Quantifying Uncertainty in Ensemble Deep Learning

Emily Diegel: diegele@my.erau.edu
Rhiannon Hicks: hicksr10@my.erau.edu
Max Prilutsky: mdp7200@gmail.com
Rachel Swan: swanr3@my.erau.edu
Project Advisor: Dr. Mihhail Berezovski

Abstract

Past research has utilized machine learning on experimental data in the material sciences and chemistry field to predict properties of metal oxides. Neural networks can determine underlying optical properties in complex images of metal oxides and capture essential features which are unrecognizable by observation. However, neural networks are often referred to as a "black box algorithm" due to the underlying process during the training of the model. Building ensemble neural networks allows for the analysis of the error bars of the prediction model. The objective is to determine the comparative differences between the predictive ability of each individual neural network versus the ensemble neural network. Overall, ensemble neural networks outperform singular networks and demonstrate areas of uncertainty and robustness in the model.

Data

Fig. 1 Absorption Spectroscopy Data for Metal Oxides

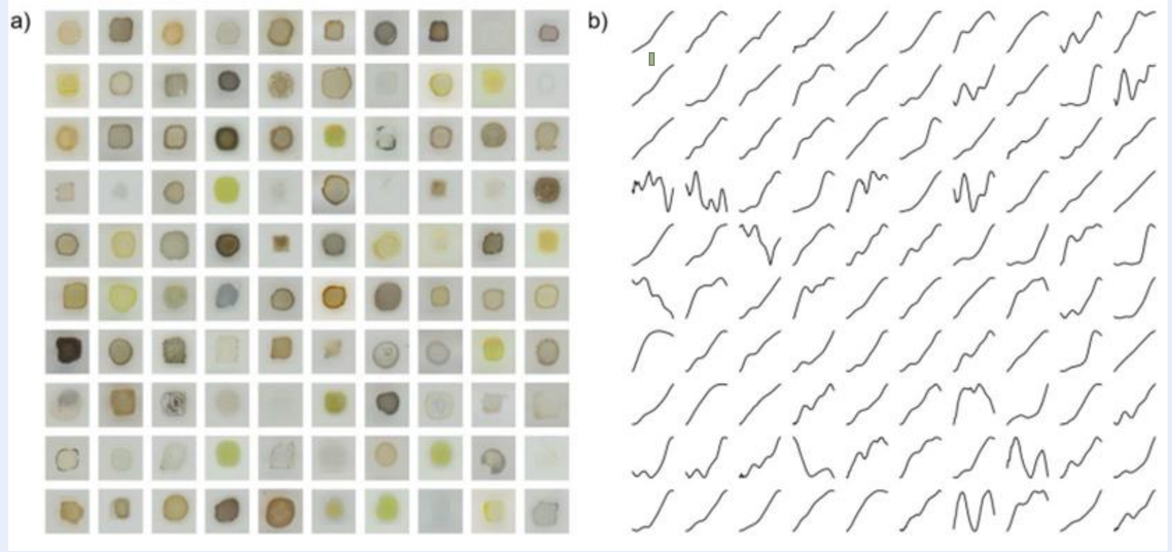


Fig. 1 shows the images of metal oxides on the left and their corresponding absorption spectrums

- 64 x 64-pixel image of metal oxide with corresponding absorption spectra
- 220 points representing energy between 1.31 and 3.1 electron volts
- 42 metals in different compositions

