Building a fatigue strength predictor for steel using an ensemble deep learning model

Chirag R Bhattad | MM14B051

Department of Metallurgical and Materials Engineering, Indian Institute of Technology Madras, Chennai 600036, INDIA

Introduction

- Material Science has traditionally been pushed forward by Experimental studies, Mathematical modelling and Computer simulations.
- This has generated lots of data which has propelled data-driven science and discovery, ushering in the fourth paradigm for Material Science.
- Sophisticated modelling techniques are employed to unlock the processing-structure-property-performance (PSPP) correlations.[2]
- Digitization of the microstructure data coupled with recording of steelmaking process data makes a database which helps analyse the PSPP correlations.
- This facilitated the creation of a fatigue strength calculator using ensembling techniques, which unlocked the PSPP correlations.[1]

Modelling procedure

The fatigue strength dataset has 437 data-points and 25 features. The dataset is extracted from National Institute of Material Science (NIMS), Japan.

Table 1: List of features in the dataset[4]

Abbreviation	Details
NT	Normalizing Temperature
THT	Through Hardening Temperature
THt	Through Hardening time
THQCr	Cooling rate for Through Hardening
CT	Carburization Temperature
Ct	Carburization time
DT	Diffusion Temperature
Dt	Diffusion time
QmT	Quenching media Temperature (for carburization)
TT	Tempering Temperature
Tt	Temperating time
TCr	Cooling rate for Tempering
C	% Carbon
Si	% Silicon
Mn	% Manganese
P	% Phsophorous
S	% Sulphur
Ni	% Nickel
Cr	% Chromium
Cu	% Copper
Mo	% Molybdenum
RedRatio	Reduction Ratio (Ingot to Bar)
dA	Area Proportion of Inclusions Deformed by Plastic Work
dB	Area Proportion of Inclusions Occurring in Discontinuous Array
dC	Area Proportion of Isolated Inclusions
Fatigue	Rotating Bending Fatigue Strength (10^7 Cycles)

Table 2: List of Machine learning models employed with their accuracy[3]

Model name	$r^2 score$ for Training data	$r^2 score$ for Test data
SVM regressor	0.964832	0.956210
Linear regression	0.972629	0.966013
Polynomial regression	0.972629	0.966013
Decision Tree regressor	0.992247	0.966527
Neural Network	0.983781	0.974342
Random Forest regressor	0.996656	0.977406
Gradient Boosting regressor	0.999802	0.979231
ADA Boost regressor	0.974254	0.974501
LightGBM regressor	0.997507	0.984731
XGBoost regressor	0.998625	0.983021

Table 3: Feature ranking by the tree-based Machine learning models[3]

S. No.	Model	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
1	Decision Tree regressor	Dt	Cr	C	TT	NT
2	ADA Boost regressor	TT	Cr	Tt	Dt	Ct
3	Random Forest regressor	QmT	Ct	Cr	NT	Tt
4	LightGBM regressor	TT	C	Cr	P	Mn
5	XGBoost regressor	TT	C	Cr	Mn	P

Using the feature rankings, the top five features have been selected. The dataset is now reduced to these five features and the Machine learning models are recalibrated to gauge their effectiveness in modelling the dataset.

Table 4: List of features selected[3]

Top 5 features selected
Tempering Temperature (TT)
% Carbon (C)
% Chromium (Cr)
% Manganese (Mn)
% Phosphorous (P)

Table 5: Accuracy of Machine learning models for just five features[3]

Model name	$r^2 score$ for Training data	$r^2 score$ for Test data
SVM regressor	0.497909	0.209385
Linear regression	0.617179	0.528545
Polynomial regression	0.976667	0.0.961134
Decision Tree regressor	0.989510	0.962687
Neural Network	0.945696	0.937432
Random Forest regressor	0.991512	0.963090
Gradient Boosting regressor	0.997648	0.976408
ADA Boost regressor	0.959724	0.946410
LightGBM regressor	0.994032	0.930288
XGBoost regressor	0.996073	0.979749

To improve the accuracy and preserve the variance and trends in the dataset, more than one models are chosen to create an Ensemble deep learning model.

