Supplementary Information

Machine learning assisted interpretation of creep and fatigue life in titanium alloys

Sucheta Swetlana[†], Ashish Rout[†], and Abhishek K. Singh[*†]

[†]Materials Research Centre, Indian Institute of Science, Bangalore-560012, India

1 Methodology

1.1 Outlier detection

Outliers are the data points that are significantly different from the rest of the data set. These observations can often skew the data distribution, leading to a misleading representation. Z-score outlier detection method is a standard deviation based process used for outlier detection. Z-score method provides information about the distance of data point from its mean value. If the Z-score of a data point is greater than 3, it indicates that the data point is quite different from the other data points. Such a data point can be an outlier. The equation for Z-score outlier detection is,

$$Z\ score = \frac{X - X_{mean}}{X_{std}} \tag{S1}$$

where, X_{mean} is the average and X_{std} is the standard deviation of the data point. The Z-score eliminates the data points that lie beyond the threshold (Z-score $>$ 3).

1.2 Machine Learning algorithm

Gradient Boost Regression (GBR) model is an ensemble of decision trees [1]. Gradient boosting minimizes the loss function by iteratively adding weak learners (regression trees) following a gradient descent procedure. GBR overcomes the limitations of weak learners by constructing multiple sequential decision trees.

Mathematically, GBR considers the initial prediction (F(x)) in the following form,

$$F(x) = \sum_{k=1}^{K} \gamma_k h_k(x) \tag{S2}$$

where, K is the total number of decision trees, γ_k is the learning rate, and $h_k(x)$ are the weak learners. Then, GBR develops the iterative model in a greedy fashion,

$$F(x) = F_{k-1}(x)\gamma_k h_k(x) \tag{S3}$$

where, the newly added tree $h_k(x)$ tries to minimize the loss function L, given the previous ensemble $F_{k-1}(x)$ such that,

$$h_k = \arg\min_h \sum_{i=1}^{n} L(y_i, F_{k-1}(x_i) + h(x_i)) \tag{S4}$$

The model is developed in a stepwise manner with each consecutive tree trained on the residuals of previous model. Therefore, it considers the overall sample space to improve the model performance.

[*]abhishek@iisc.ac.in

1.3 SHapley Additive exPlanation (SHAP)

The interpretability of ML models are explained by adapting SHapley Additive exPlanations (SHAP) method to identify the influence of each descriptor on its target properties [2]. The SHAP method is mostly used to interpret the ”black-box” machine learning models. Interpretation of these models is necessary to develop unique design principles and new theories for accelerating the discovery of promising materials. SHAP uses an additive feature attribution method, which evaluates the output model as a linear addition of input variables. Assuming the model with input variables $\{x_i\}_1^N$, where N is number of input variables, the explanation model *g(x′)* with simplified input *x′* for an original model *f(x)* is expressed as,

$$f(x) = g(x') = \phi_0 + \sum_{i=1}^{N} \phi_i x'_i \tag{S5}$$

where ϕ_i represents the feature's contribution value and ϕ_0 represents the constant value when all inputs are missing. The inputs x and x' are related through a mapping function $x' = h_x(x)$. SHAP returns a unique value related to the input feature's importance to the model output based on these three properties: local accuracy, missingness, and consistency. In order to compute SHAP values efficiently, the expected value of the function conditioned on a subset S of all features is defined as $E[f(x)|x_S]$. The resulting importance of each feature is weighted by all possible conditional expectations as denoted,

$$\phi_i = \sum_{S \subseteq N/\{i\}} \frac{|S|!(N - |S| - 1)!}{N!} \Big[f_x(S \cup \{i\}) - f_x(S) \Big] \tag{S6}$$

where, S represents the set of nonzero x' values, and N is the set of all the features. In more clear terms, for each prediction, the model generates a prediction value, and a SHAP value is assigned to each feature. Assuming that the i_{th} sample is x_i, the j_{th} feature for i_{th} sample is x_{ij} and the model's predicted value for sample is y_i. The mean of the target property is the base value, y_{base} for the entire model. Then, the SHAP value for the i_{th} sample follows this equation,

$$y_i = y_{base} + f(x_{ij}) \tag{S7}$$

where, $f(x_{ij})$ is the SHAP value of x_{ij}. In other words, $f(x_{ij})$ is the contribution of the j_{th} feature in the i_{th} sample to the predicted value y_i. If $f(x_{ij}) > 0$, it means that feature has a positive effect and improves the predicted value and vice versa.

2 Data and descriptors used for prediction of fatigue life (N_f in cycles)

The data collected for the prediction of fatigue life are taken from the experimental published literature on titanium alloys [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21]. Table S1 lists down all the descriptors considered for the development of GBR model.

Table S1: Statistics of all attributes utilized in the present study for prediction of N_f.

