

ARTICLE OPEN



Machine learning augmented predictive and generative model for rupture life in ferritic and austenitic steels

Osman Mamun^{1✉}, Madison Wenzlick^{2,3}, Arun Sathanur¹, Jeffrey Hawk² and Ram Devanathan¹

The Larson–Miller parameter (LMP) offers an efficient and fast scheme to estimate the creep rupture life of alloy materials for high-temperature applications; however, poor generalizability and dependence on the constant C often result in sub-optimal performance. In this work, we show that the direct rupture life parameterization without intermediate LMP parameterization, using a gradient boosting algorithm, can be used to train ML models for very accurate prediction of rupture life in a variety of alloys (Pearson correlation coefficient >0.9 for 9–12% Cr and >0.8 for austenitic stainless steels). In addition, the Shapley value was used to quantify feature importance, making the model interpretable by identifying the effect of various features on the model performance. Finally, a variational autoencoder-based generative model was built by conditioning on the experimental dataset to sample hypothetical synthetic candidate alloys from the learnt joint distribution not existing in both 9–12% Cr ferritic–martensitic alloys and austenitic stainless steel datasets.

npj Materials Degradation (2021)5:20; <https://doi.org/10.1038/s41529-021-00166-5>

INTRODUCTION

9–12% Cr ferritic–martensitic alloys (FMA)^{1,2} and austenitic stainless steels^{3–5} are used pervasively in high-temperature, high-pressure components in fossil energy power plants. Opportunities to improve the efficiency of power plants depend highly on the materials available for use, which makes computational materials modeling to devise high-temperature-resistant materials an important area of research^{6,7}. When designing heat-resistant alloys for high-temperature applications two challenges are apparent. The first challenge is to devise a microstructure that is stable in the power plant environment to increase the lifetime of the alloy. On the other hand, the second aspect is related to the first in that the designed microstructure must have features that present an array of strengthening obstacles that persist over a wide temperature range so the alloy can operate in the new temperature envelope of the power plant. Traditionally, the design approach has focused on alloy chemistry and thermo-mechanical processing to achieve the desired mechanical and physical response to the power plant conditions⁸. However, due to the increasing complexity of the underlying physical relationships between chemistry for high temperature strength and stability, alloy design is expensive, and time consuming⁹. To better understand the physics of creep deformation in heat-resistant alloys, several semi-empirical time–temperature relations have been developed and used, including the Larson–Miller parameter (LMP)^{10,11} and the Manson–Haferd parameter (MHP)¹¹. However, the relatively poor predictive capability of these parameters, and the dependency on empirical constants such as C in the LMP, renders them unsatisfactory for accurate estimations of the life of a component of a heat-resistant alloy from short- to medium-term creep testing¹².

To ameliorate these limitations, data analytics (DA), including statistical analysis and machine learning (ML) algorithms, have emerged as powerful tools for making reliable estimations with defined risk very quickly¹³. The DA tools can, therefore, be used to construct a reliable, predictive model for creep rupture life using a

range of experimental data. To this end, a concerted effort has been made^{14–17} to build predictive models, but these models currently possess significant limitations. These limitations include:

- (1) having an extensive and well-curated database that satisfies the three V's pertinent to creep life model building, i.e., volume, variety, and veracity;
- (2) applying a well-suited algorithm to map the underlying physics-based function; and
- (3) integrating the most comprehensive physical interpretation of the developed model with experimental results and component life performance crucial in the inverse design of heat-resistant alloys.

The first problem arises from the fact that practical considerations of collecting and compiling data—such as experimental tensile, creep, fatigue, stress, relaxation, etc.—are both very expensive and time consuming. This is evidenced by the approximately 90 years that have passed since the problem of creep life predictions first arose in designing high-temperature power plants. Even after 30–50 years of various government and non-government efforts to collect creep data, the amount of knowledge accumulated, and available in the public domain, is not enough to build reliable predictive models for materials spanning a variety of alloy classes¹⁸. To tackle this problem, a specialized database for 9–12% Cr and austenitic stainless steels has been established as part of the National Energy Technology's (NETL) effort on Extreme Environment Materials (XMAT) research. XMAT was undertaken for three primary reasons:

- (1) Develop a modeling framework that can comprehensively define an arbitrary material's behavior (life performance with associated risk) in a power plant environment.
- (2) Construct a database of information on energy materials (and the platform for housing it) appropriate for use in this activity as well as providing information for use in other DA and ML activities.
- (3) Apply these methodologies through the design of an advanced austenitic stainless steel for boiler tubing and piping

¹Energy and Environment Directorate, Pacific Northwest National Laboratory, Richland, WA, USA. ²Materials Performance Division, National Energy Technology Laboratory, Albany, OR, USA. ³Leidos Research Support Team, Albany, OR, USA. ✉email: mdosman.mamun@pnnl.gov

with an aluminum oxide scale that extends the operating temperature 50 °C above the current industry standard alloys. The dataset constructed to date contains high quality and accurate data, including information from National Institute for Materials Science (NIMS), NETL, industry, and open-source literature^{19–31}.

The second problem, i.e., applying the appropriate algorithm to the data, partly arises from the actual dataset size. Substantial data volume is required to fit all the model parameters for a complex and physically accurate model; e.g., it is not uncommon to have 50–60 million parameters for a state-of-the-art image recognition neural network architecture³². Recently published activities have used ML algorithms, such as linear regression, random forest, and support vector machines. Alternatively, the Gradient Boosting Algorithm (GBA) for the prediction of rupture life has been used in this case as it has been shown to be very accurate in case of limited data^{33,34}.

The third problem, i.e., appropriately interpreting the ML model, is a general problem that is faced by both academia and industry^{35,36}. This arises because the high accuracy obtained by complex models also makes them extremely difficult to interpret even by the experts in the field. Further, the disconnect between materials science experts and data science experts means that data science results must be compatible to materials science principles and intuition. To improve the model interpretability, the Shapley value was calculated. The Shapley value is a game theory-inspired approach where the contribution of each player is quantified based on the performance with and without the active participation of that player in a multi-player game^{37,38}. In this article, a gradient boosting algorithm³⁹ is proposed for accurate predictions of rupture life with Shapley value analysis to interpret the effect of different features on the model performance.

Even after developing an accurate predictive model for rupture life, using this model to efficiently search the vast design space is a major challenge. For example, the predictive model can be used to (inversely) design an alloy (or alloy class) by identifying the composition, microstructure, and processing parameters required that may result in superior alloy rupture life⁴⁰. However, in the inverse design scheme, the enumeration of the design space for performing an exhaustive screening of the whole space is difficult when there is no explicit rule to generate synthetic alloys. Recently, generative ML models are being used to create synthetic data from the latent space distribution, which is derived from the real available dataset^{41,42}. In this approach, the dataset can be used to learn the implicit distribution of the alloy space. At this point random alloy samples are withdrawn from the resulting distribution. Due to the immense learning capacity of this type of model architecture, they can learn hidden rules in the dataset and can later exploit these rules to create novel data with targeted properties. In several materials science fields, generative models have been successfully applied to inverse design various properties^{43,44}. Despite the success in efficiently generating samples, the usage of generative ML models is very limited, mainly due to the difference in the architecture of such models and the difficulty in training them.

To supplement this, a variational autoencoder (VAE)-based generative model has been used to create synthetic alloy samples to assist in the inverse design of the alloys. Another application of the developed generative model is for data augmentation to improve ML model performance. For example, in image processing, a generative model has been successfully applied to augment the dataset several fold by sampling images from the latent space, thereby facilitating unprecedented ML model performance^{45,46}. Finally, the application of a generative model to increase the ML model accuracy is briefly discussed.