Fig. 2 Composition Heat Map

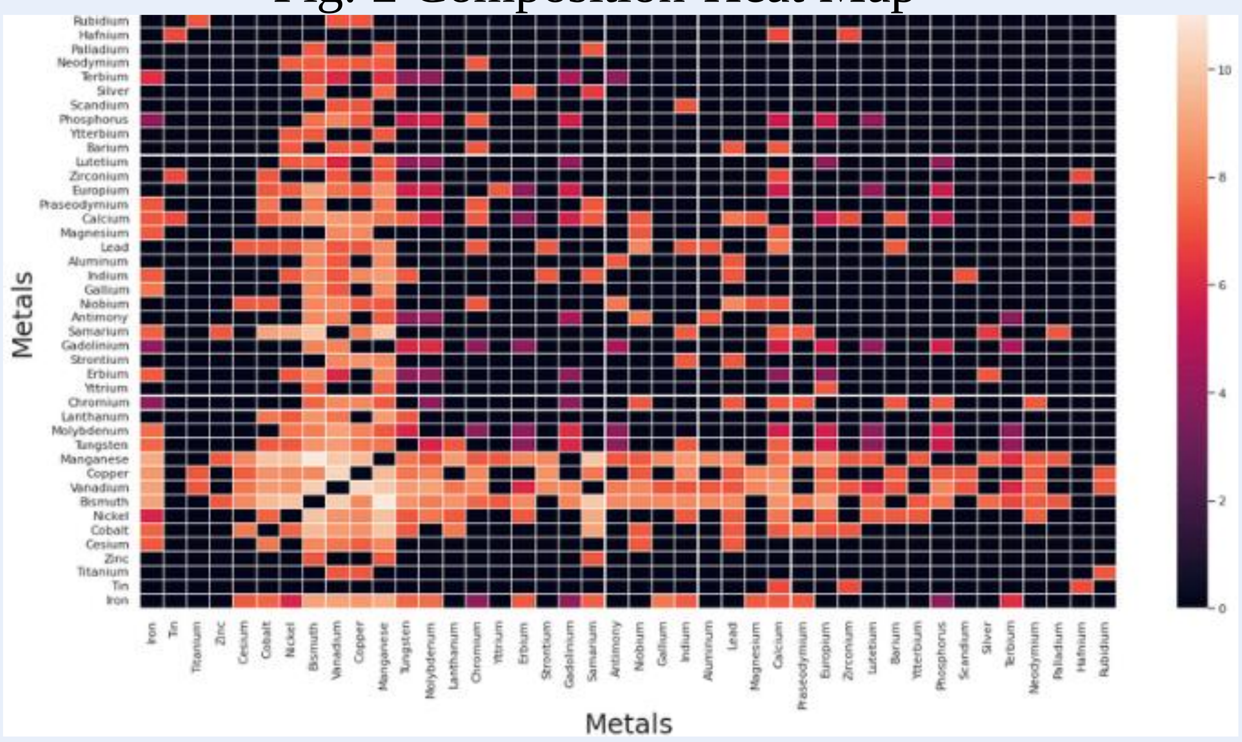
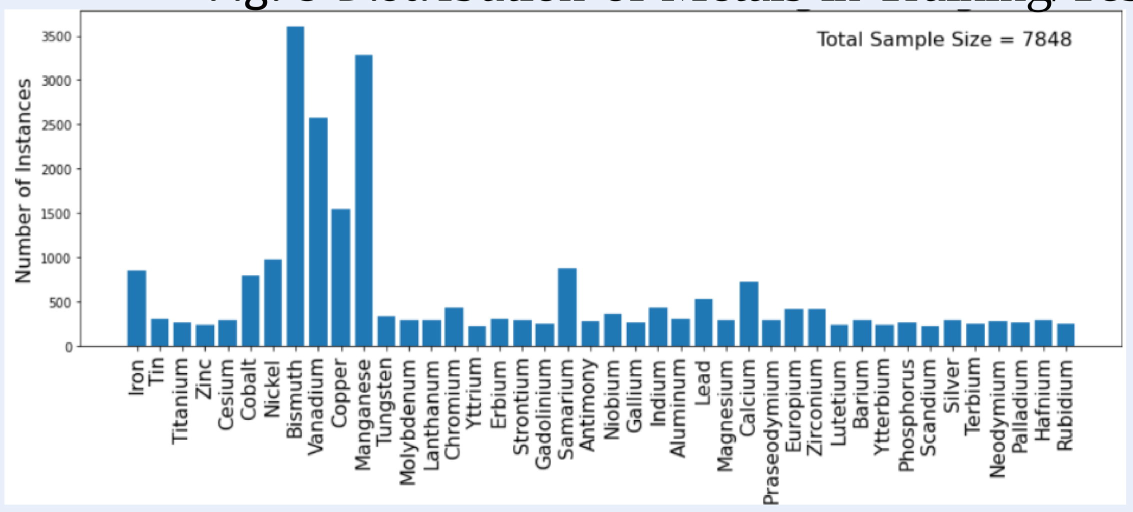


Fig. 2 is a heat map symmetrical along the diagonal and provides a visualization of how often specific metals are combined together

Fig. 3 Distribution of Metals in Training/Testing Subset



-Subset size: 7848 images out of 180,000 total (RAM limitations)
-Stratified sampling : at least 300 samples of each element