Features	Features details	Abbreviation	min	max	mean
Chemical composition	Titanium (wt %)	Ti	52.39	89.94	84.65
	Alumunium (wt %)	Al	5.75	28.67	7.38
	Vanadium (wt %)	Vn	0	4.2	1.49
	Carbon (wt %)	C	0	0.08	0.03
	Nitrogen (wt %)	N	0	0.53	0.04
	Oxygen (wt %)	O	0	0.2	0.09
	Hydrogen (wt %)	H	0	0.023	0.003
	Iron (wt %)	Fe	0	0.2	0.07
	Silicon (wt %)	Si	0	0.35	0.17
	Tin (wt %)	Sn	0	4.02	2.13
	Niobium (wt %)	Nb	0	18.7	1.43
	Molybednum (wt %)	Mo	0	6.0	0.45
	Zirconium (wt %)	Zr	0	4.35	2.12
Experimental parameters	Applied fatigue stress (MPa)	σ_f	180.72	1426.0	685.48
	Total strain (%)	ϵ_{tot}	0	2	0.71
	Temperature of measurement (cel)	T_f	25	750	205.7
	Stress ratio	R	-1	0.8	-0.37
	Frequency of load (Hz)	f	0	100	10.2
Heat treatment conditions	Aging temperature (cel)	T_{age}	0	1300	685.51
	Aging time (hrs)	t_{age}	0	24	3.36
	Solution temperature (cel)	T_{sol}	0	1050	675
	Solution time (cel)	t_{sol}	0	2	0.5

3 Prediction of N_f using different ML models

The performance of different machine learning models is carried out to select the best model with highest R^2 and lowest RMSE.

Table S2: Performance of different ML models. The model performance is evaluated using pyCaret library of python [22].

Sl No.	Models	R^2
1	Gradient Boost Regressor (GBR)	0.913
2	Random Forest Regressor (RF)	0.906
3	K Neighbors Regressor (KNN)	0.888
4	Decision Tree Regressor (DT)	0.884
5	Support Vector Regressor (SVM)	0.879

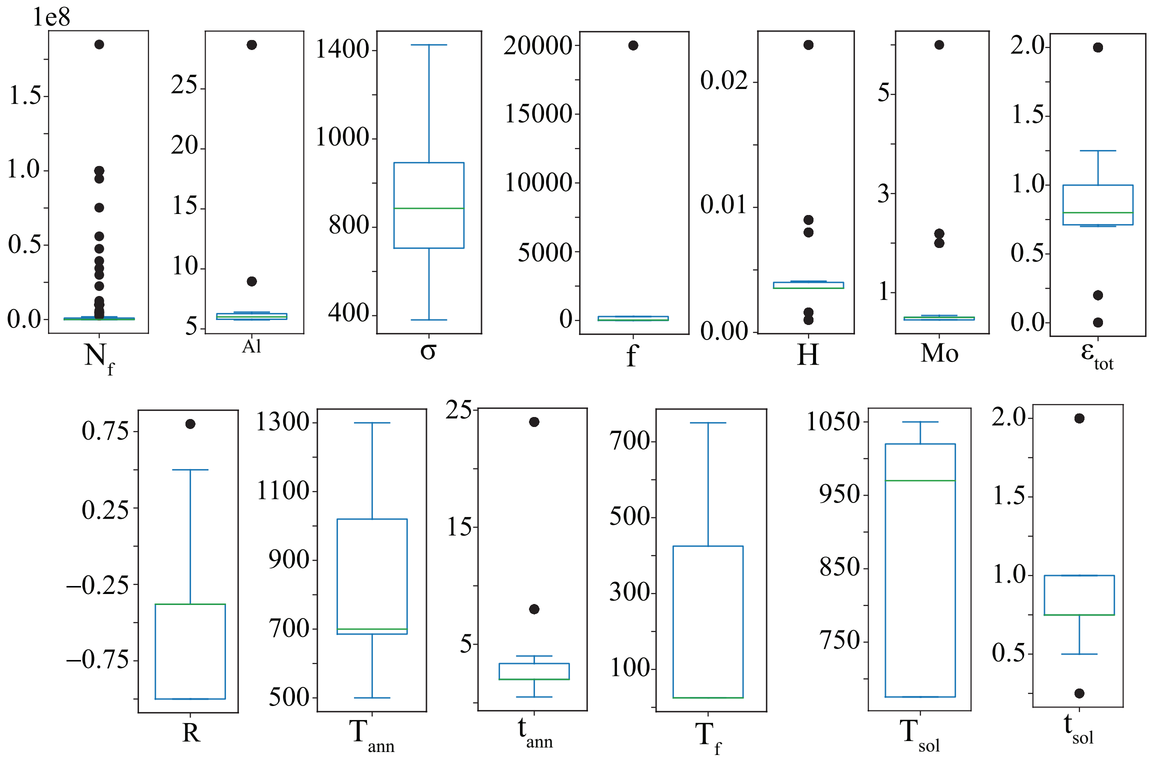

Figure S1: Box plot for all the features considered for prediction of N_f. The points that lie beyond the interquartile range are eliminated as outliers.