RESULTS AND DISCUSSION

ML model results

For the 9–12% Cr family data, a linear regression model was used to examine data on creep rupture life. For Scheme-1, Scheme-2, and Scheme-3, the resulting Pearson correlation coefficient (PCC) for the prediction of LMP ranges from 0.4 to 0.5, and for predicting rupture life the PCC varies from 0.2 to 0.4. These low correlation coefficients indicate that the relationship between the descriptors and LMP or rupture life is not a linear one. The motivation behind using a linear regression model is twofold: First, oftentimes a linear regression model can reveal a fundamental functional relation between the feature and target properties, and second, it can establish a baseline for model performance to compare other non-linear models. Also, it is of note that the linear model performance is computed using a greedy feature selection approach. In the greedy selection approach, both top-down (all the features are used and then iteratively the worst performing feature is discarded until the accuracy gain is negligible) and bottom-up approaches (the best feature is identified and then sequentially the next best feature is added until the performance reaches a plateau) are used.

Next, a Random Forest Regressor (RFR) on the training dataset was trained. However, the accuracy of the trained model was not significantly higher than the accuracy of the above linear models (PCC ranges from 0.3 to 0.5 for the test set) and, thus, is unsuitable for the purpose of building a meaningful predictive model. One important observation we note from training the linear regressor and the RFR is that even though the mean accuracy of the cross-validation scheme was low, the standard deviation was quite high, indicating that for some data the model can be very efficient. Based on these observations, the GBA was selected to train the model as it works as a decision tree, divides the data into several folds, and applies the best algorithm on that fold. Therefore, the resulting model may be able to handle a higher complexity in the datasets than either the linear regression or the random forest regression approaches.

To test the hypotheses about GBA, the ML model was trained with three different gradient boosting packages: *CatBoost*, *LightGBM*, and *XGBoost*. The performances of the three algorithms were found to be very close for the prediction of creep for the 9–12% Cr family data. However, *CatBoost* consistently performed slightly better than the other two algorithms. Additionally, the solution was robust, as the standard deviation of the prediction was found to be lower than that of the other two packages. Unless otherwise specified, in the rest of the article the *CatBoost* results are used to indicate the performance of the gradient boosting algorithm.

In Scheme-1 for the 9–12% Cr dataset, the performance of the gradient boosting algorithm for predicting the LMP, using the training and testing sets, was found to be $0.9997 \pm 9.56 \times 10^{-5}$ and $0.997 \pm 9.23 \times 10^{-3}$. Of note is that temperature was not used as a feature to build the LMP prediction model, rather it was used later to convert the LMP prediction to a rupture life prediction using the inverse LMP formula. When the LMP prediction was converted to rupture life prediction, the PCC for the training and testing sets were found to be 0.994 ± 0.003 and 0.882 ± 0.038 . In Fig. 1, the parity plot, learning curve, and feature importance for the gradient boosting algorithm are shown. From the feature importance plot, the logarithm of stress was found to be the most important feature, as expected. Interestingly, the minimum creep rate (MCR) was also found to be very important (it usually takes a day or two to measure MCR, but it gives valuable insights into the rupture life prediction).

For Scheme-2, a Gaussian Process with a combination of a RBF kernel and a white kernel was optimized based on the 75/25% split of training and testing data to predict the C-parameter for an unknown alloy material that was not present in the library. Based

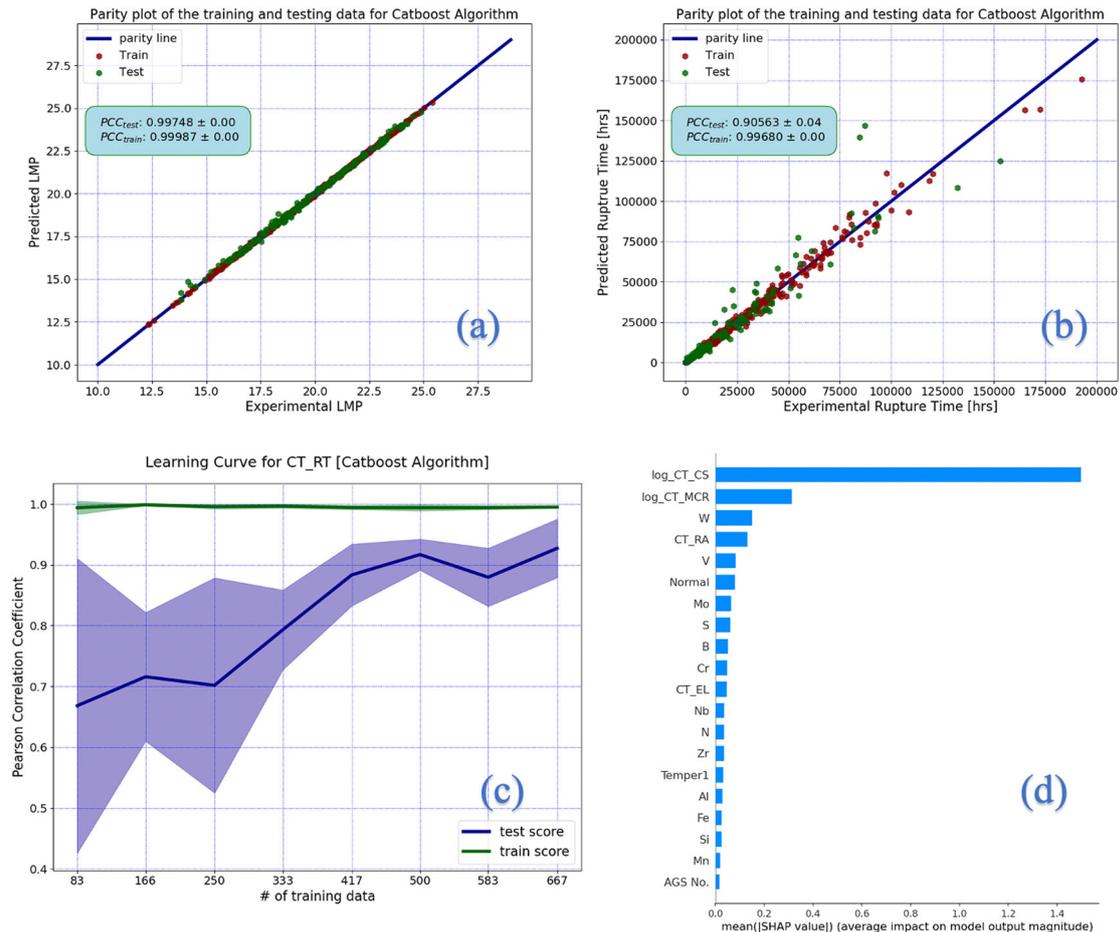


Fig. 1 Machine learning model results for 9–12% Cr dataset with Scheme-1. **a** Parity plot for the LMP prediction, **b** parity plot for the rupture life prediction, **c** learning curve for rupture life prediction, and **d** feature importance using Shapley analysis.

on the optimization, the model was capable of predicting the C -value within ± 0.68 root mean squared error (RMSE) while the testing error ranged within ± 0.80 , while the C -value ranged from 25 to 35. Considering the data available for developing the model (52 distinct alloy materials), this model was deemed accurate for future prediction of C -value for a novel alloy material. In Scheme-2, similar performance was observed for 9 to 12% Cr dataset. The correlation coefficients for modeling the LMP and rupture life of the testing set were 0.9997 ± 0.001 and 0.91 ± 0.03 , respectively. The additional parameterization of the LMP C constant showed a positive influence on the model performance, which indicates that parameterization of the constant C with first principles calculation could be a potentially important step to improve the model fidelity. In Fig. 2, the parity plots, learning curve, and feature importance bar plot are shown for Scheme-2.

Finally, the gradient boosting algorithm was trained on the 9–12% Cr data for Scheme-3, which is the direct prediction of rupture life from all the features without the intermediate model training for the LMP prediction. For Scheme-3, the model performance in terms of PCC were 0.99 ± 0.014 and 0.93 ± 0.04 , respectively, for the training and testing sets. In Fig. 3, the parity plot, learning curve, and feature importance are shown for Scheme-3. Surprisingly, Scheme-3 performed better than both Scheme-1 and Scheme-2, owing to the fact that Scheme-3 enjoys additional degrees of freedom while modeling the functional dependency of temperature on the alloys' rupture life. In contrast, for Scheme-1 and Scheme-2, the functional representation of temperature dependence is dictated by the LMP equation.