Fig. 3 is a representation of the number of instances each element appears over the stratified data set of 7848 images

Methodology

Fig. 4 Neural Network Architecture

```
tf.keras.layers.Conv2D(64, (3, 3), activation='relu', input_shape=(64, 64, 3)),
tf.keras.layers.Dense(256, activation = 'relu'),
tf.keras.layers.MaxPooling2D((2,2)),
tf.keras.layers.Dropout(.1),
tf.keras.layers.Flatten(),
tf.keras.layers.Dense(256, activation = 'relu'),
tf.keras.layers.Dense(128, activation = 'relu'),
tf.keras.layers.Dense(units=numOutputs, activation = 'sigmoid')])
```

Fig. 4 shows the TensorFlow layers in the network with their activation functions

-10 networks in parallel
-Stratified subset of 7848 images
-80-20 split on training/testing
- Epochs: Early stopping
-2,000 images validation
-Batch size: 25

Fig. 5 Single Prediction of Neural Network and Ensemble Network

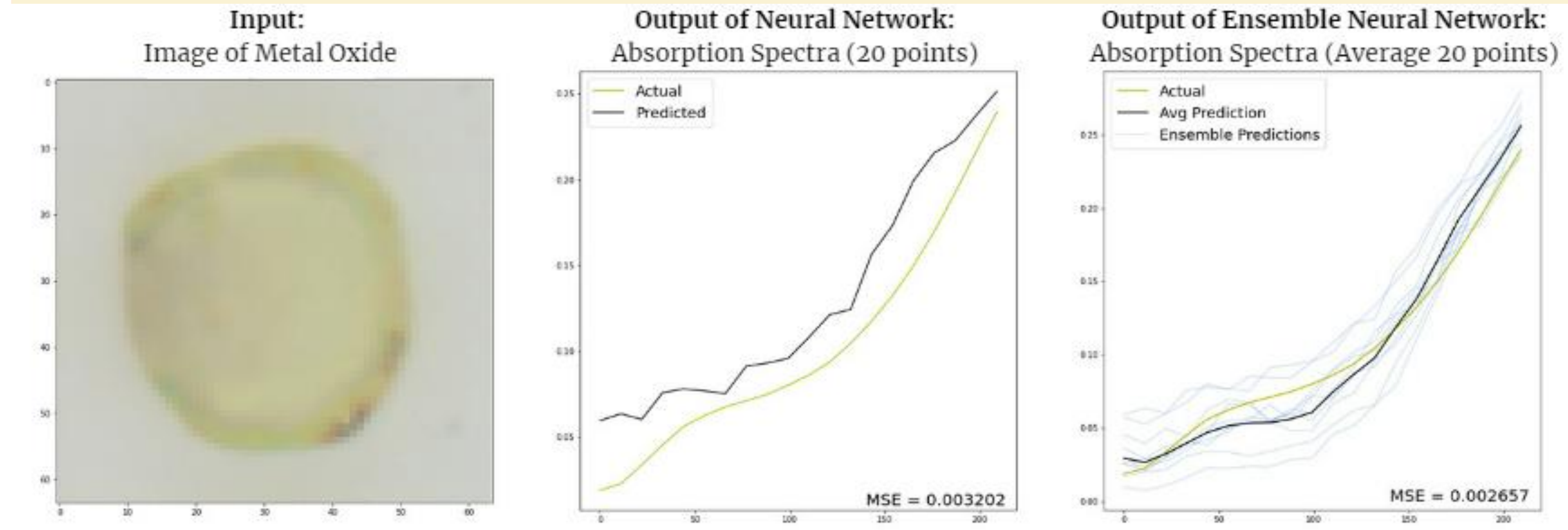


Fig. 5 compares the outputs of a metal oxide from the neural network and from the ensemble neural network

- Geometric average of the 10 predictions in the ensemble network
- Yellow – ground truth absorption spectra
- Black – prediction absorption spectra
- Ensemble network produces more accurate predictions with a smaller margin of error in comparison with a single neural network
- Smoother predictions
- Lower mean squared error