Table S3: Performance of GBR model on unseen data of titanium alloys for prediction of fatigue life cycles.

Sl No.	**Actual** N_f	**Predicted** N_f
1	6.37	6.69
2	6.73	6.63
3	6.88	6.99
4	4.59	5.12
5	4.93	5.23
6	5.82	5.41
7	6.40	5.44
8	6.57	5.49
9	6.73	5.72
10	7.00	5.73

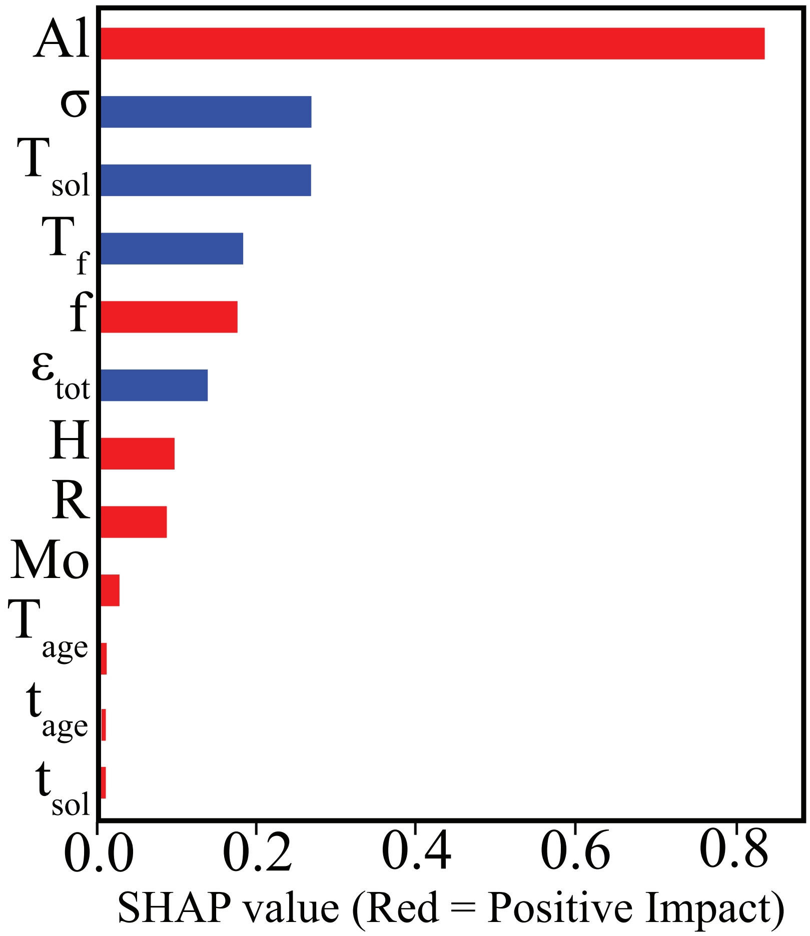

Figure S2: SHAP global plot showing the importance score of features selected using PCC values. The color red denotes a positive correlation with N_f and blue denotes negative correlation.

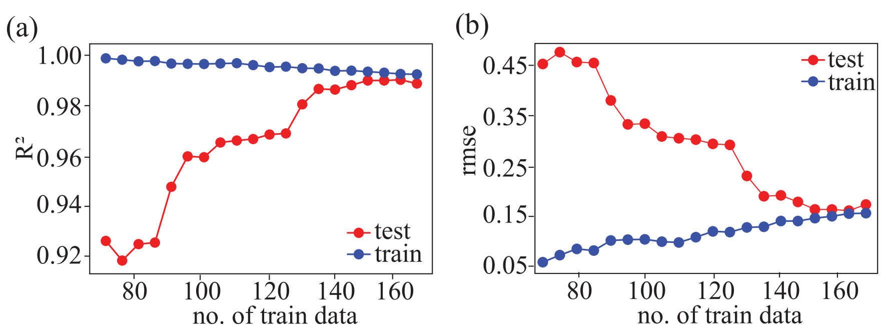

Figure S3: Learning curve plot for N_f (a) R^2 increases and gradually converges with increase in train data (b) $rmse$ decreases and gradually converges with increase in train data.