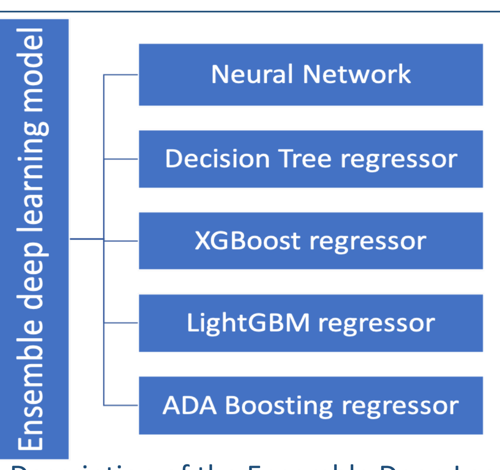


Fig. 1: Description of the Ensemble Deep Learning Model

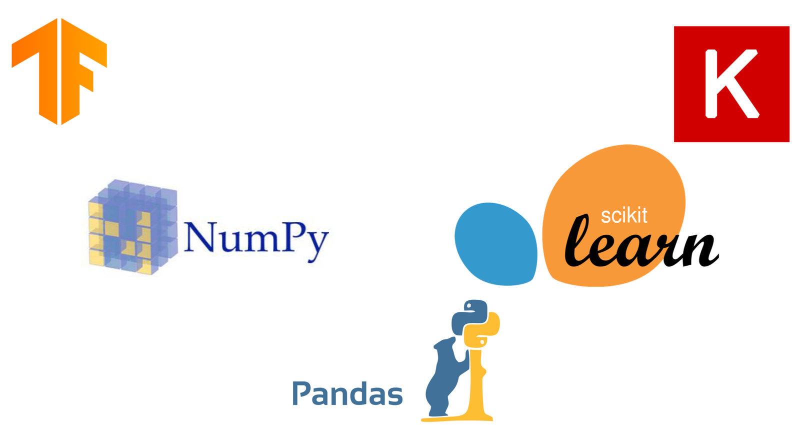


Fig. 2: Python libraries used to implement the code

Conclusions

- The fatigue strength calculator, hosted using Heroku, a python-based application, is free to use and open-for-all.
- Feature reduction is employed to reduce the number of inputs needed for calculating the fatigue strength, which was facilitated by the use of an ensemble deep learning model.
- The use of *pickle*, a python-based library, converts the trained model into a compressed package, which is hosted on the backend of the webpage, ready to predict on the values input by the user.
- The dataset used for this study is extremely small, which affects the accuracy of the calculator. This is because the outliers, which could be genuine data-points, are neglected by the model to better fit the dataset, reducing the generality of the calculator.

