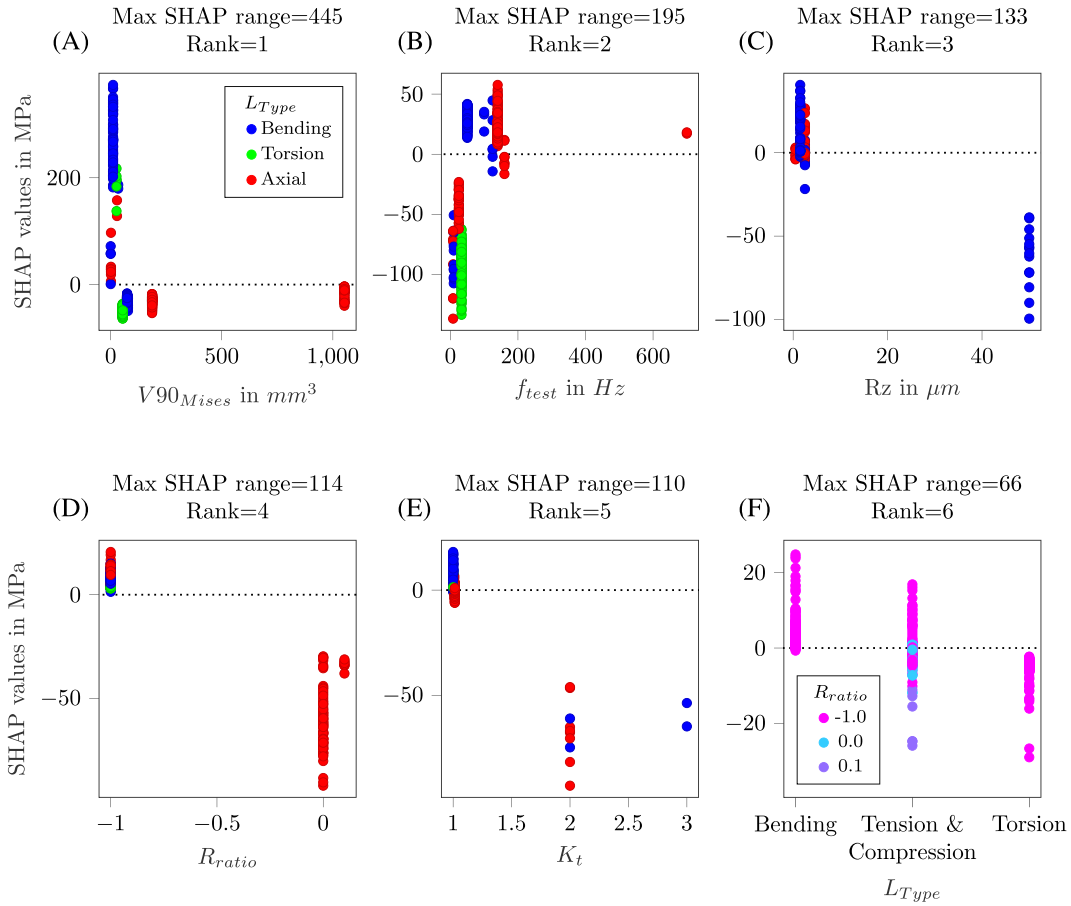

FIGURE 2 (A–F) Influence of the component design, load & testing features on the fatigue strength based on the SHAP values for *Feature Set 3* (see Table 1). Maximum SHAP ranges are reported in MPa and utilized to rank the features against each other. All subplots except of subplot F are colored according to the legend in subplot A. The dotted line is the zero reference line. [Colour figure can be viewed at wileyonlinelibrary.com]

sensitivity is higher for lower frequencies as indicated by higher scatter in the low-frequency regime, which has also been found by Takeuchi et al.[71] The database consists of a few different surface roughness Rz where a reduction of the fatigue strength can be observed for higher surface roughness values in Figure 2C. An increased surface roughness Rz constitutes microscopic notches at which stresses concentrate and fatigue damage can occur. Thus, higher surface roughness reduces the fatigue strength, which fits the trend learned by the RF model and can be explained with the SHAP values.

The database only contains three different stress ratios R_{ratio}, which are displayed in Figure 2D. Still, fully reversed loading ($R_{ratio} = -1$) shows a higher fatigue strength compared to tension–tension settings which are aligned with theory, where alternating cyclical loads with increasing portion of tension stresses open cracks and permit crack propagation which in turn reduces the fatigue strength.

The decreasing effect on the fatigue strength from higher stress concentration factors K_t can be observed in Figure 2E. Even though only a few notched specimens exist in the dataset, the decreasing fatigue strength effect can be observed. Stress concentrates at notches, causing fatigue crack initiation and failure.

Lastly, Figure 2F illustrates the influence of different loading types L_{Type} on the fatigue strength. We anticipate reducing fatigue strength, that is, gradually decreasing SHAP values when changing the loading type from rotation bending to axial and particularly torsion. The SHAP values mirror this behavior.

We conclude that the SHAP values can qualitatively reflect the influence of the component design, load, and testing features on the fatigue strength in all subplots of Figure 2. In principle, the SHAP values can extract the quantitative contribution toward a prediction.[21] The SHAP values for each feature of a dataset entry can be added sequentially to achieve the overall fatigue strength prediction as shown in Equation (6). Thus, the SHAP value on the ordinate is supposedly directly linked to the change in the fatigue strength. The validity of quantitative SHAP value analysis for our task will be

examined based on an example in the following. The lowest fatigue strength in the database for the failure probability of one part per million at 2×10^7 cycles corresponds to $\sigma_{a,10^{-4}\%,2E7} = 951$ MPa. When changing the load from bending to torsion for $R_{ratio} = -1$, the anticipated fatigue strength reduction will be between 523 and 570 MPa according to Equation (2). The maximum SHAP value reduction of the fatigue strength for different loading types for an $R_{ratio} = -1$ is the entire range from the highest bending to the lowest torsion SHAP value, which is approximately 50 MPa and thus one magnitude smaller than the reduction of 523 MPa according to the result from Equation (6). This example shows that a quantitative analysis of the SHAP value does not correctly represent the fatigue strength reduction for different loading types. This observation can have multiple reasons. SHAP measures the feature's importance to the trained model rather than the real-world problem. However, the underlying dataset covers a range of distinct materials that are difficult to represent collectively through a single model. At the same time, publicly available fatigue data are scarce, do not capture the entire feature space, and are afflicted with data bias since certain materials and standard geometries are more often available. Furthermore, the feature independence assumption is violated to some degree in our setting where, for instance, the loading type is correlated with $V90_{Mises}$. Thus, hereafter, we will analyze the influence of the material and heat treatment parameters qualitatively only, without relying on quantitative analysis. The quantitative analysis, in particular, needs to be performed with great care and demands domain expertise to avoid unrealistic interpretations. Consequently, the ranking of the features according to their relative importance in Figure 2 should also be treated with care, as the quantitative analysis is not reliable.

3.4 | Feature importance analysis of heat treatment features

The SHAP values for the heat treatment parameters are aggregated the same way as the SHAP values for the component design parameters in Figure 2. However, the carburized steels are excluded from this heat treatment analysis as the carburization process step, and the follow-up heat treatment steps of hardening and tempering significantly alter the microstructure at the surface. Thus, the microstructural changes of carburized and non-carburized steels should not be analyzed together. At the same time, analyzing the carburized steels separately is not feasible since the database only consists of 80 carburized steel entries (solely case-hardening experiments with few standardized geometries), which does not provide the basis for drawing statistically sound conclusions. Each SHAP value in Figure 2 is colored according to the corresponding specimen size ($V90_{Mises}$) of that database entry since the component size influences the time and temperature of each process step. The larger the specimen size, the longer usually a soaking time has to be to achieve the desired results across the whole cross-section, which is why the Hollomon–Jaffe parameter is a useful feature. The same image with binary coloring for distinguishing whether the heat treatment occurred is illustrated in Figure C3 in the appendix. Generally, the hardening process is restricted to martensitic transformations.