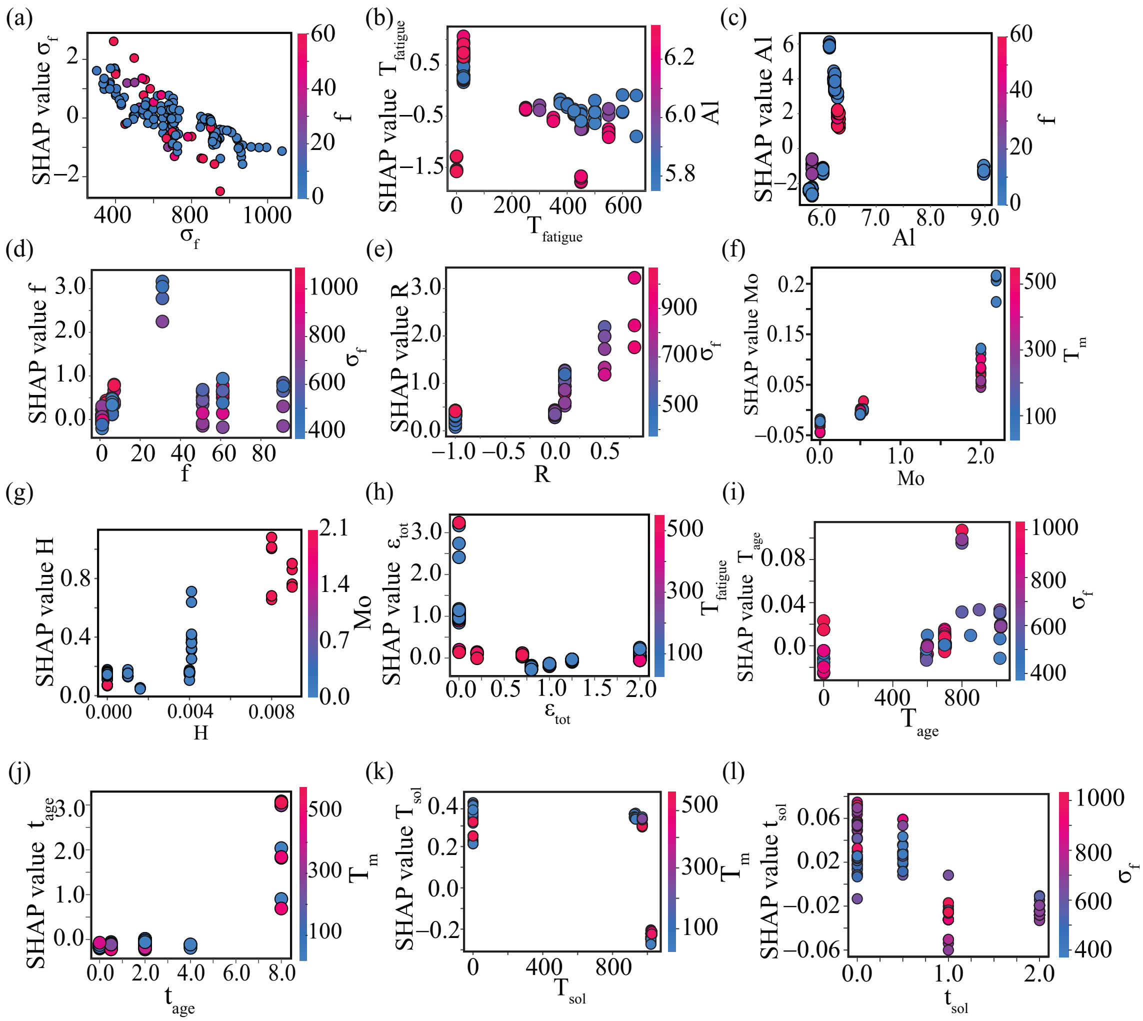

Figure S4: SHAP dependence plot showing the variation of relevant features with N_f.

4 Descriptors used for prediction of creep rupture (t_r in hours)

Experimental data are collected for the prediction of creep rupture life [23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50]. Table S4 shows the descriptors used for prediction of creep rupture life (t_r).

Table S4: Statistics of all attributes utilized in the present study for prediction of t_r.

Features	Features details	Abbreviation min	max	mean	
Chemical composition	Titanium (wt %)	Ti	48	89.98	85.99
	Alumunium (wt %)	Al	5.5	48	8.03
	Vanadium (wt %)	Vn	0	5.08	3.03
	Carbon (wt %)	C	0	0.1	0.03
	Nitrogen (wt %)	N	0	0.05	0.02
	Oxygen (wt %)	O	0	0.24	0.11
	Hydrogen (wt %)	H	0	0.40	0.011
	Iron (wt %)	Fe	0	1.18	0.39
	Silicon (wt %)	Si	0	0.3	0.07
	Tin (wt %)	Sn	0	4.8	0.52
	Niobium (wt %)	Nb	0	2.0	0.16
	Molybednum (wt %)	Mo	0	5.1	0.65
	Zirconium (wt %)	Zr	0	5.0	0.78
	Boron (wt %)	B	0	2.0	0.15
	Chromium (wt %)	Cr	0	0.5	0.02
Experimental parameters	Rupture strain (%)	ϵ_r	0	95	14.76
	Temperature of measurement (cel)	T_{creep}	315	773	568
	Steady state strain rate (1/s)	$\dot{\epsilon}$	0	$2.13x10^{-3}$	$5.83x10^{-5}$
	Applied stress (MPa)	σ	14	673	281.75
Heat treatment conditions	Aging temperature (cel)	T_{age}	0	1373	621.59
	Aging time (hrs)	t_{age}	0	150	2.78
	Solution temperature (cel)	T_{sol}	0	1100	404
	Solution time (cel)	t_{sol}	0	4	0.48

5 Prediction of creep rupture life (t_r) using different ML models

The performance of different machine learning models is carried out to select the best model with highest R^2 and lowest RMSE.