In light of these analyses, Scheme-3 was used for the austenitic stainless steel DA. In Scheme-3, from the feature importance analysis, MCR was identified as the most important feature with Cu content, stress, elongation, and temperature being the next important features. MCR is the minimum of the creep strain versus time (or creep strain) curve and typically occurs during what is defined as the secondary creep stage of an alloy⁴⁷. It takes a significantly shorter time to measure than the rupture life. The Monkman–Grant equation has long been used to predict long-term rupture life properties from a short- to medium-time creep test^{48,49}. The ML model also suggests that MCR is an essential property for creep rupture life estimation. However, unlike the Monkman–Grant equation, the model also incorporates other features, such as inclusions through materials composition of S and Cu in an indirect way. The role of S and Cu to capture the inclusion formation is also reported in a previous study⁵⁰, and this study also finds that the inclusion is a crucial feature for accurate estimation of creep rupture time; however, in the absence of direct data to incorporate inclusion, S and Cu can be used as effective surrogates to model the inclusion formation in the ML study.

For the austenitic stainless steel dataset, the PCCs obtained for the training and testing sets were 0.97 ± 0.02 and 0.83 ± 0.04 , respectively. The performance obtained was not as impressive as the 9–12% Cr family results. Two possible reasons for this are: (1) This dataset is significantly smaller than the 9–12% Cr family dataset, e.g., 9–12% Cr family dataset contains 836 data points whereas the austenitic stainless steel dataset contains only 466

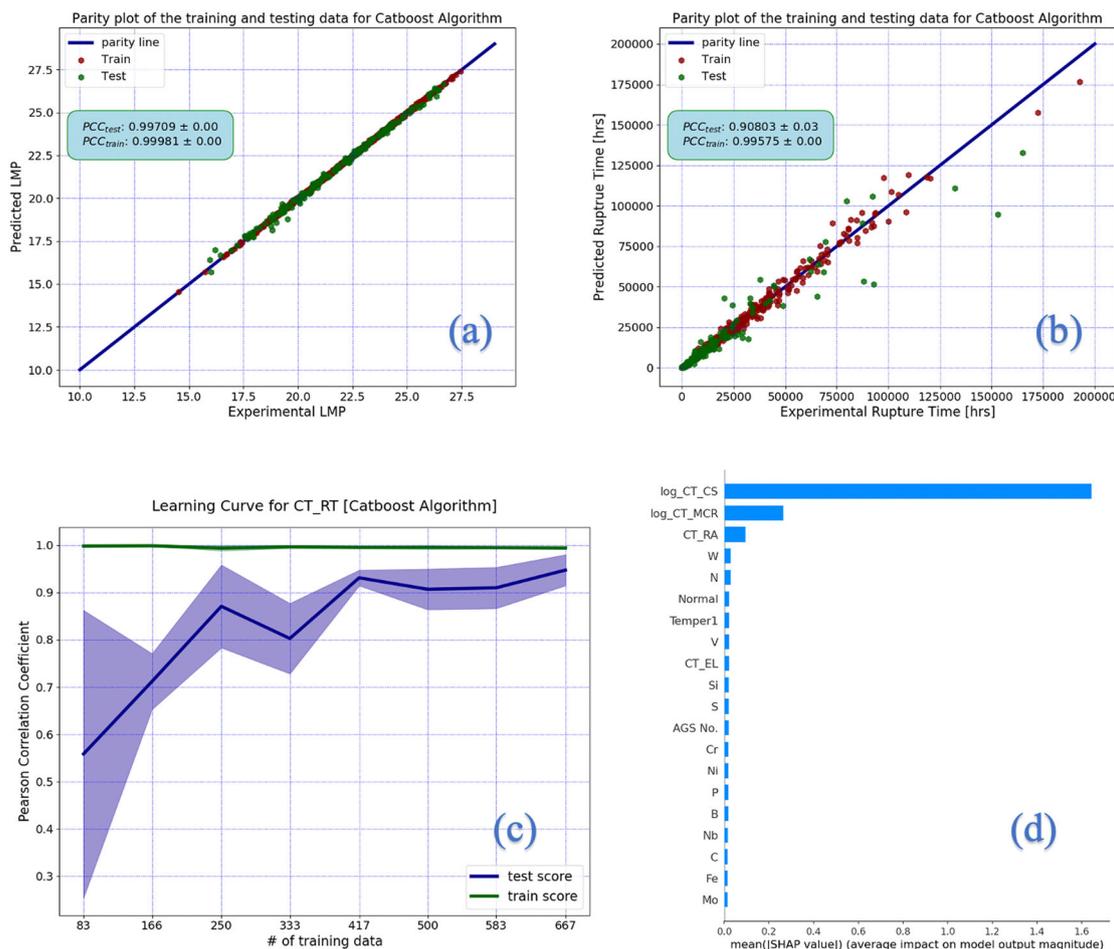


Fig. 2 Machine learning model results for 9–12% Cr dataset with Scheme-2. **a** Parity plot for the LMP prediction, **b** parity plot for the rupture life prediction, **c** learning curve for rupture life prediction, and **d** feature importance using Shapley analysis.

data points, and (2) this dataset is missing the MCR feature which is identified in the 9–12% Cr family model as an important feature. In the austenitic stainless steels dataset, a significant portion of the data points does not contain the MCR information, and as a result, this feature was completely eliminated during the preprocessing step. When MCR was eliminated from the 9–12% Cr family dataset, a decrease in ML performance was observed ($PCC_{test} = 0.86 \pm 0.02$), indicating that the inclusion of MCR is crucial in building a highly reliable ML model for rupture life prediction. From the feature importance analysis, it was found that C and B are very important features among all the chemical composition, which is expected as C is responsible for the formation of strengthening carbides and saturating the interstitial site in Fe matrix while B is effective for creep cavitation suppression by resisting grain boundary sliding⁵¹. In Fig. 4, the plots for the austenitic stainless steel dataset are shown, which includes the parity plot, learning curve, and feature importance. Finally, Table 2 summarizes the ML model results for different schemes and datasets. Next, the training of generative model for data augmentation and synthetic alloy generation is discussed.

Training a generative model

To train the VAE, both the mean squared error and the binary cross-entropy were used to compute the reconstruction loss. However, the binary cross-entropy was found to have superior performance over the mean squared error. The model was trained for 1500 epochs, although after about 1200 epochs the model

achieved convergence. In Fig. 5, the principal component analysis of the reconstructed samples and real samples for both datasets are shown. As expected, the continuous latent space was faithfully mapped, rather than the discrete space as one would expect in a vanilla autoencoder.

Generative model for novel materials discovery

In Fig. 6, the violin plot for the rupture life of the real samples and the gradient boosting model predicted rupture life of the generated samples is shown to illustrate the distribution of the data at different rupture life values. The VAE-generated samples were mainly drawn from the space where more experimental data are available, i.e., it was unable to generate alloys from the high rupture life space. Since VAE was mapped from the (learned) latent space distribution rather than the prior distribution, the generated samples may not be fully consistent with those prior. Furthermore, as opposed to a generative adversarial network, sampling using VAE is intractable due to the mismatch between the prior and posterior distribution mentioned previously. There are a few ways to improve sampling, such as Markov chain Monte Carlo (MCMC) sampling. A MCMC-based VAE is planned for future work as it is out of scope of this article.