The heat treatment process starts with the materials' normalization. Normalizing is an annealing step to homogenize the microstructure, often contributing to a higher fatigue strength of materials exhibiting pronounced segregation. The normalization time is not included in Figure 3A as there is only a single time value if normalization is conducted. This information does not provide additional information to the *RF* model when the temperature is already supplied. The SHAP values indicate that conducting no normalization systematically reduces the fatigue strength, while in the case of normalization, the fatigue strength can increase and decrease with high variation. If normalization occurs, the fatigue strength increases with the normalization temperature until 1000°C. Therefore, data scientists could conclude that higher normalization temperatures result in higher fatigue strength. In contrast, engineers who are familiar with materials processing might know that it is not primarily the normalization temperature that gives rise to the higher fatigue strength but that there is rather a spurious correlation at play, which is caused by the different alloying of the steels. Higher alloyed steels tend to have higher fatigue strength and typically require higher austenitization temperatures.[31] Thus, the higher fatigue strength is not attributed to the normalization temperature but rather to higher alloying. The normalization temperature above 1000°C belongs to the group of stainless steels, which are known to have lower fatigue strength and, therefore, reduced SHAP values. The SHAP values for hardening $H_{P,Hard}$ using the Hollomon–Jaffe parameter in Figure 3B show an increasing behavior until reaching a maximum, followed by a decrease but staying on a high level. $H_{P,Hard}$ describes the austenitization progress. Presumably, the diffusion-based transformation from ferrite to austenite has not been completed, and the remaining ferrite does not transform to martensite for low $H_{P,Hard}$ values. As a result, the hardness decreases. Murakami[32] showed a strong correlation between hardness and fatigue strength. A decreasing hardness and material fatigue strength for too high $H_{P,Hard}$ values could be associated with the disadvantageous

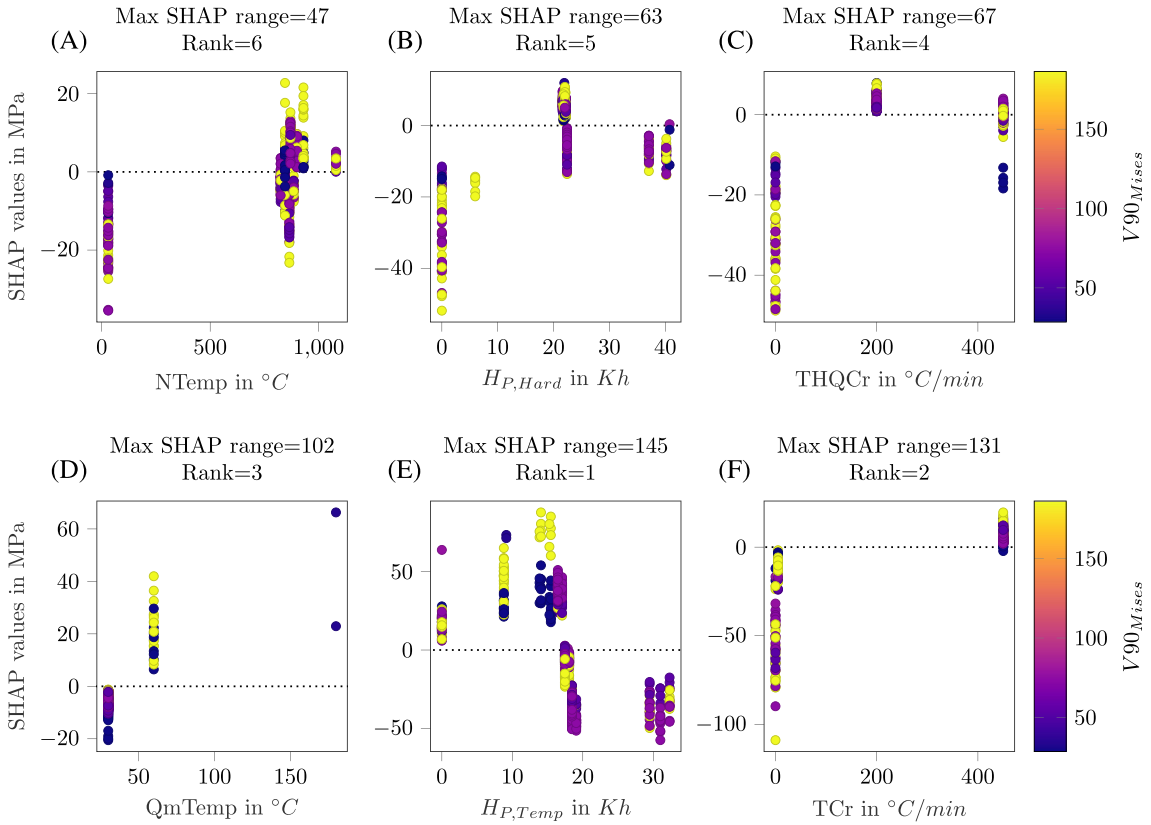

FIGURE 3 (A–F) Influence of each heat treatment step on the fatigue strength depending on the specimen size ($V90_{Mises}$) for *Feature Set 3* (see Table 1). Maximum SHapley Additive exPlanations (SHAP) ranges are reported in MPa and utilized to rank the features against each other. The subplots are ordered according to the process route. The dotted line is the zero reference line. [Colour figure can be viewed at wileyonlinelibrary.com]

coarsening of the martensitic substructures. At the maximum SHAP values, presumably, an optimal fine martensite morphology is achieved.[34,35]

The SHAP values for the hardening cooling rate *THQCr* show a convex-shaped curve progression in Figure 3C. If the cooling rate is too low, the martensitic starting temperature will not be reached without some transformation to ferrite and pearlite. Internal stresses and warping induced by too high-temperature gradients can cause crack initiation and reduce fatigue strength.

The effect of the temperature of the cooling medium *QmTemp* in Figure 3D on the fatigue strength is similar to the hardening cooling rate *THQCr*, which is also indicated by the SHAP values. If the quenching medium is too cold, the thermally induced stresses can initiate cracks, effectively reducing the fatigue strength. Additionally, the temperature of the quenching medium depends on the kind of medium itself. Water is often used at room temperature and generally has the most intensive quenching properties. Quenching oils are usually used at temperatures of 60°C and above, and salt baths are mainly operated at temperatures above 160°C[27,31].

Figure 3E illustrates the SHAP values for tempering $H_{P,Temp}$ step and shows similar behavior as the ones for hardening $H_{P,Hard}$ except for the significant SHAP value drop beyond the maximum for the tempering. The steel remains hard and brittle for too low $H_{P,Temp}$ values, which reduces the fatigue strength, indicated by smaller SHAP values. The steel becomes more ductile and loses hardness with increasing $H_{P,Temp}$ values as carbon diffuses out of the metastable martensite and forms carbides. An optimum for the fatigue strength exists that provides the ideal trade-off between ductility and hardness.[34] Furthermore, for precipitation hardening steels, optimal strengthening is determined by ideal precipitate sizes, spacing, and interface characteristics, which coincide with intermediate processing times. In the underaged state, particles fail to impede dislocation motion as they can be easily sheared. In the overaged state, precipitates are too coarse and dispersed to interact with dislocations effectively.[33]

Lastly, the SHAP values for *TCr* in Figure 3F show increased fatigue strength for higher cooling rates, while low cooling rates reduce fatigue strength. The classes of low cooling rates and no cooling, where annealing was

not conducted, overlap in Figure 3. For a better distinction between treated and nontreated groups, we refer to Figure C3 in the appendix, where the same illustration is colored in a binary fashion. There is a broad scatter for instances in which no tempering is evident. These instances correspond to stainless and carbon steels. Low cooling rates might not stop the carbon diffusion and culminate in reduced material hardness. Thus, high cooling rates enhance mechanical properties adjusted during tempering or annealing.[31]

All subplots in Figure 3 are colored according to the specimen size ($V90_{Mises}$). A significant effect of different sizes on the fatigue strength concerning different heat treatment steps cannot be observed though. Presumably, the difference in specimen size regarding the database is too small to observe a relation between the heat treatment steps and specimen size.