Results

Data Exploration

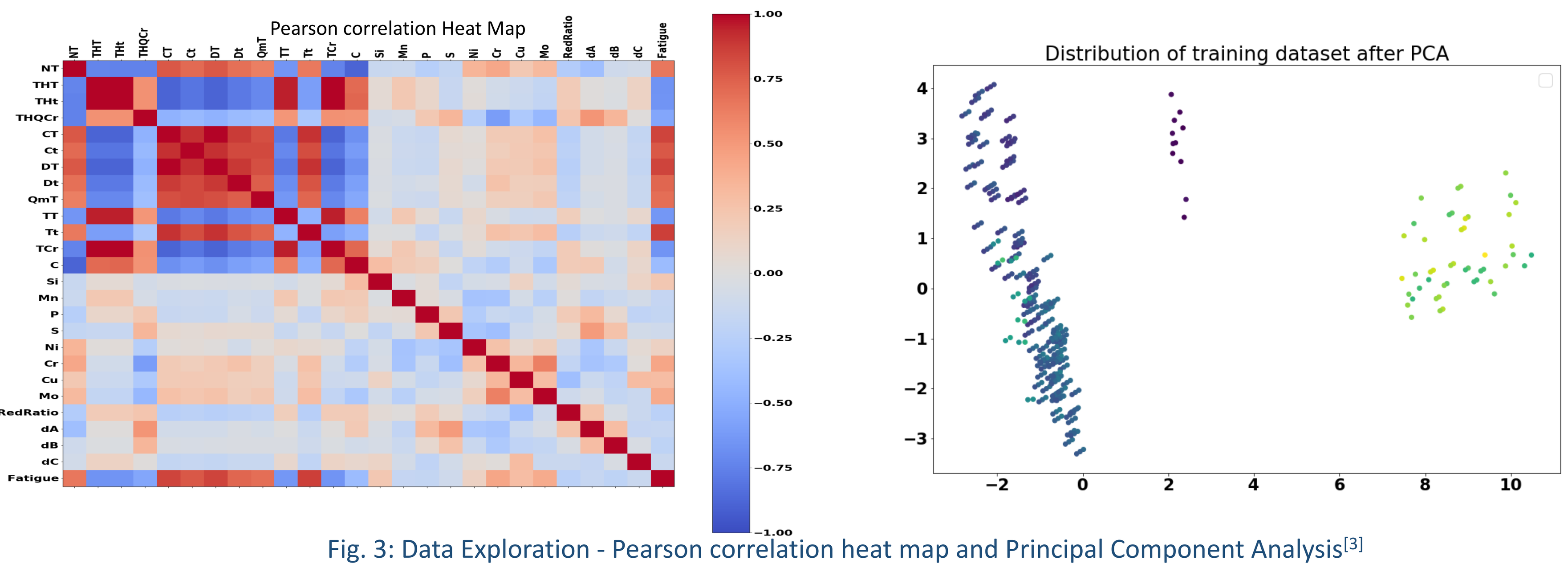
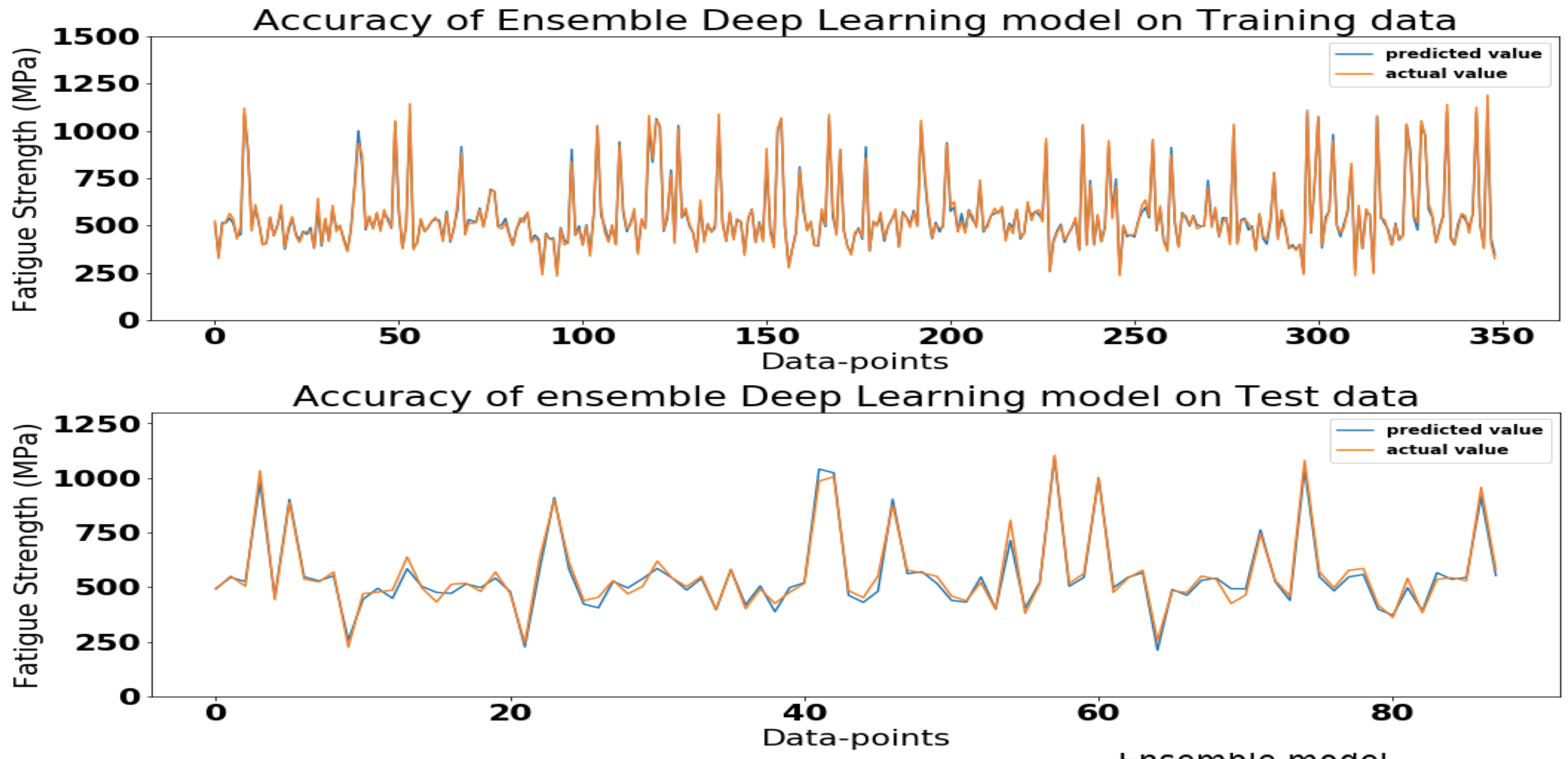