The long-term goal in building a ML-assisted predictive model is to enable novel materials discovery at a faster rate than using traditional trial and error experimentation, or first principles computational approaches. In order to achieve that goal, a reinforcement learning setup is proposed that will enable the

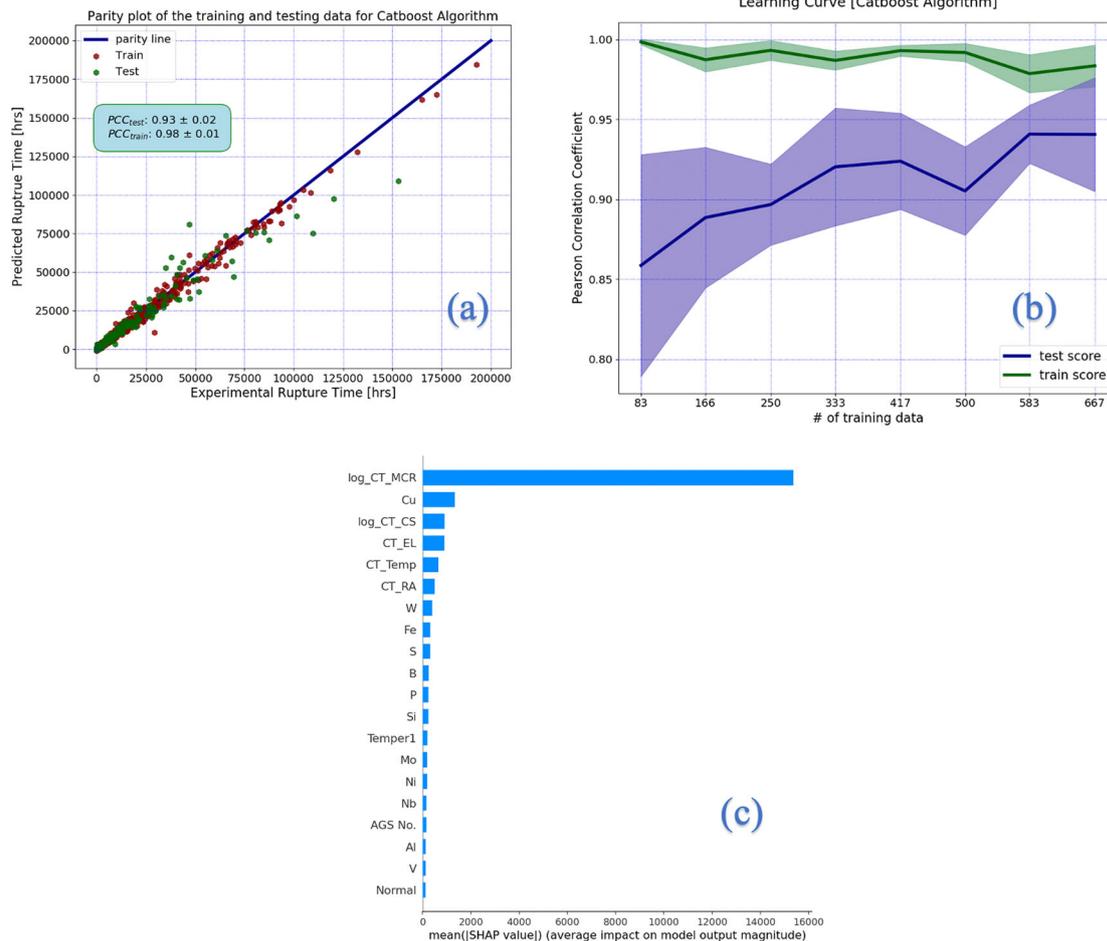


Fig. 3 Machine learning model results for 9–12% Cr dataset with Scheme-3. **a** Parity plot for the rupture life prediction, **b** learning curve for rupture life prediction, and **c** Feature importance using Shapley analysis.

inverse design of the material space. This includes estimation of the materials composition and processing parameters that will yield desired physical properties, such as rupture life or yield stress. In this approach, the alloys identified by the predictive model will be synthesized in the laboratory using high throughput methods. If the alloys have desired properties, the reinforcement learning agent will be rewarded, otherwise it will be penalized. Further, the alloy composition can be analyzed using physics-based methods to infer information regarding the microstructural phases or other attributes significant to alloy performance, thereby improving the physics-based domain knowledge-centered aspects of the design process.

Generative model for data augmentation

After successfully training the gradient boosting algorithm for rupture life prediction using both the 9–12% Cr family and austenitic stainless steel datasets, a slightly different workflow was developed as previously described in Schemes 1–3 to include synthetic data into the training set. Inclusion of generated data did not improve the performance of ML by a large margin. In most cases similar performance between models was noted, whether it included synthetic data or not, indicating the generative model was not as successful as generating samples with complete specifications (including both alloy attributes and the corresponding rupture life). As such, the gradient boosting model, already trained, should be used to predict the rupture life of the generated samples.

In summary, a ML-based predictive model has been demonstrated that faithfully estimated the rupture life of both 9–12% Cr ferritic–martensitic and austenitic stainless steels. The model showed >90% and >80% accuracy in terms of the Pearson correlation coefficient for 9–12% Cr and austenitic stainless steels, respectively. Subsequently, a VAE was built to generate synthetic alloys for both datasets. Though the impact of the augmented dataset on the ML performance was negligible, the generative model provides a means to identify hypothetical alloys that were not included in either dataset, thus allowing a wide array of candidates to be identified without first synthesizing them in the laboratory. The work shows that even with sparse data and lack of detailed microstructural information beyond austenitic grain size (AGS), ML can make interpretable predictions about creep life. As more data are added and microstructural details, such as Laves phases, carbides and inclusions, are incorporated, there is considerable promise of making reliable predictions of creep life and selecting robust alloys based on short-term tests. It is also possible to incorporate thermodynamics data and synthetic data from phase field models and crystal plasticity simulations to improve the predictive power of the model.

METHODS

Alloy data

Two well-curated datasets were used to build the ML model. The first dataset contains data on 9–12% Cr family steels and the second contains

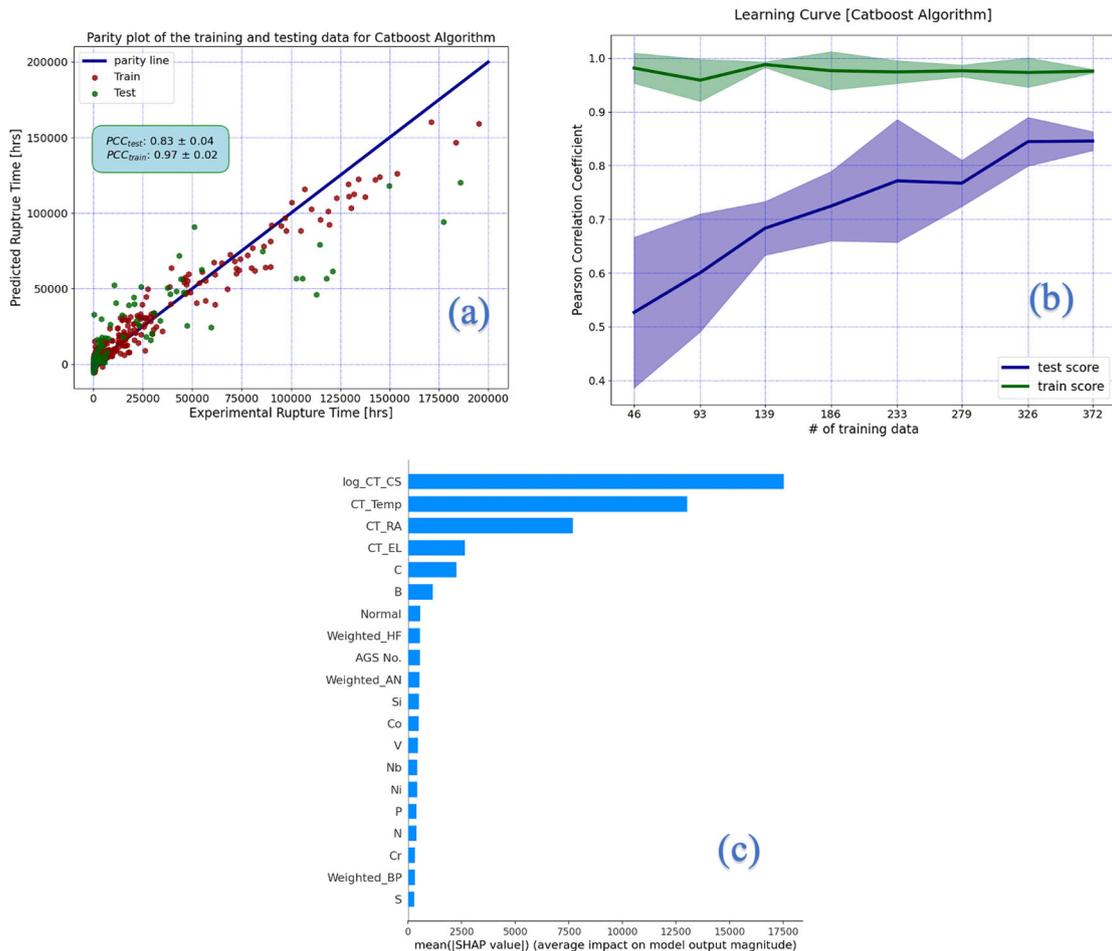


Fig. 4 Machine learning model results for the Austenitic steel dataset with Scheme-3. **a** Parity plot for the rupture life prediction, **b** learning curve for rupture life prediction, and **c** feature importance using Shapley analysis.

