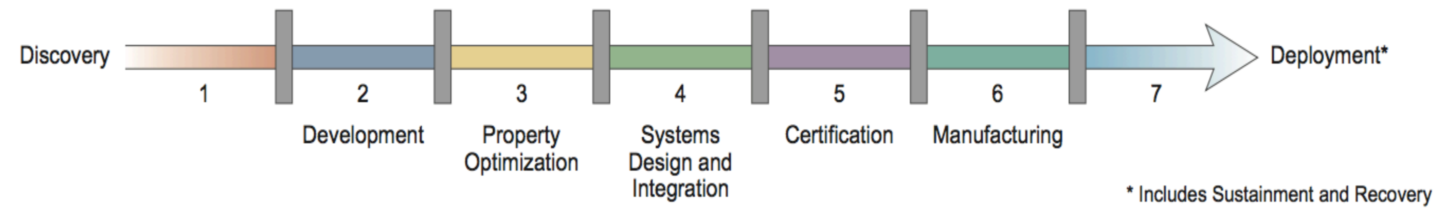
Feedstock Optimization Using Machine Learning Techniques

G. Jaman[1], D. Loveland[1], B. Gallagher[2], T. Y. Han[1]

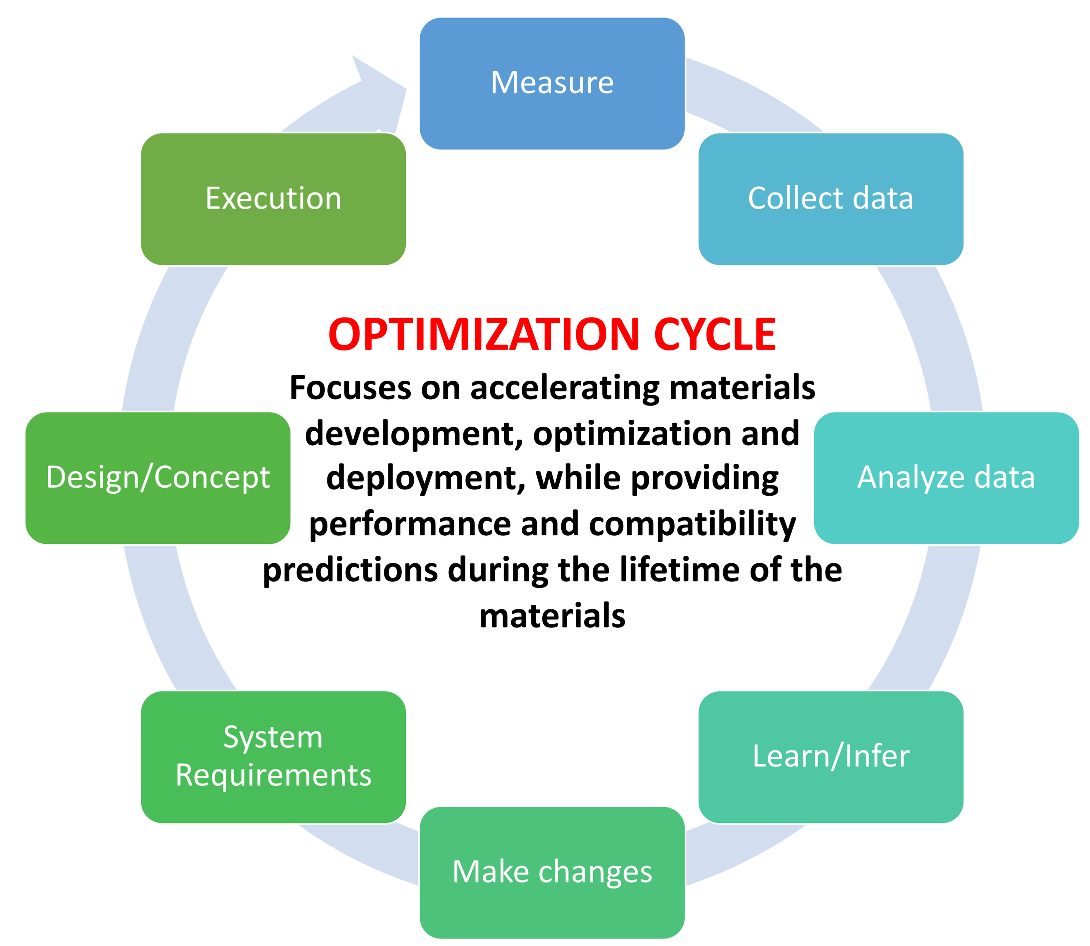
[1]Materials Science Division, PLS [2]Center for Applied Scientific Computing