Table S5: Performance of different ML models. The model performance is evaluated using pyCaret library of python [22].

Sl No.	Models	Prediction accuracy
1	Gradient Boost Regressor (GBR)	0.77
2	Random Forest Regressor (RF)	0.76
3	Ridge Regression (RR)	0.72
4	Decision Tree Regressor (DT)	0.66
5	Support Vector Regression (SVR)	0.31

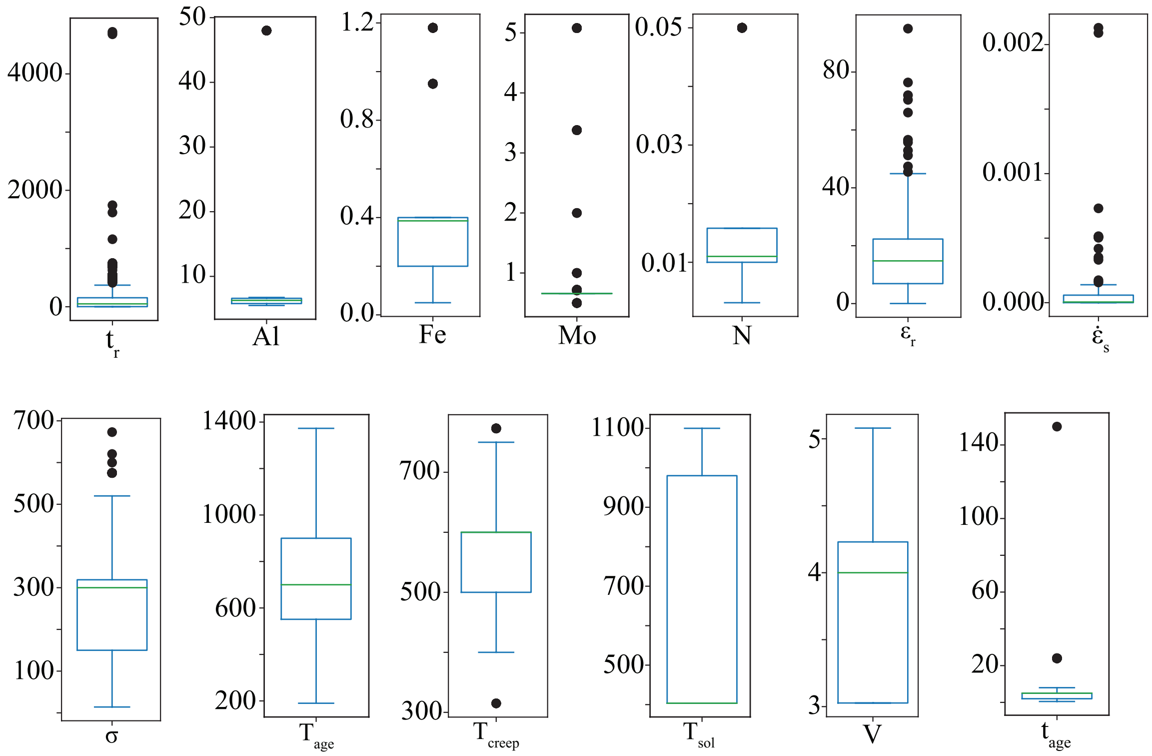

Figure S5: Box plot for all the features considered for prediction of t_r. The points that lie beyond the interquartile range are eliminated as outliers.

Table S6: Performance of GBR model on unseen data of titanium alloys for prediction of creep rupture life cycles

Sl No.	**Actual** t_r	**Predicted** t_r
1	1.94	1.89
2	2.21	2.28
3	2.48	2.31
4	2.82	2.38
5	3.24	2.24
6	3.19	2.23
7	2.55	2.35
8	2.09	2.13
9	2.41	1.93
10	3.93	2.23
11	2.75	2.34
12	2.37	2.12