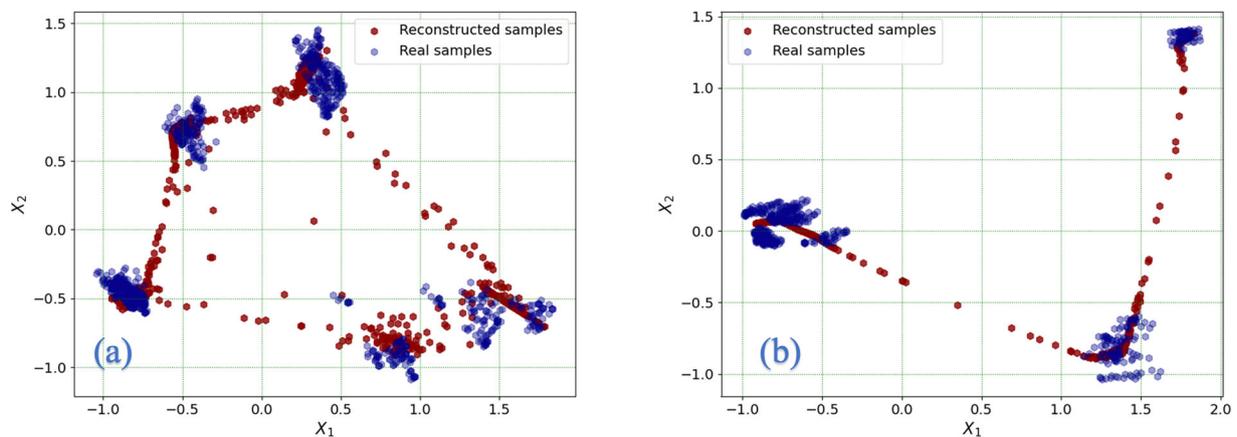


Fig. 5 Principal component analysis plot of the real samples and variational autoencoder reconstructed samples. Principal component analysis for **a** 9–12% Cr and **b** austenitic steel.

data on austenitic stainless steels. These data were collected from a variety of sources, including, but not limited to, (1) data generated from NETL's internal Advanced Alloy Development research program; (2) information contained in the NIMS data sheets^{52,53}; (3) independent research articles from the open literature; and (4) proprietary research⁵⁴ conducted by public and private research institutions. The 9–12% Cr family database

contains data on the ferritic–martensitic class of steels, which are predominantly Fe (82–90%) with 8–13% Cr and small amounts of C, Mn, Si, Ni, Co, Mo, W, Nb, Al, P, Cu, Ti, Ta, Hf, Re, V, B, N, O, S, and Zr. In total, there are 36 unique alloy composition with 836 data points, collected at different temperatures and processing conditions. The dataset contains 32 unique features, including material composition, temperature, stress, AGS,

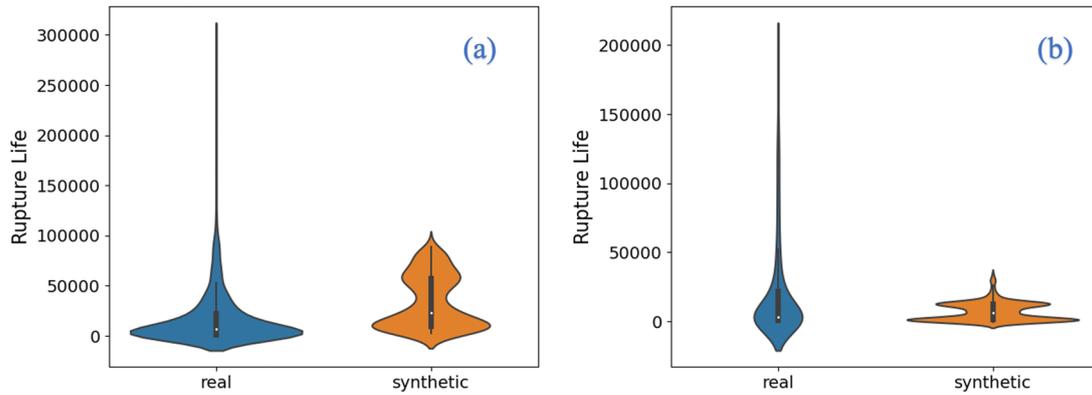


Fig. 6 Violin plot of the rupture life of the real samples and machine learning model predicted rupture life of the generated samples. **a** 9–12% Cr dataset and **b** austenitic steel dataset.

Table 1. Summary of the features present in both datasets, including feature name, description, units, mean values, and standard deviations.

| Feature name | Description | Unit | 9–12% Cr FMA | | Austenitic stainless steel | |
|--------------|---|-------------------------------------|-----------------------|-----------------------|----------------------------|-----------------------|
| | | | Mean | SD | Mean | Std. |
| Fe | Iron content | wt% | 86.70 | 2.40 | 47.88 | 26.04 |
| C | Carbon content | wt% | 0.12 | 0.04 | 0.06 | 0.02 |
| Cr | Chromium content | wt% | 10.02 | 1.48 | 17.72 | 1.54 |
| Mn | Manganese content | wt% | 0.51 | 0.11 | 1.26 | 0.63 |
| Si | Silicon content | wt% | 0.31 | 0.06 | 0.50 | 0.20 |
| Ni | Nickel content | wt% | 0.28 | 0.22 | 30.17 | 26.30 |
| Co | Cobalt content | wt% | 0.009 | 0.027 | 0.33 | 0.34 |
| Mo | Molybdenum content | wt% | 0.65 | 0.33 | 1.09 | 2.80 |
| W | Tungsten content | wt% | 0.82 | 0.84 | 0.08 | 0.23 |
| Nb | Niobium content | wt% | 0.05 | 0.03 | 0.59 | 0.34 |
| Al | Aluminum content | wt% | 0.01 | 8.69×10^{-3} | 0.08 | 0.12 |
| P | Phosphorous content | wt% | 0.01 | 6.65×10^{-3} | 0.02 | 8.11×10^{-3} |
| Cu | Copper content | wt% | 0.23 | 0.35 | 0.07 | 0.05 |
| Ti | Titanium content | wt% | 2.74×10^{-4} | 7.94×10^{-4} | 0.05 | 0.09 |
| V | Vanadium content | wt% | 0.18 | 0.07 | 0.02 | 0.02 |
| B | Boron content | wt% | 1.05×10^{-3} | 1.4×10^{-3} | 0.001 | 0.001 |
| N | Nitrogen content | wt% | 0.04 | 0.02 | 0.03 | 0.05 |
| S | Sulfur content | wt% | 3.36×10^{-3} | 3.47×10^{-3} | 8.88×10^{-3} | 5.79×10^{-3} |
| Zr | Zirconium content | wt% | 1.30×10^{-4} | 3.37×10^{-4} | | |
| Normal | Normalization or austenization heat treatment temperature | °C | 1044.46 | 29.48 | 1129.09 | 98.73 |
| Temper1 | Temper heat treatment 1 | °C | 744.58 | 56.18 | | |
| AGS No. | Austenitic grain size number | $\frac{\text{grains}}{\text{mm}^2}$ | 7.73 | 2.52 | 4.88 | 1.74 |
| CT_Temp | Creep test temperature | °C | 589 | 72 | 716 | 109 |
| CT_EL | Elongation to failure | % | 26.52 | 10.05 | 32.98 | 31.06 |
| CT_RA | Reduction in area | % | 69.50 | 24.66 | 34.02 | 24.80 |
| log_CT_CS | Logarithm of creep test stress | | 4.93 | 0.065 | 4.46 | 0.83 |
| log_CT_MCR | Logarithm of minimum creep rate | | -7.56 | 2.49 | | |