The different heat treatment steps in Figure 3 are ordered chronologically and show increasing importance along the manufacturing process chain. High importance is observed for the features associated with processes that are carried out later such as the quenching speed and the Hollomon–Jaffe parameter for tempering. Indeed, these processes are more directly related to the fatigue properties than those situated at the beginning of the process chain.

We conclude that the SHAP values can explain the effect of each heat treatment step on the fatigue strength for *Feature Set 3*. Thus, the influence of heat treatment parameters can be discerned despite their existing interactions with the chemical composition.

Figure C2 in the appendix illustrates the SHAP value analysis for the heat treatment parameters of *Feature Set 1*. Especially the SHAP values for $NTemp$ and $H_{P,Hard}$ show unrealistic results. Specifically, the plots spuriously imply that conducting normalization reduces the fatigue strength and that it is irrelevant whether austenitization is performed during the hardening process (since many SHAP values are clustered around zero). Evaluators without a materials engineering background could conclude that hardening does not affect fatigue strength. An ideal XAI method is discussed in subsection 3.6.

3.5 | Feature importance analysis of chemical composition features

Lastly, we analyze the SHAP values for the chemical composition to investigate their influence on fatigue strength. Figure 4 provides an overview of the different SHAP distributions. Stainless steels are not covered in this study. The chemical composition of stainless steels differs significantly from the other steel classes since stainless steels are characterized by a considerably higher Cr and Ni percentage. Jointly depicting such high values with the other steels would compromise the assessment of the feature sensitivity. Generally, we did not observe a positive or negative impact on SHAP values for an increased or decreased Cr or Ni concentration for stainless steels. The reason may lie within the unbalanced dataset where stainless steels represent less than 10%. Thus, their effect on the SHAP values might be overshadowed by the other steels. Surface hardening through carburization significantly changes the fatigue properties as described in subsection 3.4. Thus, the effect of different alloying elements can be different depending on whether carburization was conducted or not. Whether the steel is carburized or not is captured through a binary color-coding in Figure 4.

The SHAP values for chromium show increasing fatigue strength with an increasing amount of Cr content in Figure 4A). The slope reduces around 0.5% Cr of chromium but remains positive. Fatigue strength reduction can be observed beyond 1.5% Cr content for carburized steels. In general, chromium reduces the critical cooling rate for martensite formation and increases, the hardening depth. The reduction of fatigue strength for carburized steels for high Cr concentration could be explained by the increased potential of surface oxidation, which causes local stress concentrations[72,73]. The change of slope around 0.5% Cr for noncarburized steels can be explained by different alloying systems. Steel materials are made of specific alloying element combinations and concentrations. Clustered steel groups containing certain concentrations of chromium and molybdenum can be observed in Appendix D.4. Since this information is compounded, independent consideration of the main alloying elements for increasing the hardening depth, Cr, Ni, Mo, and Mn,[31] and their individual contribution to the fatigue strength described by the SHAP values cannot be investigated in detail.

The SHAP values and, thus, the fatigue strength of the steel component increase linear with increasing amount of carbon in Figure 4B for noncaburized steels. The slope for noncarburized steels reduces around 0.4 wt % of carbon but remains positive. The starting carbon concentration for carburized steels, which is plotted, is initially lower compared with noncarburized steels. The SHAP values also increase linearly with the carbon concentration but stop around 0.2 weight percentage of carbon. Carbon is the most critical element as it dictates the diffusionless transformation of austenite into martensite, increasing the steel's hardness.[31] Presumably, the change of slope indicates the change of lath martensite to a mixture of lath and plate martensite, which is assumed to be around 0.5% of carbon[31] for noncarburized steels. The

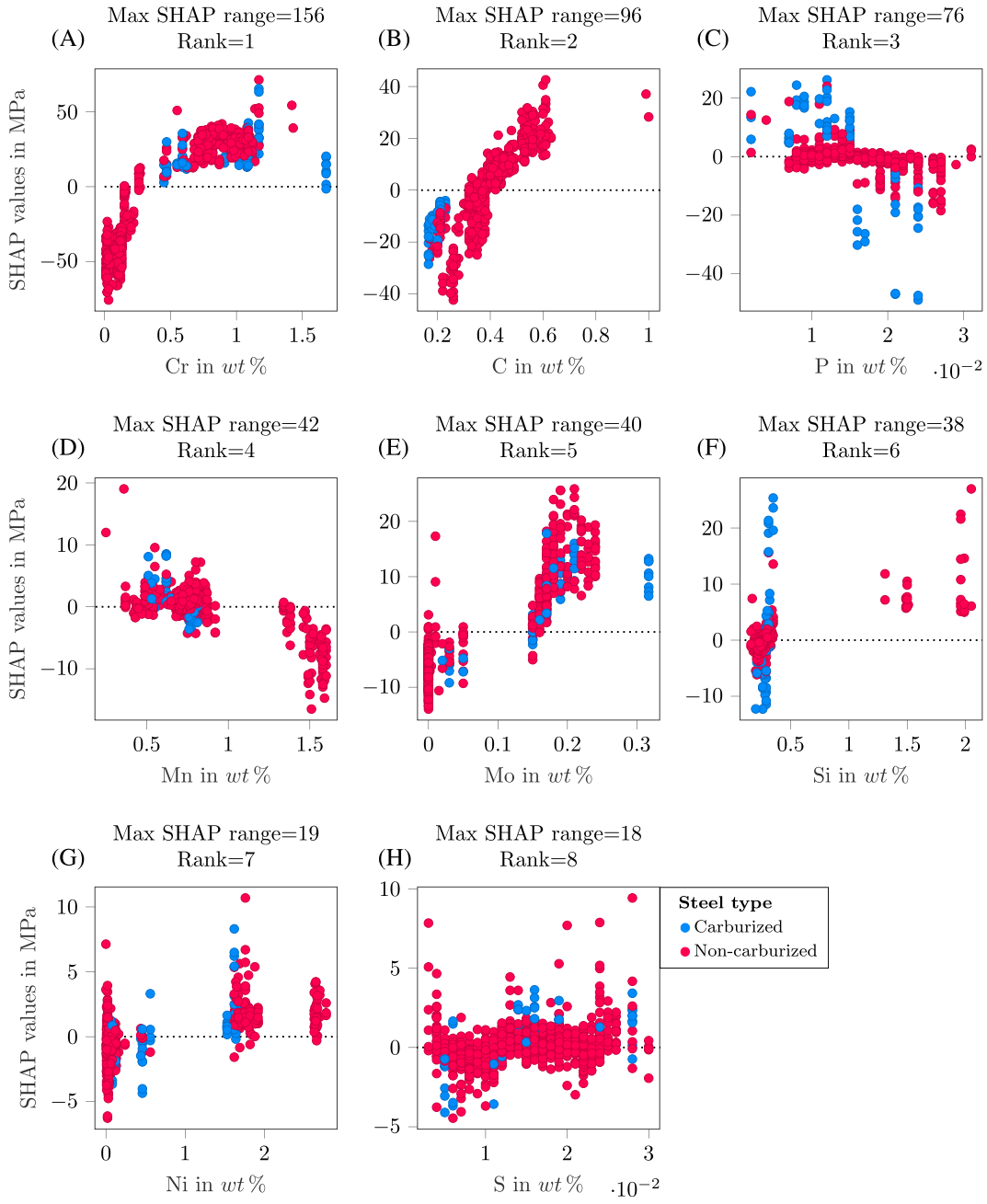

FIGURE 4 (A–H) Influence of each alloying element on the fatigue strength for *Feature Set 3* (see Table 1). Maximum SHAP ranges are reported in MPa and utilized to rank the features against each other. All subplots are colored according to the legend in subplot h) to distinguish between carburized and noncarburized steels. The dotted line is the zero reference line. [Colour figure can be viewed at wileyonlinelibrary.com]

initial carbon concentration for carburized steel is significantly lower to facilitate carbon absorption during the carburization process while maintaining the cores' ductility. The amount of carbon is increased at the surface through the carburization step, resulting in a hard surface layer.[31] Nonetheless, higher initial carbon contents seem beneficial for fatigue strength in the low carbon range of carburized steels (blue).