ABSTRACT

Traditional experiments and computation modelling often consume tremendous time and resources. A typical time frame for materials development from initial research state to commercial use can take 10-20 years. It is crucial to develop methods for accelerating the discovery and design process of novel materials. **This poster addresses automating some expensive traditional approach.** Machine learning has received significant attention in materials discovery, development, optimization and design.



- ***Features from SEM images and human assessed data are used to predict the peak stress, the peak strain and the initial slope.***
- **Experiments suggests, domain knowledge based features captured from the SEM images aids prediction accuracy.**
- **Binarized Statistical Image Feature (BSIF) is a hand engineered feature that is motivated through domain knowledge.**
- **Fusion models can lead to improve accuracy with decrease in error of predicted performance parameters in the intermediate stage.**



APPROACH

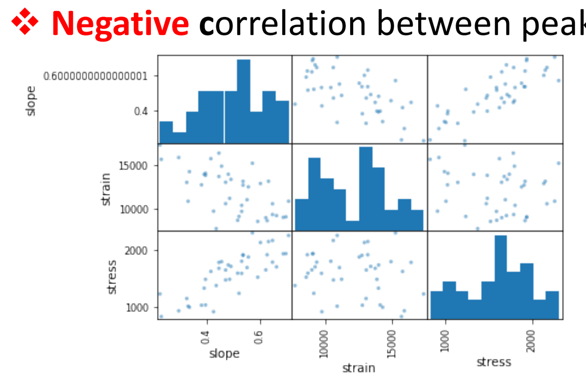
BSIF motivated through domain knowledge, applied in the experiments in order to analyze the domain knowledge chosen method over Deep Learning on images, which automates feature extraction. Hand engineered features that target specific attributes (texture) help to gain a level of interpretability that can be lost in deep learning feature extraction. Followings are key steps considered:

- *Correlation among performance parameters evaluated.*
- *Linear Regression and Random Forest prediction performance compared*
- *Analysis of prediction performance across different data types and fusion models*
- *Pseudo error assessment to seek motivation in reducing error from scalar data-based ML models*
- *Heading towards Deep learning*

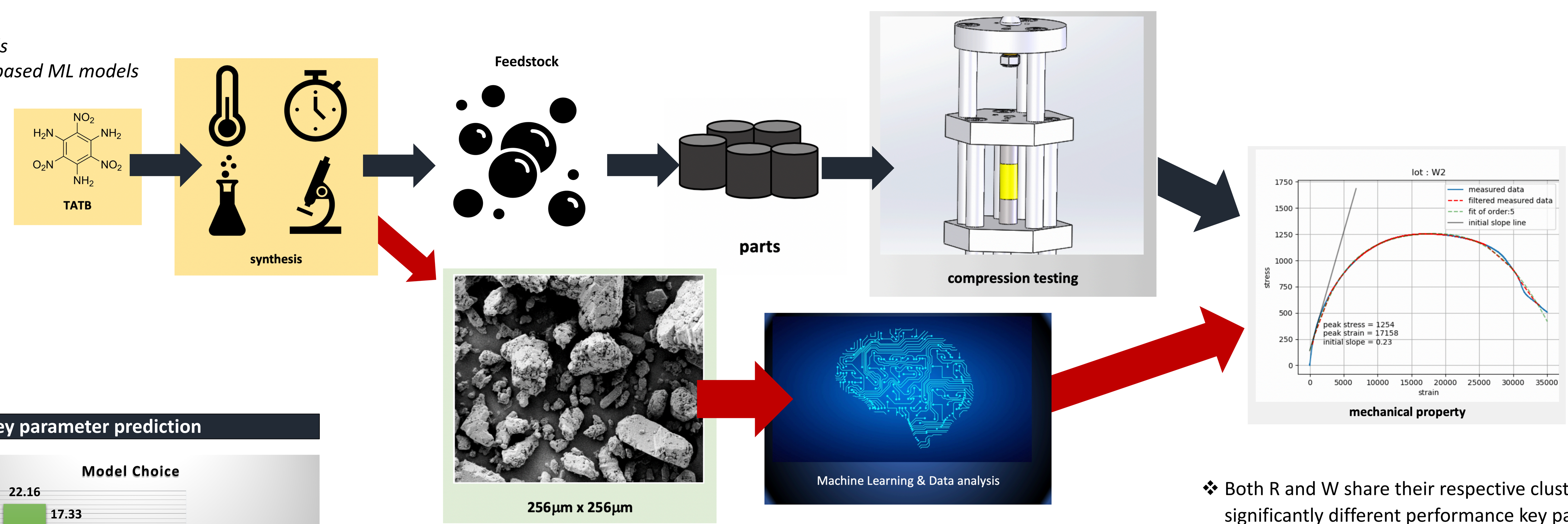
Linear correlation of performance parameters