Fig. 6 Ensemble Predictions and Associated Confidence Interval

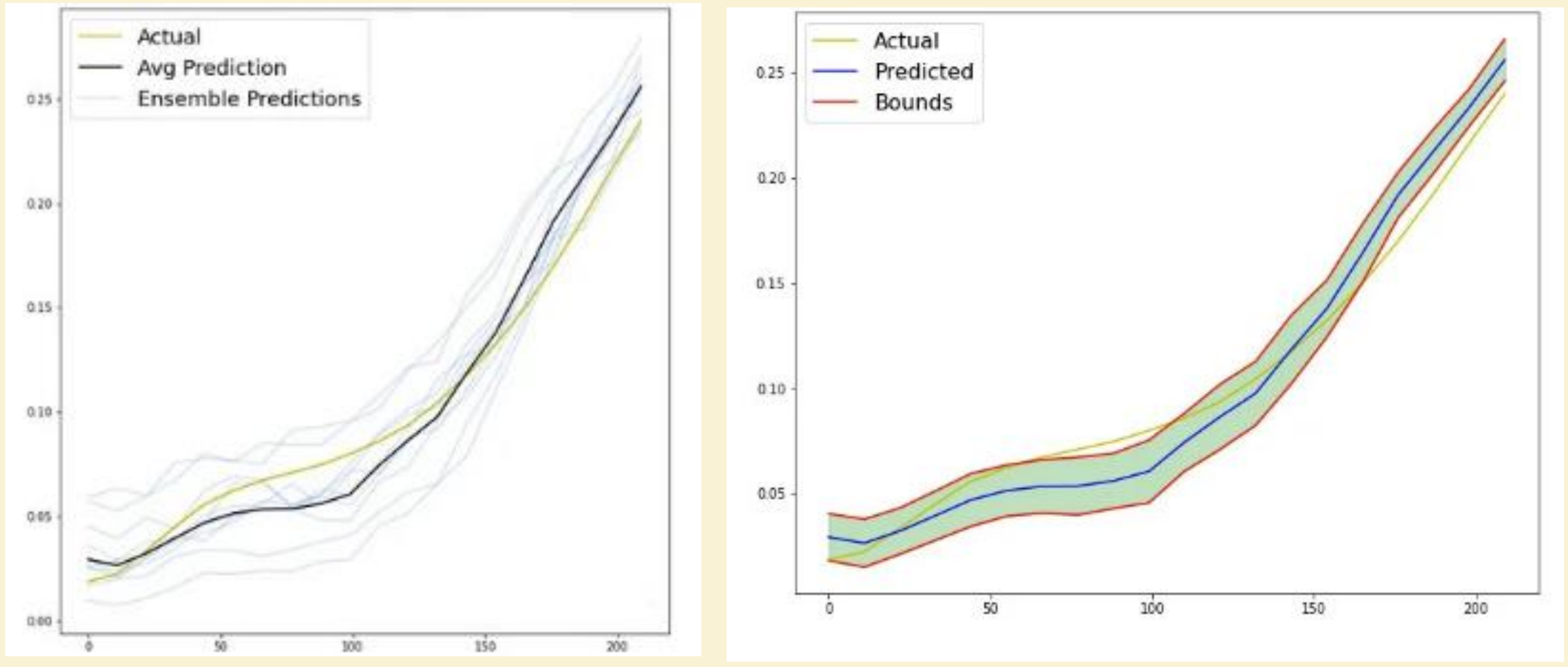


Fig. 6 demonstrates a 95% confidence interval around a singular prediction of the ensemble model. It also shows the range of confidence of model

- Thicker areas correspond to higher levels of uncertainty
  - Where individual networks mostly disagree
- Thinner areas correspond to lower levels of uncertainty
  - Where individual networks mostly agree
- Shows flaws of neural network algorithms
  - Upper right-hand corner: network had low uncertainty in its prediction (narrow bounds), but the prediction was inaccurate (ground truth outside of bounds)