The fatigue strength reduces significantly with increasing phosphorus concentration for carburized steels in Figure 4C, while there is only a weak trend for some noncarburized steels. Phosphorus contaminates steel components, and a decrease in fatigue strength is expected with rising phosphorus concentration. Phosphorus is known to segregate in steel microstructure, especially at grain boundaries, and reduce the fatigue

strength[31,74]. Information about the underlying microstructure, failure mechanisms, or fracture analysis is unavailable. Thus, it remains speculative why the correlation between phosphorus and the fatigue strength decrease is not stronger for noncarburized steels.

The SHAP values for manganese decrease with increasing Mn concentration in Figure 4D. Manganese Mn increases similar to Ni, Cr, and Mo the hardening depth.[31] Thus, an increase in fatigue strength would be expected with rising Mn concentration, which is not reflected within the SHAP values. However, the effect of manganese is not as strong as for the other alloying elements, which might explain why no significant increase in fatigue strength is observed. The SHAP value decreases with higher manganese concentration, which could be explained by manganese sulfide inclusions, which often act as crack initiation sites.[74] However, a correlation with the amount of sulfur S was not found within the dataset. Also, an embrittlement effect and thus reduced fatigue strength due to diffusion of primary austenite grains for higher temperatures during tempering cannot be reliably detected in the manganese and tempering temperature dependence plot in Appendix D.5.

The SHAP values for molybdenum and nickel increase with rising Mo and Ni concentration in Figure 4E,G, respectively. Both elements increase the hardenability of steel and the hardening depth for carburized steels, which is beneficial for the fatigue strength of both steel types.[31] Further interpretations about different or changing slopes cannot be reliably determined due to the dependency on the alloying system as discussed in Figure 4A.

The SHAP values for silicon increase for some noncarburized steels with higher silicon content in Figure 4F, while most noncarburized steels are clustered at low silicon concentration and SHAP values around zero. In contrast, the SHAP values for carburized steels show a large scatter at low silicon concentrations. Silicon is a solid solution-strengthening element essential for spring steels and increases yield strength.[31] Crack initiation in fatigue is highly correlated with irreversible dislocation movement and persistent slip band formation within grains.[33] Increasing the critical resolved shear stress increases the required stress for dislocation movement, which increases the fatigue strength. Additionally, silicon contributes to form retained austenite (RA) due to its low solubility in cementite[75,76]. RA can close cracks due to transformation-induced plasticity (TRIP) and thus increase the fatigue strength in case of fine dispersion.[77] Interesting SHAP values can be observed for the influence of silicon on carburized steels. Schmiedhofer[78] found that silicon significantly reduces the carbon diffusion depth for carburized steels and thus reduces the fatigue strength. Silicon is also known for promoting surface oxidation by forming silicon dioxide SiO_2. Surface oxides can be incorporated into the bulk during cyclic loading, affecting the fatigue properties due to an increased local stress concentration[73,79]. These aspects favor a reduction of fatigue strength. However, contrary to the expectation, the SHAP values increase for some carburized steels. One reason could be finely dispersed RA that hinders crack growth. Another reason could be that not all process steps are documented within the database. Razim[73] reported that removing the surface oxidated layer can increase the fatigue strength by 20%. Since the information on whether the oxidated layer was removed and microstructural and/or fractural analysis are not available, a conclusive statement about the silicon influence on carburized steels cannot be made.

Lastly, a correlation of the SHAP values with increasing sulfur S concentration cannot be observed in Figure 4H. Even though corresponding empirical evidence exists—especially the interplay with manganese to form manganese sulfide inclusions.[74] Each subplot in Figure 4 was colored depending on the steel class, carburized or noncarburized. A significant difference for both steel classes can only be observed for phosphorus and silicon content. There could be multiple reasons for that: The effect of specific alloying elements on the fatigue strength is affected by the following heat treatment procedure and can also not be investigated completely independently due to the different alloying systems. Furthermore, even from a physical point of view, it is yet not clear whether and how the fatigue damage mechanisms potentially change concerning alloying elements (see discussion on silicon in Figure 4F.

The relative importance analysis shows on one hand high importance of carbon followed by the hardening depth increasing alloying elements such as manganese and molybdenum and great importance for steel contamination like phosphorus and silicon. However, the high importance of chromium cannot be explained here and requires further investigation.

We conclude that the SHAP values can reflect the effect on the fatigue strength for most alloying elements, while especially the high effect of chromium and manganese sulfide inclusions induced failure cannot be observed.

3.6 | Implications of interpretability analysis for materials science and engineering

In our study, some results cannot be explained or observed with the underlying database, such as the not feasible quantitative analysis of the component design, load and testing features, and the contamination effects

of phosphorus/sulfur. Collecting more component design, load, and testing features could resolve this issue. However, these data are usually limited, expensive to generate and often limited to standard geometries which restricts their expressiveness. It is unclear if increasing the database could resolve the problem of the chemical composition features, as an increased concentration of phosphorus and sulfur increases the chance of causing fatigue damage but is not necessarily the source of crack initiation. This could be the reason for not detecting a decrease in SHAP values with increasing *Si* content and only a weak trend for increasing *P* content for noncarburized steels. As additional microstructural/fracture analysis is unavailable, the hypotheses can not be validated conclusively. Microstructure information is often used as a foothold to bridge the large gap between composition/processing and properties as well as to attain interpretability and generalization. The composition-process design space is high-dimensional and exhibits many interactions where even minute changes in the processing pipeline can culminate in notable property changes. On this end, the microstructure allows validation and capturing fluctuations throughout the process chain. This role is also partially fulfilled by the recorded macroscopic properties. Despite the need for microstructural information and more specimen geometry data, we conclude based on the results and discussions from the previous subsections 3.3–3.5 that the SHAP values based on linear regression surrogates and the reduced feature set show physically reasonable trends and presumably can explain the fatigue influencing effects among the different scales. This even holds despite the existing cross-correlations between the alloying elements, the heat treatment parameters, and the component design parameters which harms the SHAP feature independence assumption. The study's experimental design, that is, the utilized broad steel data distribution and the framing of the prediction task, affects the deducible relations to a major extent. Specifically, the setting presented here facilitates the identification of first-order relations, which tend to apply to a wide range of steel materials. It thus is relevant for *materials selection*. When exploration of mechanisms for *materials optimization* through microstructure engineering is the objective, the data range needs to be confined to a specific material, and local microstructure representations could be applied as proposed by Thomas et al.[17] Future work should consider collecting more microstructural, fractural information and specimen geometries to enhance and extend fatigue databases for data-driven causal analysis and potentially increase the expressiveness of XAI methods. The need for XAI methods and their expressiveness beyond standard data exploration tools such as correlation and variance analyses is discussed in the following. While the detailed assessment of these techniques is beyond the scope of this work, we perform a simple comparison between a $Sd_{50\%,2E7}(f_{test})$ scatter plot and the corresponding SHAP plot, which are both provided in the appendix in Figure B1. Notably, the positive impact of higher frequency on the endurance limit, which is ubiquitous in literature[70] and clearly confirmed by SHAP, cannot be easily deduced from the data scatter plot. This behavior holds for many of the influencing factors. It is presumably ascribed to the superposition of many influencing factor contributions, which yet can be effectively decomposed through SHAP. This simple investigation justifies the usage of XAI methods in fatigue analysis. While the SHAP value showed reliable results for most investigated fatigue influencing factors, their limitation due to the feature independence assumption cannot be aligned with the complex fatigue failure process. More advanced XAI methods should be used in the Future. A proposal is described in the following.