Dataset of mechanical properties derived from compression test carries

- **Positive** correlation between peak stress and initial slope
- **Negative** correlation between peak strain and initial slope



	slope	strain	stress
slope	1.000000	-0.617616	0.811004
strain	-0.617616	1.000000	-0.162001
stress	0.811004	-0.162001	1.000000



Stages involved in extracting performance parameters

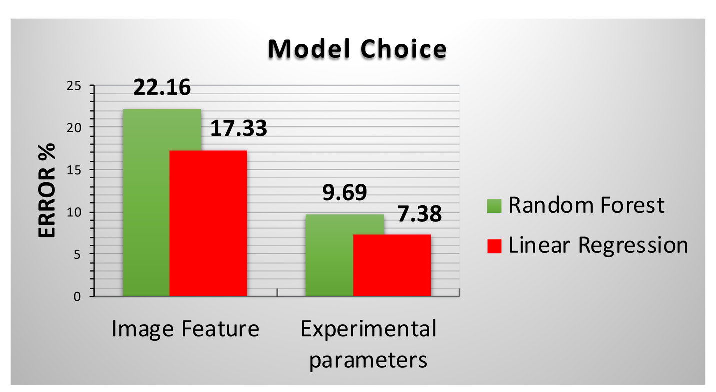
Linear regression generates lower error relative to Random Forest in key parameter prediction

Performance parameter prediction from scalar data

- Linear Regression (LR) showed greater accuracy than Random Forest (RF)
- Result further supports correlation table

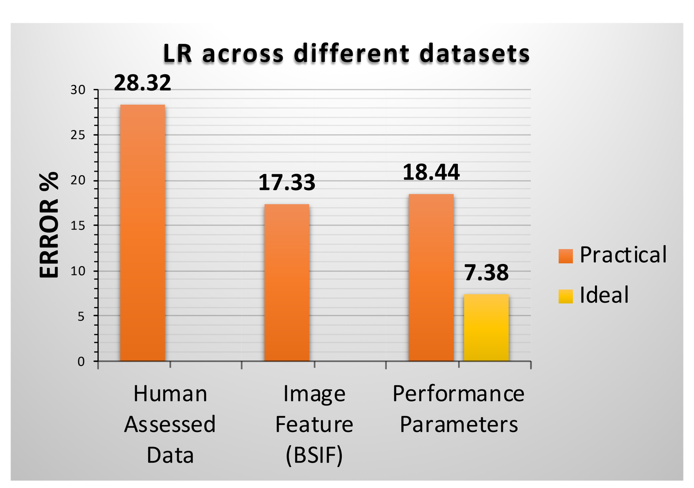
Performance parameter prediction from image feature