MCR, elongation, and other processing parameters. In the austenitic steel dataset, there are 466 unique data points with 28 unique features, including most of the features available in the 9–12% Cr family dataset. However, it is noteworthy that due to our preprocessing approach, some data were pruned as it did not satisfy initial requirements, e.g., features with $\geq 50\%$ missing values or with low data variation. The data statistics reported above are the final data description after the preprocessing step. Table 1 tabulates the features and nomenclature used in the ML model building, and the mean and standard deviation of the features.

Time–temperature parameter

Following the rate process theory formulation by Eyring in 1936 (ref. 55), several parametric models emerged over the years to represent the creep behavior of metals and alloys, including the LMP¹⁰, the MHP¹¹, and the Dorn–Sherby parameter⁵⁶. In this article, the LMP is used as it is widely implemented and demonstrated to be satisfactory for comparing a wide range of different classes of alloys:

$$\text{LMP} = 1 \times 10^{-3} \times T [^{\circ}\text{C}] \times (C + \log(t_{\text{rupture}}[h])) \quad (1)$$

where LMP is the Larson–Miller parameter, T is the temperature in °C, C is a constant, and t_{rupture} is the rupture time in h. Application of LMP hinges on only one constant (C) which can be readily established ($C = \frac{\text{LMP}}{T} - \log(t)$) using data at an iso-stress value. Originally, Larson and Miller also proposed the value of C to be range from 20 to 30 for many metallic materials. Herein, the value of $C = 25$ was used to define the LMP parametric relationship. Owing to the constant offset usually seen for LMP values for different classes of materials, it was hypothesized that there is an LMP master curve capable of representing all alloys with a single parametric equation. To implement that idea, a workflow was developed:

1. Build a library of C parameters based on the LMP optimization (discussed later).
2. If the C -parameter is available for the alloy, use the library value to make the rupture time prediction.
3. If the C -parameter is not available, use a Gaussian Process to approximate the C -parameter.

For the LMP optimization, a base model was selected from all available alloys. The C -value was then adjusted when the base model was applied to other alloys until sufficient accuracy was reached. This enables a unified LMP based model to be built, as well as generating a library of C -values for future prediction activities. When the C -value does not exist in the library, a Gaussian process was designed to approximate the C -value. For a single alloy sample, the procedure to find the alloy-adjusted C -value is shown in Fig. 7. Here, R^2 represents the goodness of fit and *scipy.optimize* is an optimization function in Python. At a later point, the performance of both the constant C and the alloy-adjusted C are compared in the ML model.

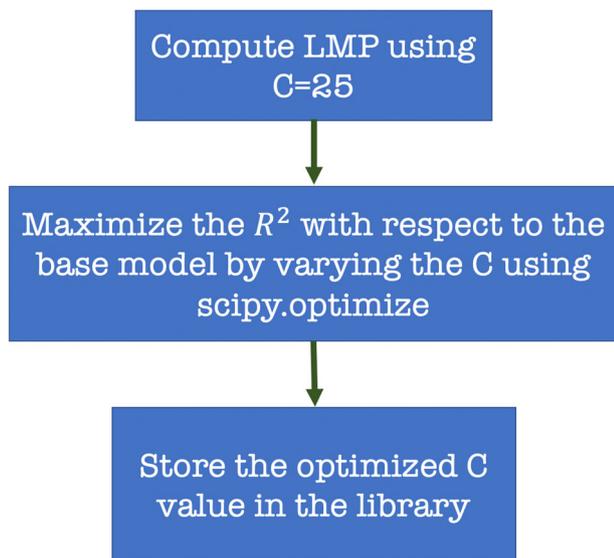


Fig. 7 The optimization scheme to build the LMP master curve. First the LMP is initialized with $C = 25$ and then it is optimized w.r.t. the master curve equation to maximize the r^2 value.

ML modeling strategy

To make reliable and interpretable predictions of creep rupture time for Fe-based 9–12% Cr ferritic–martensitic steels and austenitic stainless steels, three schemes were designed to build the predictive model for rupture life:

1. Scheme-1: First, the LMP is computed using a fixed value of C ($C = 25$) and temperature. Next, ML algorithms are trained with the remaining features to predict the LMP. Finally, the prediction is translated back to rupture life using the inverse LMP. This approach can be construed to be a hybrid one where a semi physics-based model is integrated in a brute force way into the ML model.
2. Scheme-2: Similar to Scheme-1, but instead of using a constant value of C , the alloy-adjusted C -value was fit into a universal master curve for LMP.
3. Direct prediction of creep rupture life using brute force ML algorithms.

In order to build the ML model, three different ML algorithms were used: (1) linear regression using important features, as identified by the greedy search; (2) random forest regression (RFR); and (3) gradient boosting algorithm. To implement the linear regression and random forest regression, the scikit-learn Python package⁵⁷ was used. For the gradient boosting algorithm, three different open-source libraries were employed each with slightly different implementations of the gradient boosting algorithm on decision trees (GBDT), i.e., XGBoost⁵⁸, CatBoost⁵⁹, and LightGBM⁶⁰. One aspect of the gradient boosting machine^{39,61} that makes it so powerful is that it combines several weak learners into a strong learner by iteratively dividing the data into smaller sets and focusing more on the points that were not fitted properly in the previous iteration. For ML model training, a fivefold cross-validation was used with default hyperparameters for each model. For random forest regression, $n_estimators = 100$, $max_depth = None$, and $max_features = Auto$ were used within the scikit-learn program package. For LightGBM, $boosting_type = gbdt$, $num_leaves = 1000$, $n_estimators = 100$, and $num_boost_round = 100$ were used. For XGBoost program package, $booster = gbtree$, $eta = 0.3$, $max_depth = 20$, and $num_boost_round = 500$ were used. Finally, for CatBoost predictive model generation, $learning_rate = 0.03$, $l2_leaf_reg = 3$, $depth = 6$, and $num_boost_round = 500$ were used. To compare the performance of different ML models, the Pearson correlation coefficient (PCC) between the real and predicted quantities was used as a performance metric, because PCC is a scale-invariant metric. This allows model performance comparison irrespective of the scale for both LMP and rupture life, while common metrics such as root mean squared error (RMSE) and mean absolute error (MAE) are scale-dependent metrics. In Eq. (2), the mathematical equation of PCC is shown,

$$PCC = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2(y_i - \bar{y})^2}} \quad (2)$$

where \bar{x} and \bar{y} are the mean values of the predicted and actual quantities, respectively. PCC has some very important properties which makes it easier to interpret a linear regression fit. For a perfect linear fit, the difference between the individual observation and the mean values for both the actual and predicted values will be same, leading to a PCC value of 1.0, while for perfect random correlation, the PCC value would be 0. The closer the value of PCC to 1.0, the better the linear fit of the actual and predicted values of rupture life. In order to evaluate the goodness of fit as a percentage of the actual value, mean absolute percentage error (MAPE) is also reported in the Table 2. The formula to

Table 2. Summary of the ML model results in terms of Pearson correlation coefficient (PCC) for both the training and holdout testing set for different schemes.