Generalizability, scalability, and transparency are desired properties of XAI methods to understand the decision-making of the arbitrary black box models. In materials and engineering applications, a graph representation that holistically and symbolically describes the reasoning path and interdependencies between features and targets would be a desired representation. This could potentially enable validating known relations, capturing unknown patterns, and ultimately inverting the way of materials exploration, development, and design. To date, however, neither those XAI methods nor the required material science data to train such data-driven models exist. New research movements such as causal deep learning[80] propose new XAI methods to address this demand. For instance, feature interactions can be represented through a directed acyclic graph whose adjacency matrix is learned.[81] In terms of data availability, several large-scale materials digitalization initiatives have formed which prospectively will improve data sustainability and thereby fuel such data-driven exploration[82–84].

The aim of acquiring new fatigue data should be to preserve all relevant information. Aside from capturing all defining processes and their metadata, one crucial aspect is information on the sequence in which the processes were performed. This information is neither retrievable from most databases nor can it be considered by the standard ML models. To address both limitations, we advocate a more integrated approach—modeling the process chain data through knowledge graphs, which not only preserves the sequence information but also enables neurosymbolic AI using knowledge graph embeddings. This could not only help to infuse background knowledge but also, by extension, improve the model's range of generalization.

4 | CONCLUSION

This work compared different feature combinations for data-driven predictions of fatigue strength with a RF model. We further analyzed the RF model with the SHAP values, which is one of the most popular XAI methods, to achieve insights into the RF black box model and understand how predictions are made and how each feature contributes to that prediction. Incorporating chemical composition, heat treatment parameters, mechanical properties, and component design parameters achieves the best predictive performance. Collecting more meaningful data, such as hardening depth, residual stresses, inclusion size, and distribution, could improve prediction fidelity and facilitate new insights through interpretability analysis.

Suitable feature selection, where quasi-static mechanical properties were discarded, enabled a meaningful SHAP analysis. The SHAP values represent the effect of macroscopic influencing factors such as the component design parameters and different loading conditions correctly. Furthermore, the SHAP values could identify trends associated with microscopic influencing factors conforming to the literature. For instance, solid-solution strengthening through elements like nickel and molybdenum and different martensite morphologies depending on the carbon concentration is implied. Thus, SHAP can validate and build trust in fatigue strength predictions and identify first-order influencing factors irrespective of size scale.

We conclude that for qualitative assessments, it is not required to strictly fulfill SHAP's feature independence assumption, as the analysis depicts meaningful results despite the cross-correlation of chemical composition and the heat treatment parameters. Nonetheless, the feature independence assumption is a major drawback, and alternative methods that avoid such methodological constraints while comprehensively and symbolically describing the reasoning path and feature interdependencies would be preferable.

Materials science applications, especially fatigue prediction, often exhibit intricate and multiscale feature interactions. Moreover, scarce, heterogeneous, and incomplete data in this domain poses major challenges. These properties render extracting causal relations difficult. Materials engineering domain knowledge is indispensable for the reliable assessment of XAI results to identify spurious correlations, limitations, and advantages. At the same time, a data scientist's expertise is required to select or develop appropriate methods that handle the aforementioned challenges. Therefore, our study underpins the importance of thoughtful integration of domain-specific knowledge and advanced data science techniques to ensure the accurate and reliable application of data-driven approaches in fatigue strength prediction and promoting successful outcomes in engineering applications[85–87].

AUTHOR CONTRIBUTIONS

Christian Frie: Conceptualization; methodology; software; formal analysis; investigation; writing—original draft. **Ali Riza Durmaz:** Writing—original draft. **Chris Eberl:** Supervision; writing—review & editing.

ACKNOWLEDGMENTS

The authors thank Thomas Waldenmaier (Robert Bosch GmbH, Germany) for the fruitful discussions which supported this research. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no potential conflict of interests.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ORCID

Ali Riza Durmaz https://orcid.org/0000-0002-0916-5990

REFERENCES

1. Durmaz AR, Natkowski E, Arnaudov N, Sonnweber-Ribic P, Weihe S, Münstermann S, Eberl C, Gumbsch P. Micromechanical fatigue experiments for validation of microstructure-sensitive fatigue simulation models. *Int J Fatigue*. 2022;160:106824.
2. Hennessey C, Castelluccio GM, McDowell DL. Sensitivity of polycrystal plasticity to slip system kinematic hardening laws for al 7075-t6. *Mater Sci Eng: A*. 2017;687:241-248.
3. McDowell DL, Dunne FPE. Microstructure-sensitive computational modeling of fatigue crack formation. *Int J Fatigue*. 2010;32(9):1521-1542.
4. Natkowski E, Sonnweber-Ribic P, Münstermann S. Determination of fatigue lifetimes with a micromechanical short crack model for the high-strength steel SAE 4150. *Int J Fatigue*. 2022;156:106621.
5. Natkowski E, Sonnweber-Ribic P, Münstermann S. Surface roughness influence in micromechanical fatigue lifetime prediction with crystal plasticity models for steel. *Int J Fatigue*. 2022;159:106792.
6. Schäfer BJ, Sonnweber-Ribic P, Ul Hassan H, Hartmaier A. Micromechanical modelling of the influence of strain ratio on fatigue crack initiation in a martensitic steel-a comparison of different fatigue indicator parameters. *Materials*. 2019;12(18):2852.
7. Agrawal A, Deshpande PD, Cecen A, Basavarsu GP, Choudhary AN, Kalidindi SR. Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters. *Integr Mater Manuf Innovat*. 2014;3:90-108.