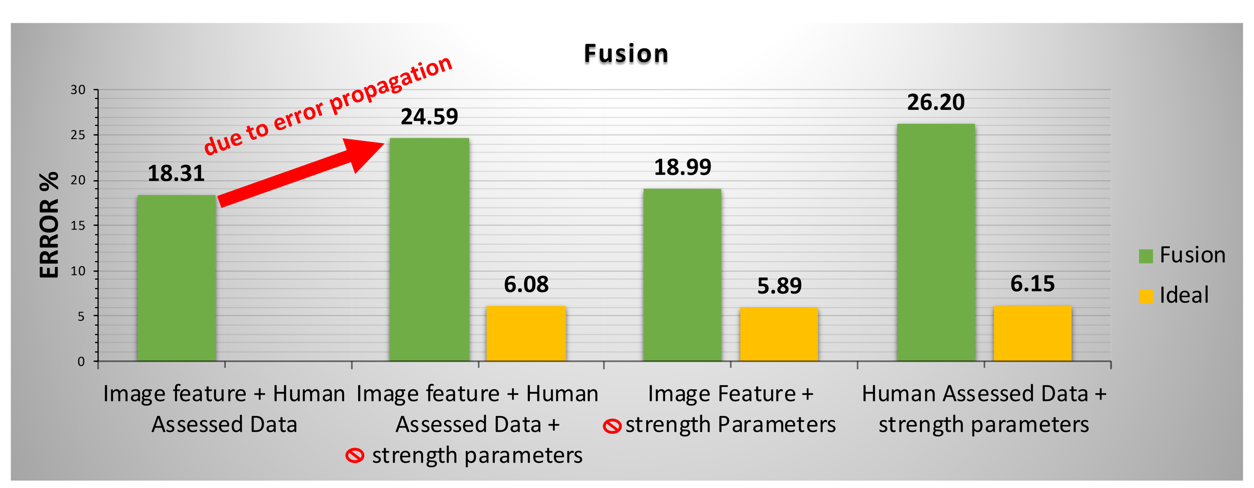
- Linear Regression generated lower percentage error relative to Random Forest in case of strain and slope prediction.