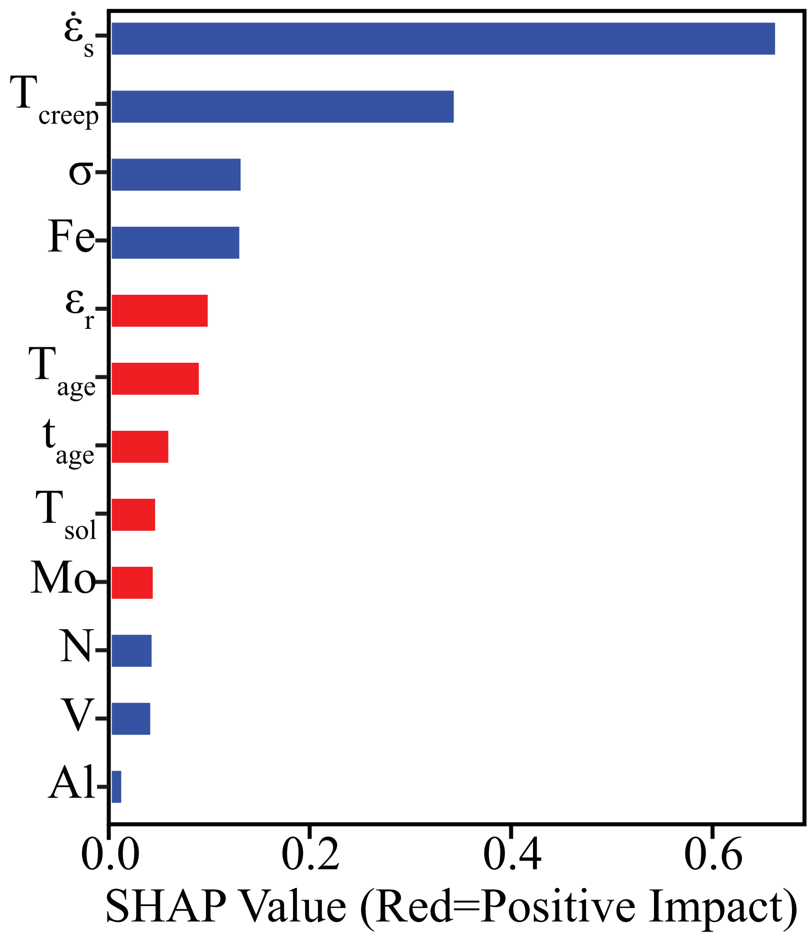

Figure S6: SHAP global plot showing the importance score of features selected using PCC values for prediction of t_r. The color red denotes a positive correlation with N_f and blue denotes negative correlation.

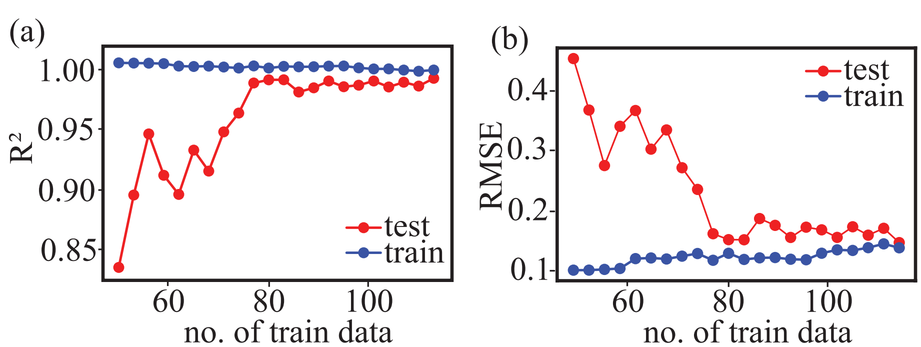

Figure S7: Learning curve plot for t_r (a) R^2 increases and gradually converges with increase in train data (b) $rmse$ decreases and gradually converges with increase in train data