8. Chen J, Liu Y. Fatigue modeling using neural networks: a comprehensive review. *Fatigue Fract Eng Mater Struct*. 2022;45(4):945-979.
9. He L, Wang Z, Akebono H, Sugeta A. Machine learning-based predictions of fatigue life and fatigue limit for steels. *J Mater Sci Technol*. 2021;90:9-19. https://linkinghub.elsevier.com/retrieve/pii/S1005030221002607
10. He N, Ouyang R, Qian Q. Learning interpretable descriptors for the fatigue strength of steels. *AIP Adv*. 2021;11(3):35018.
11. Kolyshkin A, Frie C, Froschmeier T. Datenbasierte lebensdauervorhersage. 48 Tagung des DVM-Arbeitskreises Betriebsfestigkeit, https://dvm-wissen.de/gb/betriebsfestigkeit-2022/240-BF-2022-017.html; 2022.
12. Weichert D, Kister A, Houben S, Ernis G, Wrobel S. Robustness in fatigue strength estimation. arXiv preprint arXiv:221201136; 2022.
13. Xiong J, Zhang T, Shi S. Machine learning of mechanical properties of steels. *Sci China Technol Sci*. 2020;63(7):1247-1255.
14. Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. *Neural Netw*. 1989;2(5):359-366.
15. Jansen T, Rozenberg G, Bäck T, Kok JN. Computational complexity of evolutionary algorithms; 2012.
16. Bishop CM, Nasrabadi NM. *Pattern Recognition and Machine Learning*, Vol. 4: Springer; 2006.
17. Thomas A, Durmaz AR, Alam M, Gumbsch P, Sack H, Eberl C. Materials fatigue prediction using graph neural networks on microstructure representations. *Sci Rep*. 2023;13(1):12562.
18. Burkart N, Huber MF. A survey on the explainability of supervised machine learning. *J Artif Intell Res*. 2021;70:245-317.
19. Du M, Liu N, Hu X. Techniques for interpretable machine learning. *Commun ACM*. 2019;63(1):68-77.
20. Ribeiro MT, Singh S, Guestrin C. "why should i trust you?" explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016:1135-1144.
21. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Adv Neural Inform Process Syst*. 2017;30.
22. Shapley LS, et al. A value for n-person games. 1953.
23. Lundberg SM, Erion GG, Lee S-I. Consistent individualized feature attribution for tree ensembles. arXiv preprint arXiv:180203888; 2018.
24. Furuya Y, Nishikawa H, Hirukawa H, Nagashima N, Takeuchi E. Catalogue of NIMS fatigue data sheets. *Sci Technol Adv Mater*. 2019;20(1):1055-1072. https://doi.org/10.1080/14686996.2019.1680574
25. Ellmer F, Esderts A, Eulitz K-G, Hinkelmann K. *Forschungskuratorium maschinenbau: Datenbank und auswertesystem betriebsfestigkeit*: FKM-Vorhaben; 2011.
26. GmbH RB. Robert Bosch internal fatigue database; 2023.
27. Bargel H-J, Schulze G. *Werkstoffkunde*: Springer-Verlag; 2008.
28. Gudehus H, Zenner H. Leitfaden für eine betriebsfestigkeitsrechnung. Verlag Stahleisen GmbH, Düsseldorf; 1999.
29. Haibach E. *Betriebsfestigkeit*: Springer; 2006.
30. Sander M. *Sicherheit und betriebsfestigkeit von maschinen und anlagen*: Springer; 2008.
31. Schlegel J. *Die welt des stahls*: Springer; 2021.
32. Murakami Y. *Metal Fatigue: Effects of Small Defects and Nonmetallic inclusions*, Referex Engineering: Elsevier; 2002. https://books.google.de/books?id=zcJmlAEACAAJ
33. Gottstein G. *Physikalische grundlagen der materialkunde*: Springer-Verlag; 2007.
34. Hollomon JH. Time-temperature relations in tempering steel. *Trans AIM*. 1945;162:223-249.
35. Fritz S. Untersuchungen zum Optimierungspotential von Erwärmzeit und Eigenschaftsprofil von Stählen im Presshärteprozess; 2022.
36. Garwood MF. Interpretation of tests and correlation with service. *ASM*. 1951;1.
37. Meyers MA, Chawla KK. *Mechanical Behavior of Materials*: Cambridge university press; 2008.
38. Pang JC, Li SX, Wang ZG, Zhang ZF. General relation between tensile strength and fatigue strength of metallic materials. *Mater Sci Eng: A*. 2013;564:331-341.
39. Suresh S. *Fatigue of Materials*: Cambridge university press; 1998.
40. Kloos KH. Einfluß des oberflächenzustandes und der probengröße auf die schwingfestigkeitseigenschaften.; 1976.
41. Radaj D, Vormwald M. *Ermüdungsfestigkeit*: Springer; 2007.
42. Billaudeau T, Nadot Y, Bezine G. Multiaxial fatigue limit for defective materials: mechanisms and experiments. *Acta Materialia*. 2004;52(13):3911-3920.
43. ASM Handbook Volume. 19: fatigue and fracture. ASM international. Online Edition, Materials Park, USA; 1996.
44. Wächter M, Müller C, Esderts A. *Angewandter festigkeitsnachweis nach fkm-richtlinie*: Springer; 2017.
45. Awatani J, Katagiri K. Study of ultrasonic fatigue by microscopic observation. *Trans Japan Soc Mech Eng*. 1968;34:833-841.
46. Bayraktar E, Garcias IM, Bathias C. Failure mechanisms of automotive metallic alloys in very high cycle fatigue range. *Int J Fatigue*. 2006;28(11):1590-1602.
47. Heimbach H, Heimbach H. Zum einfluß der belastungsfrequenz auf die zeit-und dauerfestigkeit von stahl/influence of loading frequency on the fatigue strength of steel for finite and infinite life/l influence de la fréquence des contraintes sur la résistance à la fatigue de p acier pour une durée de vie finie et infinie. *Mater Test*. 1970;12(11):377-380.
48. Kikukawa M, Ohji K, Ogura K. Results of tensioncompression fatigue test for mild steel up to 100kc/s. *Trans Japan Soc Mech Eng*. 1966;32:363-370.
49. Leyendecker T. *Untersuchungen zum einfluss des prüfsignals auf die schwingfestigkeit des common-rail-systems*: Shaker; 2009.
50. Tsutsumi N, Murakami Y, Doquet V. Effect of test frequency on fatigue strength of low carbon steel. *Fatigue Fract Eng Mater Struct*. 2009;32(6):473-483.
51. Zhao A, Xie J, Sun C, Lei Z, Hong Y. Effects of strength level and loading frequency on very-high-cycle fatigue behavior for a bearing steel. *Int J Fatigue*. 2012;38:46-56.
52. Eufinger J, Schneider N, Pyttel B, Berger C. Bauteilauslegung unter berücksichtigung von beanspruchungen mit variablen amplituden und sehr hohen schwingspielzahlen: Vorhaben nr. 317. Forschungskuratorium Maschinenbau FKM 292; 2012.
53. He N, Ouyang R, Qian Q. Learning interpretable descriptors for the fatigue strength of steels. *AIP Adv*. 2021;11(3):035018. https://doi.org/10.1063/5.0045561
54. Xiong J, Zhang T, Shi S. Machine learning of mechanical properties of steels. *Sci China Technol Sci*. 2020;63(7):1247-1255.
55. Breiman L. Random forests. *Mach Learn*. 2001;45:5-32.