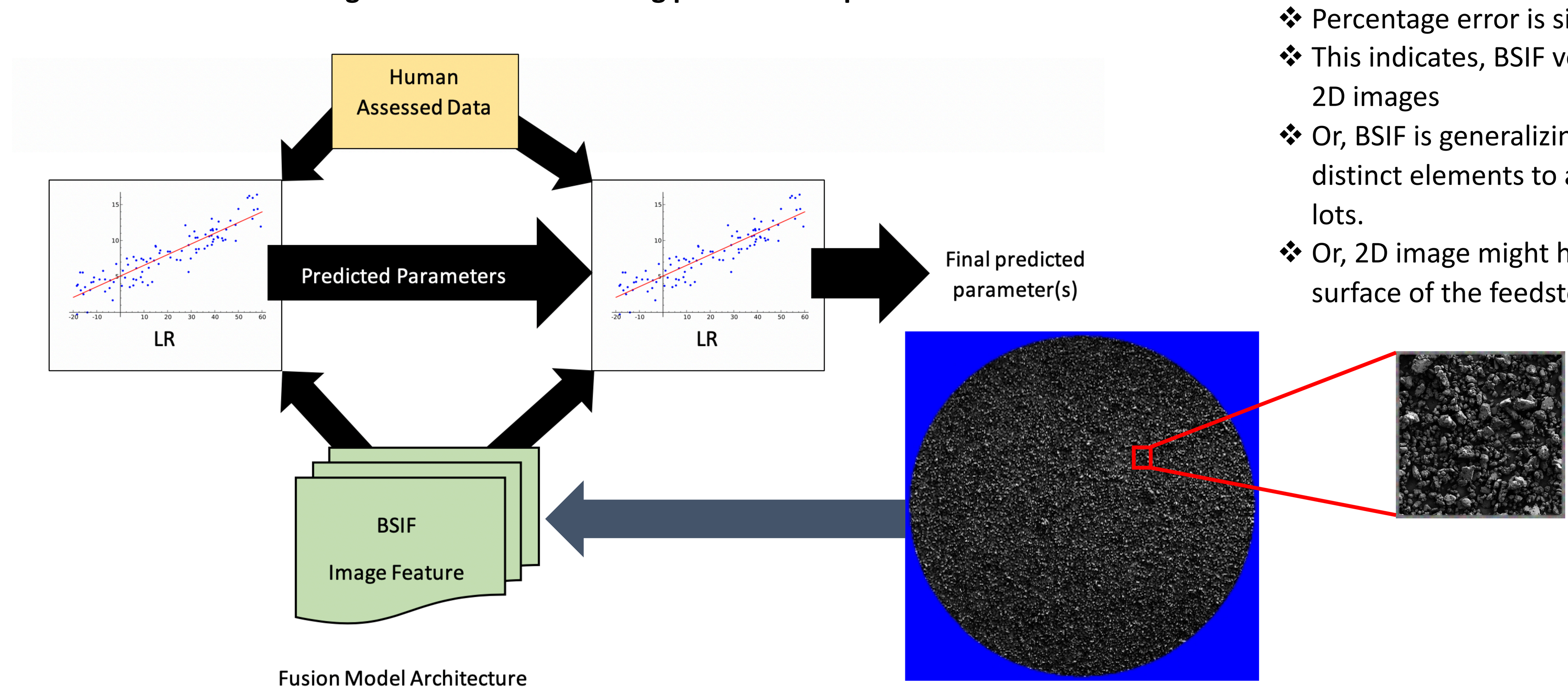


Fusion model performance depends on prediction accuracy of individual datasets

Binarized Statistical Image Feature (BSIF) is fed to multiple standard and fusion models. Human assessed parameters are also applied to observe any growth in learning.

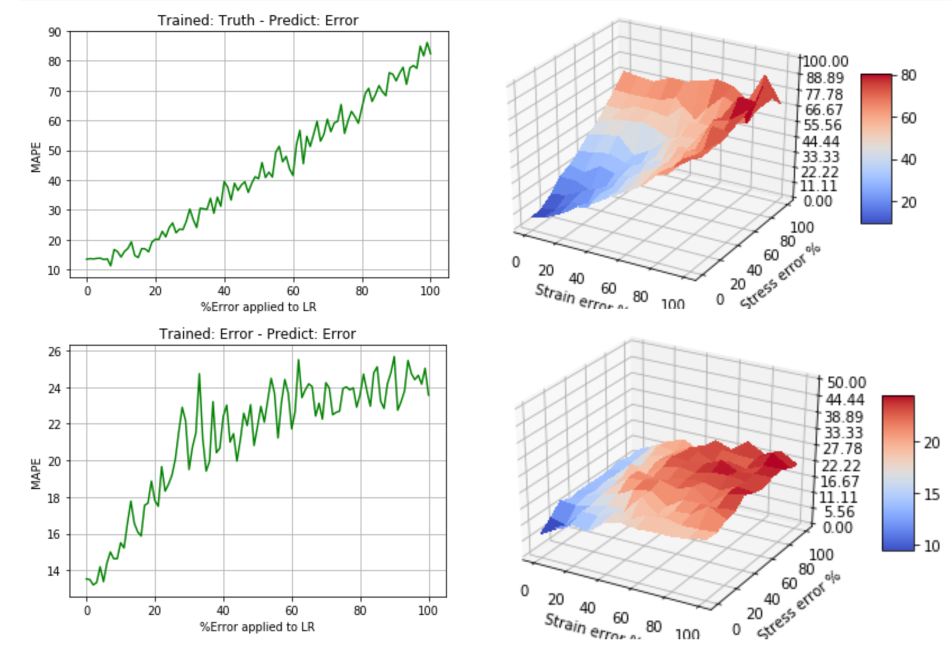






Fusion Model Architecture

Reduction in performance parameter prediction error would lead to greater accuracy in fusion models



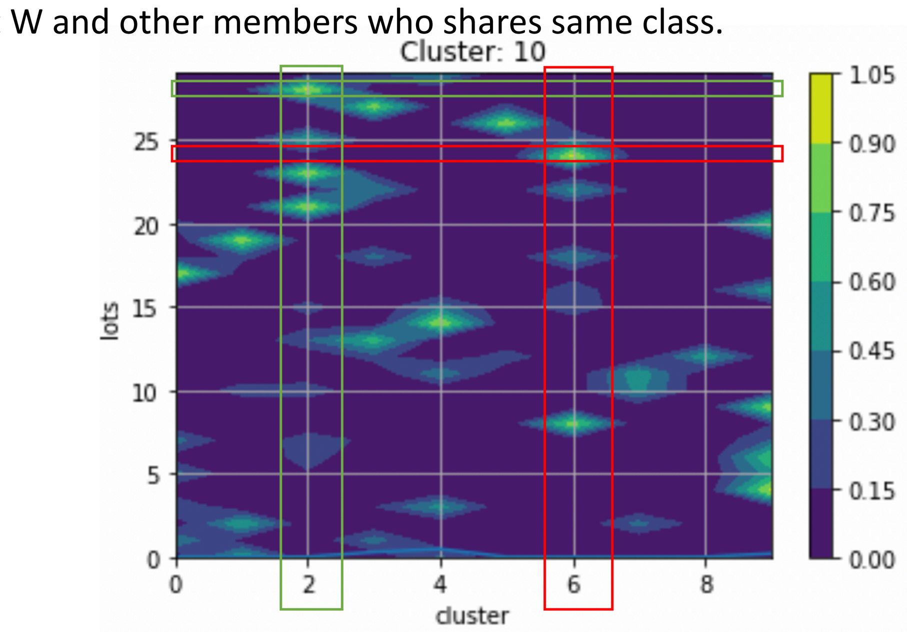
Since multiple models are using the predicted parameters, it is important to observe how LR model performs during inference of noisy input. This assessment is conducted using four different configurations as follows:

- Trained with ground truth input with noise applied to peak stress and strain during inference (2D)
- Trained with noisy input along with noise applied to peak stress and strain during inference (2D)
- Trained with ground truth input with noise applied to peak stress during inference (1D)
- Trained with noisy input along with noise applied to peak stress during inference (1D)

Predicted stress and strain from ML models explored so far falls within ~16% to ~24% in terms of percentage error. The pseudo error plots indicate that reducing predicted error by any amount to 20% or less would reduce inference error linearly.

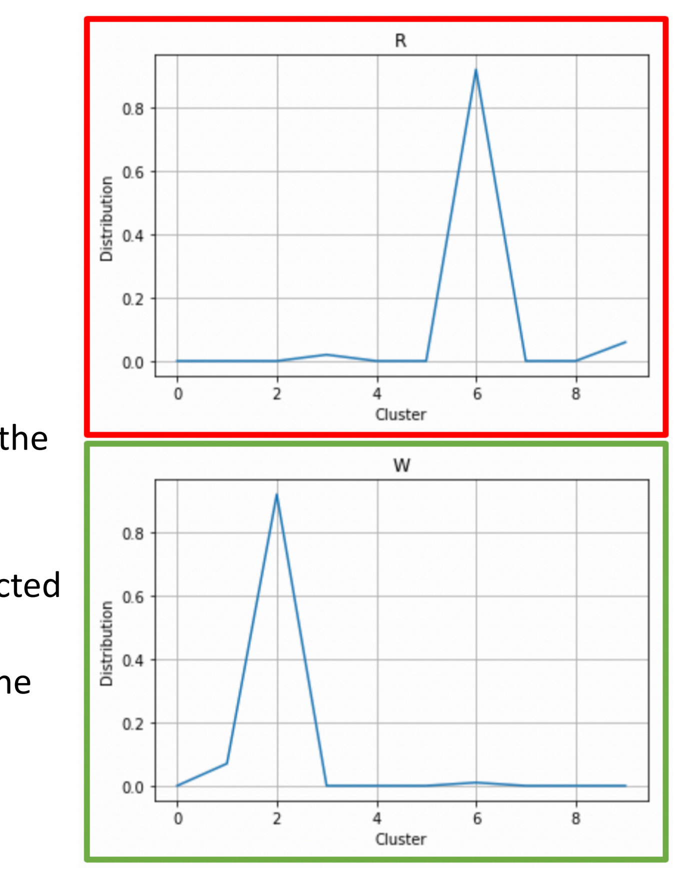
Is BSIF missing any crucial content?

Gaussian Mixture Model (GMM) with varying class size is explored to spot any pattern in classifying lots. The filled contour shows 10 class 30 lots presentation of BSIF data. Each lot is addressed from the range of 0 to 29. Red rectangles points at lot R and other members who share same class. Similarly, green rectangles points at lot W and other members who shares same class.



Gaussian mixture model to visualize classification probability

- Both R and W share their respective cluster with lots containing significantly different performance key parameter from ground truth perspective but close in predicted readings.
- Percentage error is significantly high relative to other lots.
- This indicates, BSIF vector either lacking some other key details from the 2D images
- Or, BSIF is generalizing image content to an extend that there are no distinct elements to aid learning when it comes to some poorly predicted lots.
- Or, 2D image might have missed some key feature that lies beneath the surface of the feedstock.



Gaussian mixture model classification probability on single lot

To address source of error better, it would be interesting to explore Deep Learning (DL) approach on entire image to see if we need more than the image of surface content of the feedstock.

Lawrence Livermore National Laboratory
Idaho State University
LDRD Innovation for the Nation
U.S. DEPARTMENT OF ENERGY
National Nuclear Security Administration
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.
LLNL-POST-783678