| | Scheme | PCC [train set] | MAPE [train set] | PCC [test set] | MAPE [test set] |
|----------------------------|--------|-----------------|------------------|----------------|-----------------|
| 9–12% Cr | 1 | 0.99 ± 0.003 | 0.06 ± 0.012 | 0.88 ± 0.038 | 0.19 ± 0.02 |
| | 2 | 0.99 ± 0.002 | 0.07 ± 0.016 | 0.91 ± 0.032 | 0.21 ± 0.02 |
| | 3 | 0.99 ± 0.012 | 0.36 ± 0.07 | 0.92 ± 0.022 | 0.44 ± 0.09 |
| Austenitic stainless steel | 3 | 0.97 ± 0.018 | 0.84 ± 0.07 | 0.83 ± 0.051 | 1.05 ± 0.44 |

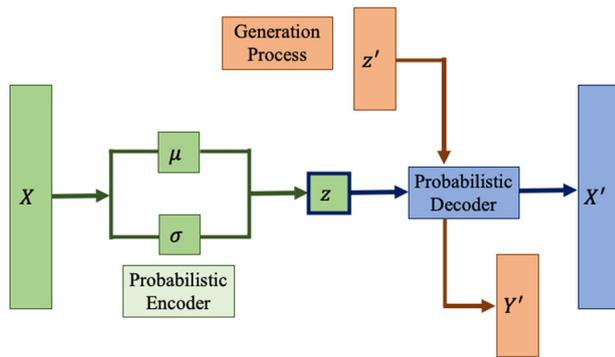


Fig. 8 Architecture of the variational autoencoder used for rupture life prediction in 9–12% Cr and austenitic steel materials. It is composed of an encoder and decoder. After the training was completed, a random sample from the distribution of z was used to generate hypothetical alloy materials.

compute MAPE is,

$$\text{MAPE} = \frac{1}{n} \sum \left| \frac{x_i - y_i}{x_i} \right| \quad (3)$$

where x_i and y_i are the actual and predicted value for the i th observation, respectively.

Synthetic data generation using a VAE

One of the major roadblocks in ML model generation is that the model performance is limited by the volume of data available for training, especially if the data generation process is very expensive and time consuming. For example, the data used in this article were collected over several decades and then painfully compiled into a consistent set of data spanning over a pre-specified composition range. However, the data that were collected are still very sparse from a ML model building perspective. To generate synthetic data, several ML algorithms, e.g., VAE⁴², generative adversarial networks⁴¹, and recurrent neural network⁶² can be used, where each have their own merits and drawbacks.

In this work, a VAE was used to generate synthetic data by sampling hypothetical alloy attributes from the learnt latent space. An autoencoder is a neural network-based model where the feature space is compressed into a latent space. In an autoencoder, a compressor/encoder and a decompressor/decoder are stacked, and the model is trained to minimize the reconstruction loss (along with other losses, i.e., regularization loss). VAE is a special kind of autoencoder where a prior distribution is enforced on the latent space. In the implementation of VAE, the negative of the Kullback–Liebler divergence between the latent space distribution and the predefined prior distribution, preferably a normal distribution $\mathcal{N}(0, 1)$, is added to the reconstruction loss. After the training process is completed, new data can be sampled from the prior distribution. However, the optimized posterior distribution may not exactly match the prior distribution, depending on the loss function value at the optimized state. This could potentially result in inefficient sampling of the latent space. In Fig. 8, a schematic depiction of VAE is outlined. First, dataset X was passed to the probabilistic encoder to get the mean (μ) and variance (σ) of the latent space distribution z . Next, an anti-symmetric encoder network was used to decode the latent space distribution z . The architecture we used to train the encoder contained two hidden layers in addition to the input layer and the output layer. The input layer was of the same size as the dataset size, and the two hidden layers had 12 and 6 neurons, respectively. Finally, two output neurons were used for both the mean and variance of the latent space. In the generation process, we sampled data points z' from the predefined prior distribution and passed it through the decoder to get the generated alloy attribute Y' . To train the VAE model, the keras⁶³ Python package with TensorFlow backend was used.

To train the ML model with the augmented dataset, the original dataset was divided into training and testing sets, and the generated data were then added to the training set to train the model. To achieve statistical significance, this process was repeated five times to mimic fivefold cross-validation. Finally, the model performance was evaluated against the withheld dataset. In Fig. 9, the ML training process using the VAE augmented data is illustrated.



Fig. 9 Schematic diagram of the construction method of the training and testing data for the Gradient Boosting Algorithm machine learning model. In every iteration, the original data are divided into train and test set. Then the generated samples are added to the train set and the GBM is trained on the combined train set and tested on the holdout test set.

DATA AVAILABILITY

The data used that support the findings of this study are available upon request to J.H. (Jeffrey.Hawk@NETL.DOE.GOV).

Received: 1 December 2020; Accepted: 12 March 2021;
Published online: 16 April 2021

REFERENCES

- Hald, J. Microstructure and long-term creep properties of 9–12% Cr steels. *Int. J. Press. Ves. Pip.* **85**, 30–37 (2008).
- Abe, F. Analysis of creep rates of tempered martensitic 9%Cr steel based on microstructure evolution. *Mater. Sci. Eng. A* **510–511**, 64–69 (2009).
- Holmström, S., Pohja, R., Nurmela, A., Moilanen, P. & Auerkari, P. Creep and creep-fatigue behaviour of 316 stainless steel. *Procedia Eng.* **55**, 160–164 (2013).
- He, J. & Sandström, R. Basic modelling of creep rupture in austenitic stainless steels. *Theor. Appl. Fract. Mech.* **89**, 139–146 (2017).
- Sakthivel, T. et al. Comparison of creep rupture behaviour of type 316L(N) austenitic stainless steel joints welded by TIG and activated TIG welding processes. *Mater. Sci. Eng. A* **528**, 6971–6980 (2011).
- Ozturk, K., Zhong, Y., Liu, Z.-K. & Luo, A. A. Creep resistant Mg-Al-Ca alloys: computational thermodynamics and experimental investigation. *JOM* **55**, 40–44 (2003).
- Levy, A. & Pifko, A. B. On computational strategies for problems involving plasticity and creep. *Int. J. Numer. Methods Eng.* **17**, 747–771 (1981).
- Kim, Y.-K., Kim, D., Kim, H.-K., Oh, C.-S. & Lee, B.-J. An intermediate temperature creep model for Ni-based superalloys. *Int. J. Plast.* **79**, 153–175 (2016).
- Agrawal, A. & Choudhary, A. Perspective: Materials informatics and big data: realization of the “fourth paradigm” of science in materials science. *APL Mater.* **4**, 53208 (2016).
- Larson, F. R. & Miller, J. *A Time-Temperature Relationship for Rupture and Creep Stresses* (1952).
- Pink, E. Physical significance and reliability of Larson–Miller and Manson–Haferd parameters. *Mater. Sci. Technol.* **10**, 340–346 (1994).
- Abdallah, Z., Gray, V., Whittaker, M. & Perkins, K. A critical analysis of the conventionally employed creep lifeing methods. *Materials (Basel)* **7**, 3371–3398 (2014).
- Mamun, O., Winther, K. T., Boes, J. R. & Bligaard, T. A Bayesian framework for adsorption energy prediction on bimetallic alloy catalysts. *npj Comput. Mater.* **6**, 177 (2020).
- Shin, D., Yamamoto, Y., Brady, M. P., Lee, S. & Haynes, J. A. Modern data analytics approach to predict creep of high-temperature alloys. *Acta Mater.* **168**, 321–330 (2019).
- Liu, Y. et al. Predicting creep rupture life of Ni-based single crystal superalloys using divide-and-conquer approach based machine learning. *Acta Mater.* **195**, 454–467 (2020).
- Mamun, O., Wenzlick, M., Hawk, J. & Devanathan, R. A machine learning aided interpretable model for rupture strength prediction in Fe-based martensitic and austenitic alloys. *Sci. Rep.* **11**, 5466 (2021).
- Wenzlick, M., Mamun, O., Devanathan, R., Rose, K. & Hawk, J. Data science techniques, assumptions, and challenges in alloy clustering and property prediction. *J. Mater. Eng. Perform.* **30**, 823–838 (2021).