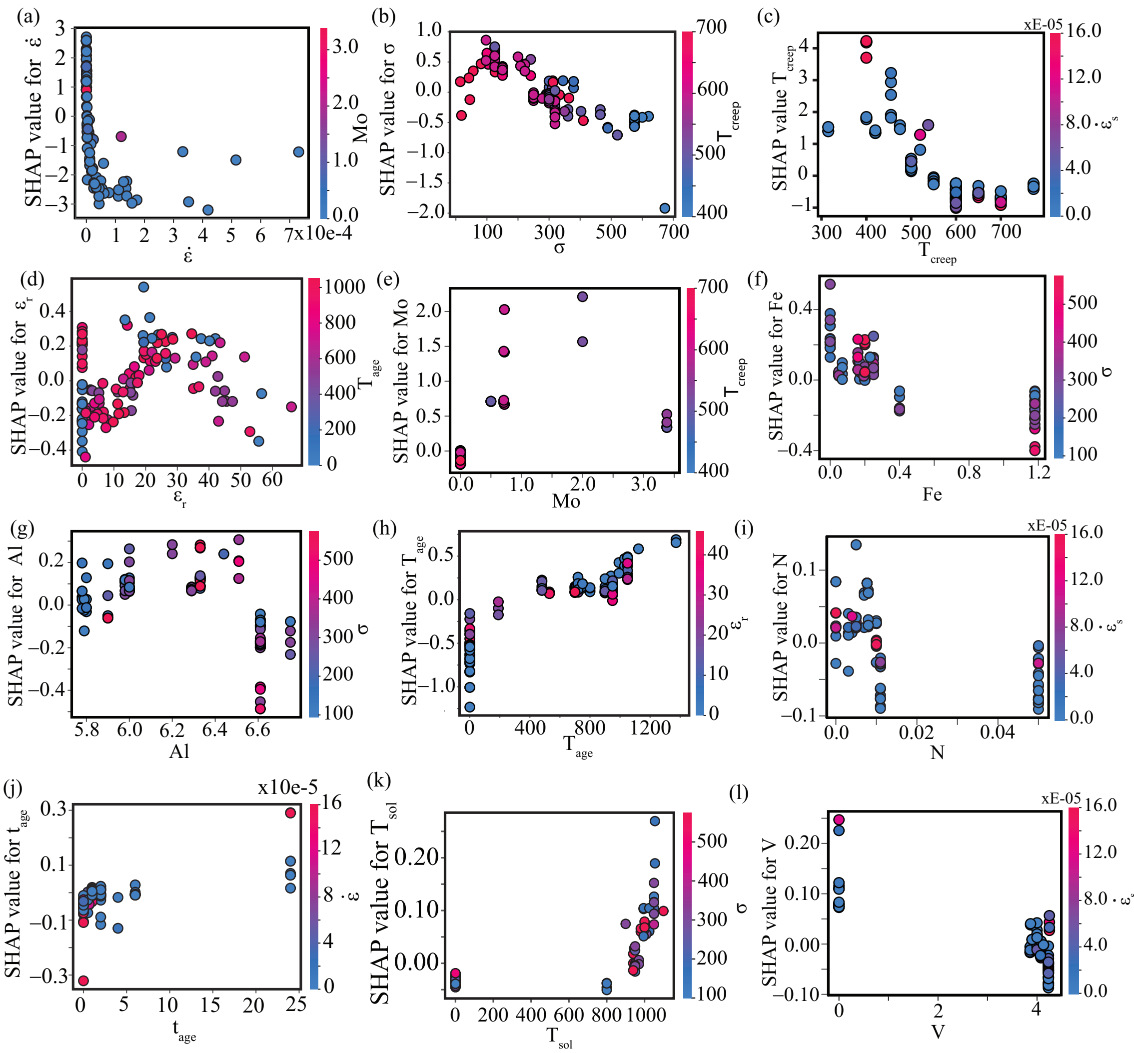

Figure S8: SHAP dependence plot showing the variation of relevant features with t_r.

References

[1] J. H. Friedman, Computational statistics & data analysis **38**, 367 (2002).

[2] S. M. Lundberg, G. G. Erion, and S.-I. Lee, arXiv preprint arXiv:1802.03888 (2018).

[3] A. Kumar, N. Singh, and V. Singh, Materials characterization **51**, 225 (2003).

[4] H. Maier, R. Teteruk, and H.-J. Christ, Metallurgical and Materials Transactions A **31**, 431 (2000).

[5] T. Heckel, A. G. Tovar, and H.-J. Christ, Experimental mechanics **50**, 483 (2010).

[6] S. Fujishiro and D. Eylon, Thin Solid Films **54**, 309 (1978).

[7] G. Lütjering, Materials Science and Engineering: A **243**, 32 (1998).

[8] Z. Liang, Z. Sun, W. Zhang, S. Wu, and H. Chang, Journal of Alloys and Compounds **782**, 1041 (2019).

[9] K. Prasad, R. Sarkar, P. Ghosal, and V. K. Varma, Materials Science and Engineering: A **494**, 227 (2008).

[10] I. Altenberger, R. K. Nalla, Y. Sano, L. Wagner, and R. O. Ritchie, International Journal of Fatigue **44**, 292 (2012).

[11] R. S. Bellows, S. Muju, and T. Nicholas, International Journal of Fatigue **21**, 687 (1999).

[12] R. Nalla, R. Ritchie, B. Boyce, J. Campbell, and J. Peters, Metallurgical and Materials Transactions A **33**, 899 (2002).

[13] J. Zuo, Z. Wang, and E. Han, Materials Science and Engineering: A **473**, 147 (2008).

[14] J. Hines, J. Peters, and G. Lütjering, *Microcrack propagation in Ti-6Al-4V alloys*, Tech. Rep. (Technical Univ. Hamburg-Harburg (DE), 1999).

[15] S. Adachi, L. Wagner, and G. Lütjering, in *Strength of Metals and Alloys (ICSMA 7)* (Elsevier, 1986) pp. 2117–2122.

[16] M. Peters, A. Gysler, and G. Lütjering, Metallurgical and Materials Transactions A **15**, 1597 (1984).

[17] C. Sauer and G. Lütjering, Materials Science and Engineering: A **319**, 393 (2001).

[18] Y. Zhou, J. Wang, B. Zhang, W. Ke, and E. Han, Intermetallics **24**, 7 (2012).

[19] P. Pototzky, H. Maier, and H.-J. Christ, Metallurgical and Materials Transactions A **29**, 2995 (1998).

[20] S. Rajan, P. Wanjara, J. Gholipour, and A. S. Kabir, Materials **14**, 30 (2021).