56. Biau G, Scornet E. A random forest guided tour. *Test*. 2016;25:197-227.
57. Borisov V, Leemann T, Sessler K, Haug J, Pawelczyk M, Kasneci G. Deep neural networks and tabular data: a survey. IEEE Transactions on Neural Networks and Learning Systems, PP; 2021.
58. Grinsztajn L, Oyallon E, Varoquaux G. Why do tree-based models still outperform deep learning on tabular data? 2022.
59. Köder M. Schwingfestigkeitsnachweis für innendruckbelastete common-rail-bauteile aus 100cr6 unter berücksichtigung hochzyklischer betriebsbeanspruchungen; 2014.
60. Canale LCF, Yao X, Gu J, Totten GE. A historical overview of steel tempering parameters. *Int J Microstruct Mater Propert*. 2008;3(4-5):474-525.
61. Thomas GA, Speer JG, Matlock DK, Krauss G, Hackenberg RE. Time-temperature equivalence in martensite tempering. In: Proceedings of the International Conference on Olson, GB, Lieberman, DS & Saxena, A; 2008:595-600.
62. Shlyakman B, Yampolskii O, Ratushev D. A method for determining constant C in the Hollomon parameter. *Metal Sci Heat Treat*. 2011;52.
63. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830.
64. Lundberg S, Lee S-I. Shap - GitHub repository. https://github.com/shap/shap; 2017.
65. Agrawal A, Choudhary A. An online tool for predicting fatigue strength of steel alloys based on ensemble data mining. *Int J Fatigue*. 2018;113:389-400.
66. Bach S, Binder A, Montavon G, Klauschen F, Müller K-R, Samek W. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PloS One*. 2015;10(7):e0130140.
67. Samek W, Montavon G, Lapuschkin S, Anders CJ, Müller K-R. Explaining deep neural networks and beyond: a review of methods and applications. *Proc IEEE*. 2021;109(3):247-278.
68. Siddiqui SA, Mercier D, Munir M, Dengel A, Ahmed S. TSVIZ: demystification of deep learning models for time-series analysis. *IEEE Access*. 2019;7:67027-67040.
69. Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks; 2017.
70. Hong Y, Hu Y, Zhao A. Effects of loading frequency on fatigue behavior of metallic materials a literature review. *Fatigue Fract Eng Mater Struct*. 2023;46(8):3077-3098.
71. Takeuchi E, Furuya Y, Nagashima N, Matsuoka S. The effect of frequency on the giga-cycle fatigue properties of a ti–6al–4v alloy. *Fatigue Fract Eng Mater Struct*. 2008;31(7):599-605.
72. Beumelburg W. *Das verhalten von einsatzgehärteten proben mit verschiedenen oberflächenzuständen und randkohlenstoffgehalten im umlaufbiege-, statischen biege-und schlagbiegeversuch*; 1974.
73. Razim C. Some facts and considerations of trends in gear steels for the automotive industry. *Alloys Eighties*. 1980:9-23.
74. Schuster A. Charakterisierung des faserverlaufs in umgeformten stählen und dessen auswirkungen auf mechanische eigenschaften; 2013.
75. Bhadeshia HKDH, Edmonds DV. Bainite in silicon steels: new composition–property approach part 1. *Metal Sci*. 1983;17(9):411-419.
76. Kozeschnik E, Bhadeshia HKDH. Influence of silicon on cementite precipitation in steels. *Mater Sci Technol*. 2008;24(3):343-347.
77. Gulbay O, Ackermann M, Gramlich A, Durmaz AR, Steinbach I, Krupp U. Influence of transformation temperature on the high-cycle fatigue performance of carbide-bearing and carbide-free bainite. steel research international; 2023.
78. Schmiedhofer U. Untersuchung des niederdruckaufkohlverhaltens von korrosionsträgen, randschichthärtbaren lagerwerkstoffen für den einsatz in flugzeugtriebwerken; 2007.
79. Boyce BL, Michael JR, Kotula PG. Fatigue of metallic microdevices and the role of fatigue-induced surface oxides. *Acta Mater*. 2004;52(6):1609-1619.
80. Berrevoets J, Kacprzyk K, Qian Z, van der Schaar M. Causal deep learning; 2023.
81. Luo Y, Peng J, Ma J. When causal inference meets deep learning. *Nature Mach Intell*. 2020;2(8):426-427.
82. Bayerlein B. Innovations-plattform material digital; 2020.
83. de Pablo JJ, Jackson NE, Webb MA, et al. New frontiers for the materials genome initiative. *npj Comput Mater*. 2019;5(1):41.
84. O'Meara S. Materials science is helping to transform china into a high-tech economy. *Nature*. 2019;567(7748):S1-S1.
85. Cakiroglu C, Aydın Y, Bekdaş G, Geem ZW. Interpretable predictive modelling of basalt fiber reinforced concrete splitting tensile strength using ensemble machine learning methods and shap approach. *Materials*. 2023;16(13):4578.
86. Feng D-C, Wang W-J, Mangalathu S, Taciroglu E. Interpretable XGBoost-SHAP machine-learning model for shear strength prediction of squat rc walls. *J Struct Eng*. 2021;147(11):04021173.
87. Mangalathu S, Hwang S-H, Jeon J-S. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. *Eng Struct*. 2020;219:110927.

How to cite this article: Frie C, Riza Durmaz A, Eberl C. Exploration of materials fatigue influence factors using interpretable machine learning. *Fatigue Fract Eng Mater Struct*. 2024;47(8):2752-2773. doi:10.1111/ffe.14315

APPENDIX A: RANDOM FOREST HYPERPARAMETERS

TABLE A1 Scikit-learns'[63] random forest regressor default hyperparmeters with version=1.3.0.

Hyperparameter	Entry
n_estimators	100
max_depth	None
min_samples_split	2
min_samples_leaf	1
min_weight_fraction_leaf	0.0
max_Features	1.0
max_leaf_nodes	None
min_impurity_decrease	0.0

APPENDIX B: COMPONENT DESIN, LOAD, AND TESTING

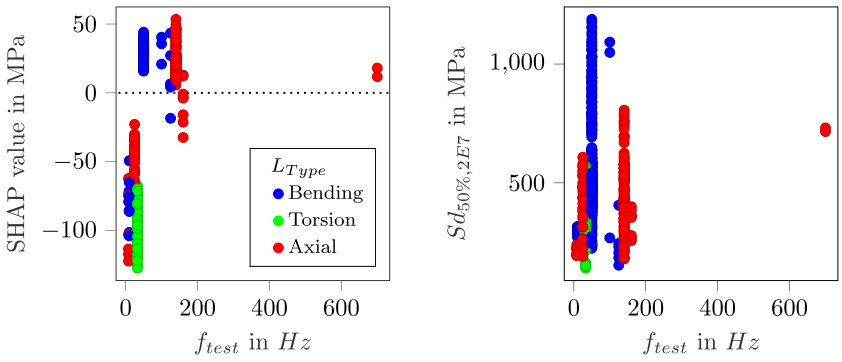

FIGURE B1 Comparison of SHapley Additive exPlanations (SHAP) value analysis from the frequency compared with the explorative data analysis of fatigue range and frequency. [Colour figure can be viewed at wileyonlinelibrary.com]