18. Himanen, L., Geurts, A., Foster, A. S. & Rinke, P. Data-driven materials science: status, challenges, and perspectives. *Adv. Sci.* **6**, 1900808 (2019).
19. NIMS Creep Data Sheet, No. 48a (National Institute for Materials Science, 2012).
20. NIMS Creep Data Sheet, No. 13b (National Institute for Materials Science, 1994).
21. NIMS Fatigue Data Sheet Series, No. 30 (National Institute for Materials Science, 1982).
22. NIMS Fatigue Data Sheet Series, No. 29 (National Institute for Materials Science, 1982).
23. NIMS Fatigue Data Sheet Series, No. 35 (National Institute for Materials Science, 1983).
24. NIMS Fatigue Data Sheet Series, No. 88 (National Institute for Materials Science, 2002).
25. NIMS Creep Data Sheet, No. 43a (National Institute for Materials Science, 2014).
26. NIMS Creep Data Sheet, No. 10b (National Institute for Materials Science, 1998).
27. NIMS Creep Data Sheet, No. 51a (National Institute for Materials Science, 2013).
28. NIMS Creep Data Sheet, No. 52a (National Institute for Materials Science, 2013).
29. NIMS Creep Data Sheet, No. 46a (National Institute for Materials Science, 2005).
30. NIMS Creep Data Sheet, No. 44 (National Institute for Materials Science, 1997).
31. NIMS Creep Data Sheet, No. 19b (National Institute for Materials Science, 1997).
32. Szegedy, C. et al. Going deeper with convolutions. In *Proc. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.1–9 (Boston, MA, USA, 2015) <https://doi.org/10.1109/CVPR.2015.7298594>.
33. Choudhary, K., DeCost, B. & Tavazza, F. Machine learning with force-field-inspired descriptors for materials: fast screening and mapping energy landscape. *Phys. Rev. Mater.* **2**, 83801 (2018).
34. de Jong, M. et al. A statistical learning framework for materials science: application to elastic moduli of k-nary inorganic polycrystalline compounds. *Sci. Rep.* **6**, 34256 (2016).
35. Doshi-Velez, F. & Kim, B. Towards a rigorous science of interpretable machine learning. Preprint at <https://arxiv.org/abs/1702.08608> (2017).
36. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **1**, 206–215 (2019).
37. Shapley, L. S. & Roth, A. E. *The Shapley Value: Essays in Honor of Lloyd S. Shapley* (Cambridge University Press, 1988).
38. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. In *Proc. 31st International Conference on Neural Information Processing Systems (NIPS'17)*. 4768–4777 (Curran Associates Inc., Red Hook, NY, USA, 2017).
39. Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).
40. Zunger, A. Inverse design in search of materials with target functionalities. *Nat. Rev. Chem.* **2**, 121 (2018).
41. Goodfellow, I. NIPS 2016 tutorial: generative adversarial networks. Preprint at <https://arxiv.org/abs/1701.00160> (2016).
42. Kingma, D. P. & Welling, M. Stochastic gradient VB and the variational auto-encoder. In *Second International Conference on Learning Representations, ICLR*. Vol. 19 (2014).
43. Sanchez-Lengeling, B. & Aspuru-Guzik, A. Inverse molecular design using machine learning: generative models for matter engineering. *Science* **361**, 360–365 (2018).
44. Ferguson, A. L. Machine learning and data science in soft materials engineering. *J. Phys. Condens. Matter* **30**, 43002 (2017).
45. Shin, H.-C. et al. Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In *Simulation and Synthesis in Medical Imaging. SASHIMI 2018. Lecture Notes in Computer Science*. Vol. 11037 (eds Gooya, A., Goksel, O., Oguz, I., & Burgos, N.) 1–11 (Springer, Cham, 2018). https://doi.org/10.1007/978-3-030-00536-8_1.
46. Mikołajczyk, A. & Grochowski, M. Data augmentation for improving deep learning in image classification problem. In *2018 International Interdisciplinary PhD Workshop (IIPHDW)*. 117–122 (IEEE, Świnoujście, Poland, 2018). <https://doi.org/10.1109/IIPHDW.2018.8388338>.
47. Dieter, G. E. & Bacon, D. J. *Mechanical Metallurgy*. Vol. 3 (McGraw-Hill, 1986).
48. Radhakrishnan, V. M. The relationship between minimum creep rate and rupture time in Cr-Mo steels. *J. Mater. Eng. Perform.* **1**, 123–128 (1992).
49. Klueh, R. L. & Nelson, A. T. Ferritic/martensitic steels for next-generation reactors. *J. Nucl. Mater.* **371**, 37–52 (2007).
50. Eggeler, G., Earthman, J. C., Nilsvang, N. & Ilchner, B. Microstructural study of creep rupture in a 12% chromium ferritic steel. *Acta Metall.* **37**, 49–60 (1989).
51. Laha, K., Kyono, J. & Shinya, N. Some chemical and microstructural factors influencing creep cavitation resistance of austenitic stainless steels. *Philos. Mag.* **87**, 2483–2505 (2007).
52. Sawada, K. et al. Catalog of NIMS creep data sheets. *Sci. Technol. Adv. Mater.* **20**, 1131–1149 (2019).
53. Yagi, K. Acquisition of long-term creep data and knowledge for new applications. *Int. J. Press. Vessel. Pip.* **85**, 22–29 (2008).
54. Krishnamurthy, N. et al. *Data Analytics for Alloy Qualification*. <https://www.osti.gov/biblio/1456238> (2018).
55. Eyring, H. Viscosity, plasticity, and diffusion as examples of absolute reaction rates. *J. Chem. Phys.* **4**, 283–291 (1936).
56. Sherby, O. D. & Dorn, J. E. Creep correlations in alpha solid solutions of aluminum. *JOM* **4**, 959–964 (1952).
57. Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
58. Chen, T., He, T., Benesty, M., Khotilovich, V. & Tang, Y. Xgboost: extreme gradient boosting. R Packag. version 0.4-2 1–4 (2015).
59. Dorogush, A. V., Ershov, V. & Gulin, A. CatBoost: gradient boosting with categorical features support. Preprint at <https://arxiv.org/abs/1810.11363> (2018).
60. Ke, G. et al. Lightgbm: a highly efficient gradient boosting decision tree. In *Proc. International Conference on Neural Information Processing Systems (NIPS'17)*. 3149–3157 (Curran Associates Inc., Red Hook, NY, USA, 2017).
61. Friedman, J., Hastie, T. & Tibshirani, R. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *Ann. Stat.* **28**, 337–407 (2000).
62. Gregor, K., Danihelka, I., Graves, A., Rezende, D. & Wierstra, D. DRAW: a recurrent neural network for image generation. In *Proceedings of the 32nd International Conference on Machine Learning, PMLR* **37**, 1462–1471. <http://proceedings.mlr.press/v37/gregor15.html> (2015).
63. Chollet, F. Keras: The python deep learning library. *ascl* ascl-1806 (2018).

ACKNOWLEDGEMENTS

This work was supported by the U.S. Department of Energy, Office of Fossil Energy, eXtreme environment MATerials (XMAT) consortium. This research used resources of the Pacific Northwest National Laboratory, which is supported by the U.S. Department of Energy.

AUTHOR CONTRIBUTIONS

R.D. and O.M. conceived the study. J.H. provided the dataset. O.M. performed machine learning training. O.M., M.W., and A.S. analyzed the data. O.M. and M.W. drafted the manuscript. J.H., R.D., and A.S. reviewed the manuscript.

COMPETING INTERESTS

The authors declare that they have no known competing or non-competing financial interests that could have appeared to influence the work reported in this paper. This manuscript was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference therein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed therein do not necessarily state or reflect those of the United States Government or any agency thereof.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to O.M.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons

Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© Battelle Memorial Institute 2021