Fig. 3: Data Exploration - Pearson correlation heat map and Principal Component Analysis[3]

Ensemble Deep Learning model



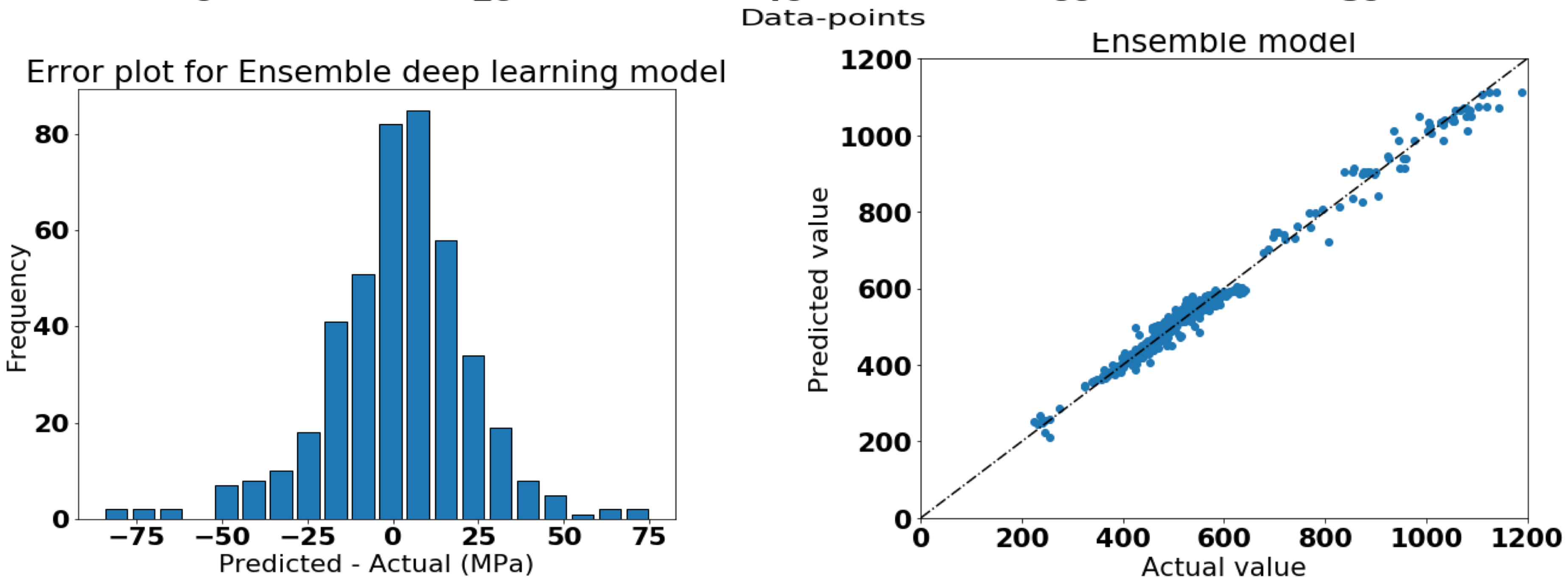


Fig. 4: Analysis of the Ensemble deep learning model[3]

Table 6: Accuracy of the Ensemble Deep Learning Model[3]

Accuracy of Ensemble deep Learning model	$r^2 score$
Training Data	0.989414
Test Data	0.975358