APPENDIX C: HEAT TREATMENT SHAP ASSESSMENT FOR THE FEATURE SET 1

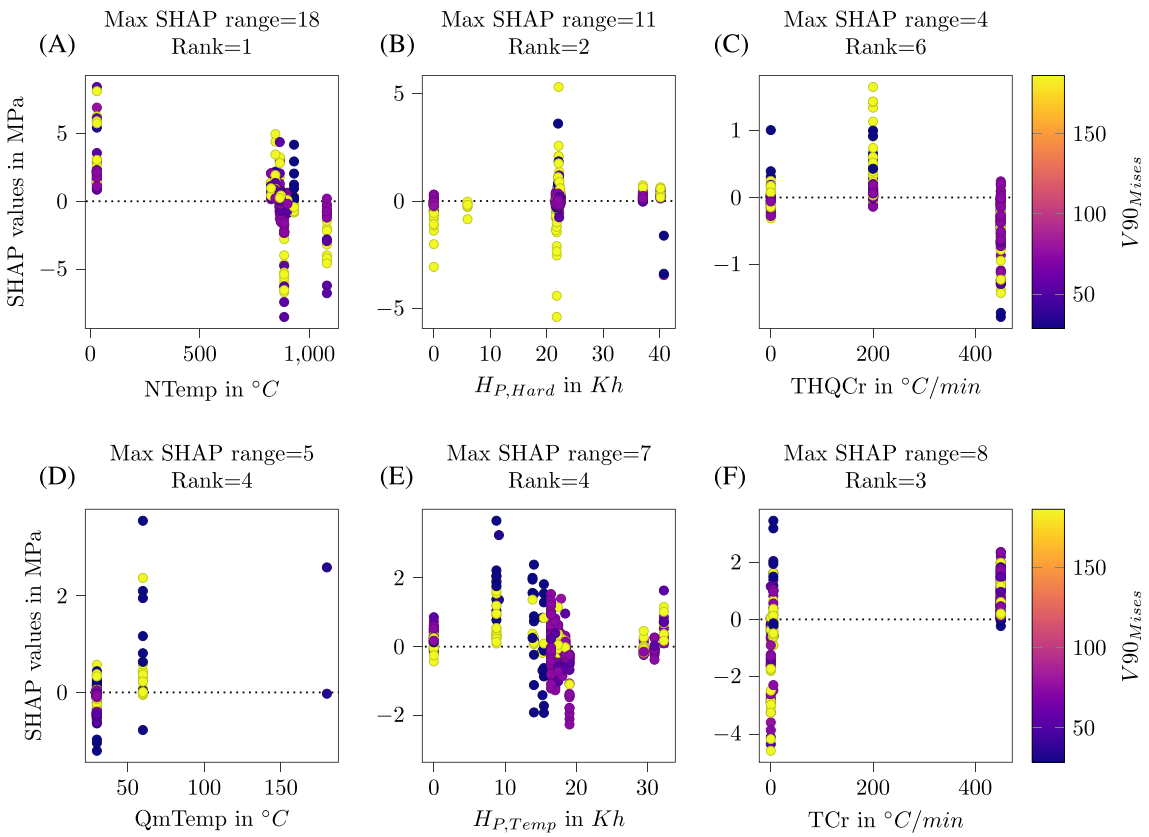

FIGURE C2 (A–F) Influence of each heat treatment step on the fatigue strength depending on the specimen size ($V90_{Mises}$) for *Feature Set 1* (see Table 1). Maximum SHapley Additive exPlanations (SHAP) ranges are reported in MPa and utilized to rank the features against each other. The subplots are ordered according to the process route. The dotted line is the zero reference line. [Colour figure can be viewed at wileyonlinelibrary.com]

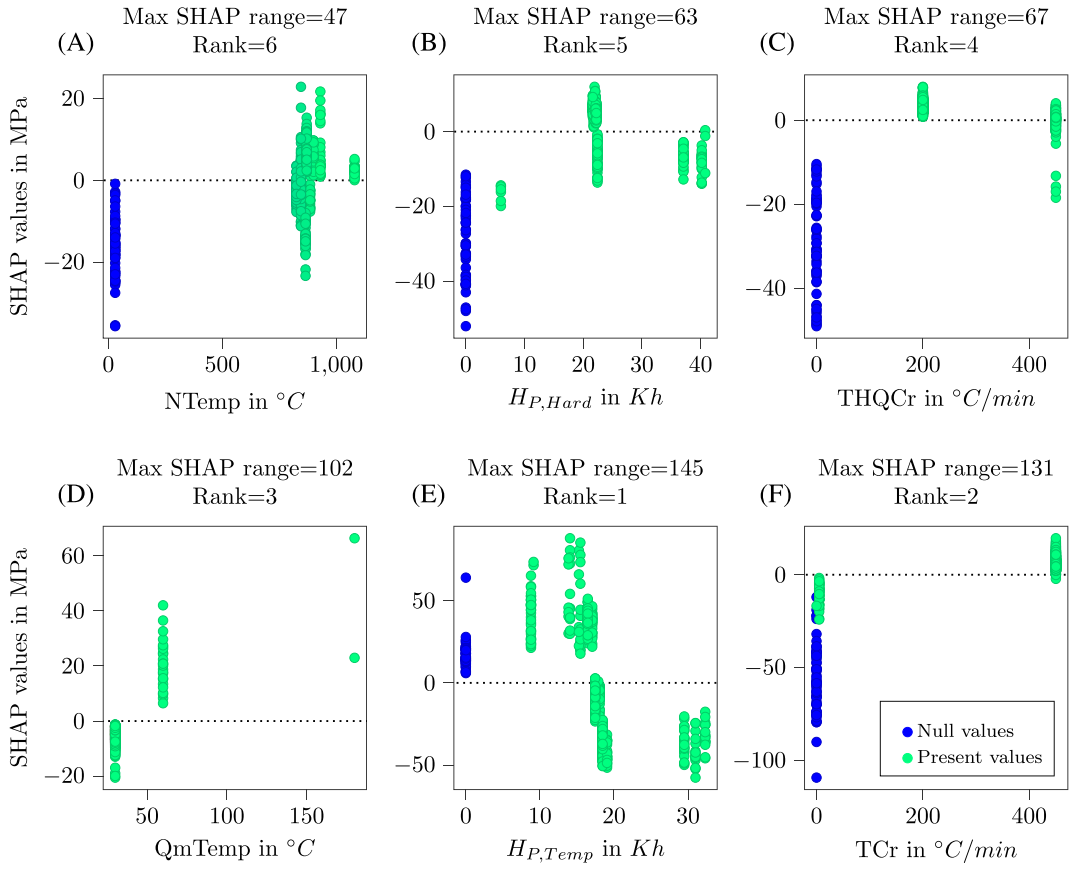

FIGURE C3 (A–F) Influence of each heat treatment step on the fatigue strength classified whether the heat treatment was present or absent based on the SHapley Additive exPlanations (SHAP) values for *Feature Set 3* (see Table 1). Maximum SHAP ranges are reported in MPa and utilized to rank the features against each other. The subplots are ordered according to the process route from (A) to (F). All subplots are colored according to the legend in subplot F. *Null values* indicate that the process step was not conducted while *Present* values denote the presence of this step. The dotted line is the zero reference line. [Colour figure can be viewed at wileyonlinelibrary.com]

APPENDIX D: ALLOYING ELEMENTS

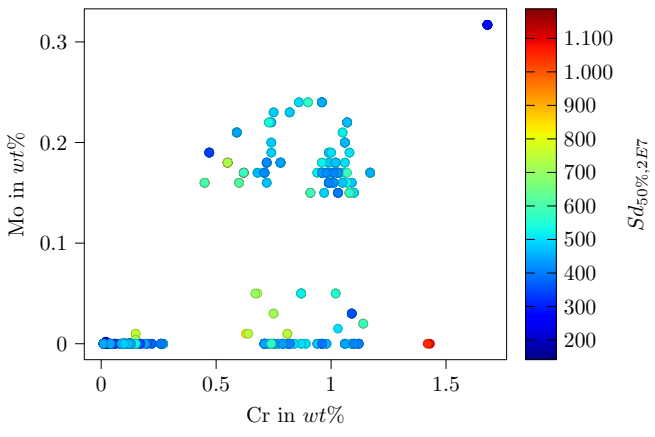

FIGURE D4 Illustration of clustered steel groups due to different alloying systems of chromium and molybdenum colored by the fatigue strength. [Colour figure can be viewed at wileyonlinelibrary.com]

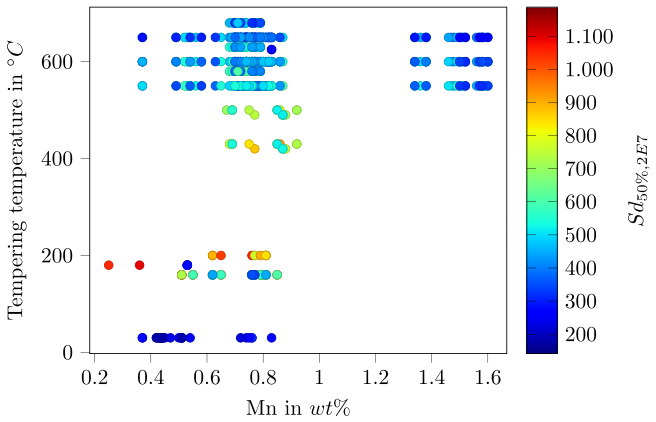

FIGURE D5 Illustration of the effect of the tempering temperature and manganese concentration on the fatigue strength. [Colour figure can be viewed at wileyonlinelibrary.com]