[21] Y. Wu, J. Liu, H. Wang, S. Guan, R. Yang, and H. Xiang, Journal of materials science & technology **34**, 1189 (2018).

[22] M. Ali, PyCaret version **2** (2020).

[23] G. Singh, D. Satyanarayana, R. Pederson, R. Datta, and U. Ramamurty, Materials Science and Engineering: A **597**, 194 (2014).

[24] V. Oliveira, A. Vazquez, C. Aguiar, A. Robin, and M. Barboza, Materials Science and Engineering: A **670**, 357 (2016).

[25] X. Li, T. Sugui, B. Xianyu, and C. Liqing, Materials Science and Engineering: A **529**, 452 (2011).

[26] X. Li, T. Sugui, B. Xianyu, and C. Liqing, Materials Science and Engineering: A **559**, 401 (2013).

[27] L. Briguente, A. A. Couto, N. M. Guimarães, D. A. Reis, C. de Moura Neto, and M. Barboza, in *Defect and Diffusion forum*, Vol. 326 (Trans Tech Publ, 2012) pp. 520–524.

[28] P. Liu, Y. Zong, D. Shan, and B. Guo, Materials Science and Engineering: A **638**, 106 (2015).

[29] M. Barboza, C. M. Neto, and C. Silva, Materials Science and Engineering: A **369**, 201 (2004).

[30] A. G. dos Reis, D. A. P. Reis, C. de Moura Neto, M. J. R. Barboza, and J. Oñoro, Materials Science and Engineering: A **577**, 48 (2013).

[31] D. Reis, C. Silva, M. Nono, M. Barboza, F. P. Neto, and E. Perez, Materials Science and Engineering: A **399**, 276 (2005).

[32] S. Nishino, K. Shiozawa, and Y. Aikawa, Journal of the Society of Materials Science, Japan **47**, 206 (1998).

[33] M. Barboza, E. Perez, M. Medeiros, D. Reis, M. Nono, F. P. Neto, and C. Silva, Materials Science and Engineering: A **428**, 319 (2006).

[34] L. Badea, M. Surand, J. Ruau, and B. Viguier, University Polytechnica of Bucharest Scientific Bulletin, Series B **76**, 185 (2014).

[35] W. Xiong, J. Luo, B. Wang, X. Li, and J. Chen, International Journal of Lightweight Materials and Manufacture **3**, 298 (2020).

[36] Y.-K. Kim, S.-H. Park, J.-H. Yu, B. AlMangour, and K.-A. Lee, Materials Science and Engineering: A **715**, 33 (2018).

[37] F. Seco and A. Irisarri, Fatigue & Fracture of Engineering Materials & Structures **24**, 741 (2001).

[38] A. Szkliniarz and W. Szkliniarz, Archives of Metallurgy and Materials **65** (2020).

[39] L. A. N. d. S. Briguente, F. P. Briguente, D. A. P. Reis, C. d. Moura Neto, and F. A. Resende, Tecnologia em Metalurgia, Materiais e Mineração **10**, 324 (2014).

[40] F. E. de Freitas, A. G. dos Reis, and D. A. P. Reis, Tecnologia em Metalurgia, Materiais e Mineração **16**, 0 (2019).

[41] G. F. Almeida, A. A. Couto, D. A. Reis, M. Massi, A. S. da Silva Sobrinho, and N. B. De Lima, Metals **8**, 618 (2018).

[42] J. Zhu, P. Liaw, J. Corum, and H. McCoy, Metallurgical and Materials Transactions A **30**, 1569 (1999).

[43] W. Harrigan, Metallurgical Transactions **5**, 565 (1974).

[44] Y. Gu, F. Zeng, Y. Qi, C. Xia, and X. Xiong, Materials Science and Engineering: A **575**, 74 (2013).

[45] K. Prasad, R. Sarkar, P. Ghosal, D. Satyanarayana, S. Kamat, and T. Nandy, Materials Science and Engineering: A **528**, 6733 (2011).

[46] X.-a. Nie, Z. Hu, H.-q. Liu, D.-q. Yi, T.-y. Chen, B.-f. Wang, Q. Gao, and D.-c. Wang, Materials Science and Engineering: A **613**, 306 (2014).

[47] Y.-K. Kim, S.-J. Youn, S.-W. Kim, J. Hong, and K.-A. Lee, Materials Science and Engineering: A **763**, 138138 (2019).

[48] Z. Chen, J. Li, J. Liu, Q. Wang, J. Liu, and R. Yang, Journal of Materials Science & Technology **26**, 564 (2010).

[49] V. Chandravanshi, R. Sarkar, P. Ghosal, S. Kamat, and T. Nandy, Metallurgical and Materials Transactions A **41**, 936 (2010).

[50] C. Quesne, C. Duong, F. Charpentier, J. Fries, and P. Lacombe, Journal of the Less Common Metals **68**, 133 (1979).