Fatigue strength calculator

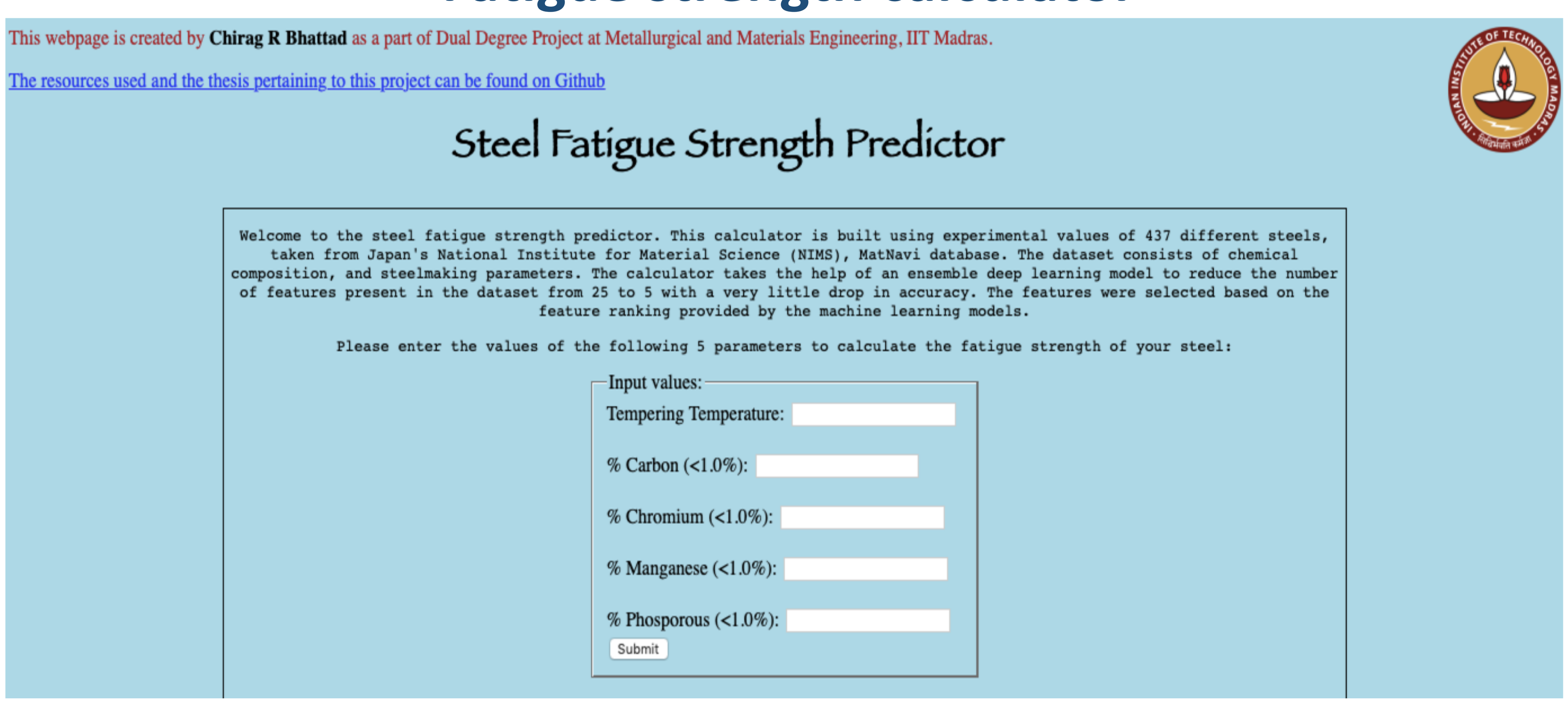


Fig. 5: Snapshot of the fatigue strength calculator hosted at www.fatigue-strength-model.herokuapp.com

References

1. **Agrawal, A., P. D. Deshpande, A. Cecen, G. P. Basavarsu, A. N. Choudhary,** and **S. R. Kalidindi** (2014*a*). Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters. *Integrating Materials and Manufacturing Innovation*, **3**(1), 8. ISSN 2193-9772.
2. **Bessa, M., R. Bostanabad, Z. Liu, A. Hu, D. W. Apley, C. Brinson, W. Chen,** and **W. Liu** (2017). A framework for data-driven analysis of materials under uncertainty: Countering the curse of dimensionality. *Computer Methods in Applied Mechanics and Engineering*, 320, 633 – 667. ISSN 0045-7825.
3. **Bhattad, Chirag R**. "Fatigue-strength-calculator." *GitHub*. February 04, 2019.
4. NIMS Materials Database (MatNavi).