Results

Fig. 7 Ensemble Neural Network vs Neural Network Scatter Plot Error

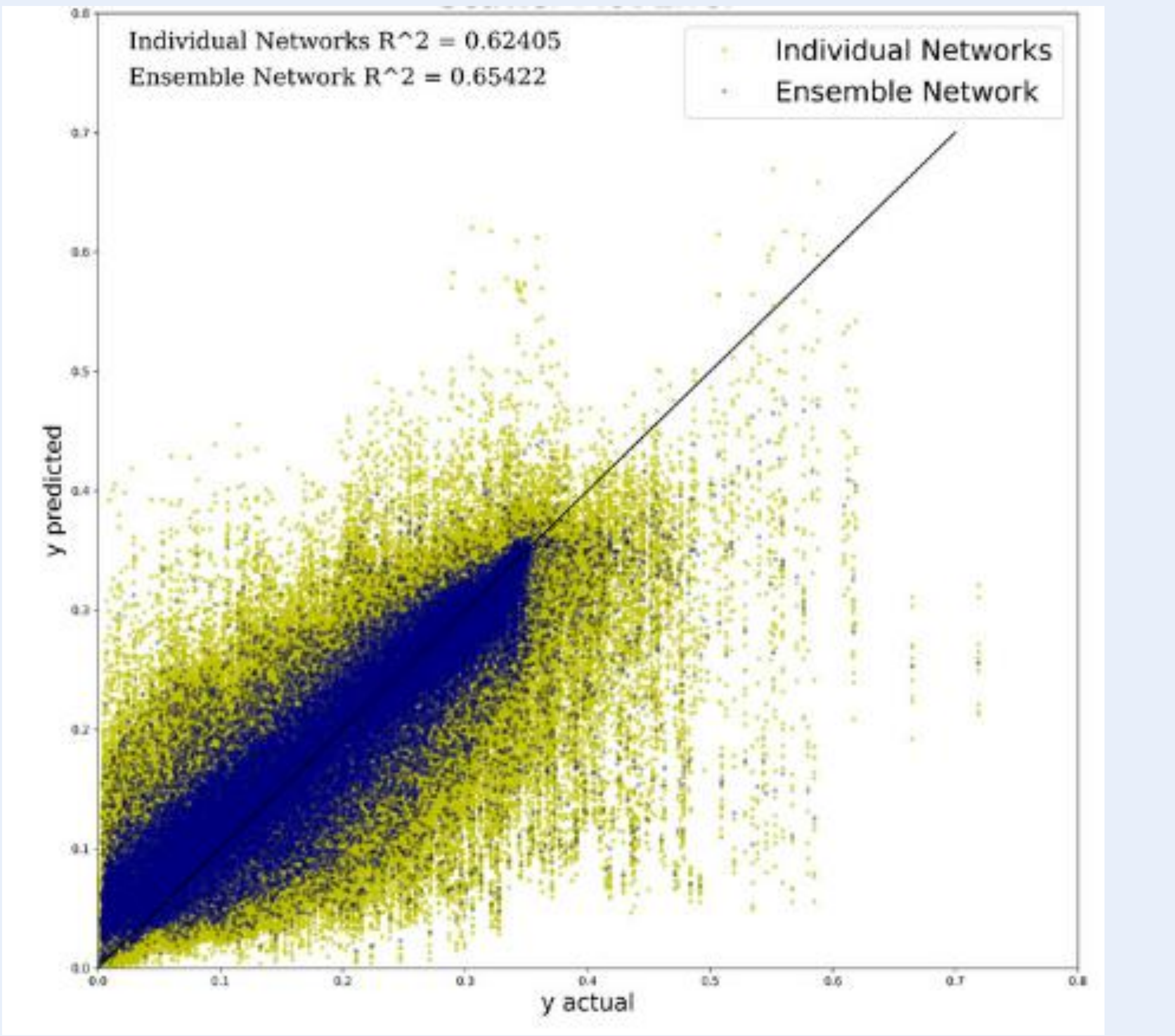


Fig. 7 shows the standard error of 2,000 predictions made by the 10 networks in the ensemble in yellow and the average prediction error of the ensemble in the blue

- Vertical distance from the line y=x to the point represents error
- Yellow represents the 10 different network predictions
- Blue represents the ensemble averaged output
- Y=x represents the desired value of the prediction
- Ensemble predictions have a higher R^2, therefore outperform singular networks

Fig. 8 Table of Error

	Mean Squared Error	R Squared Value
Single Neural Network	0.003202	0.56252
Ensemble Neural Network	0.002657	0.65422

Fig. 8 displays the mean squared error and R squared values for both the singular network and the ensemble model

- Mean squared error for ensemble network is smaller than the neural network
  - -Ensemble overall performed better than the single network
- R^2 value is closer to 1.0 for the ensemble network
  - -The model has less observed error
- Ensemble neural networks have a smaller mean squared error and larger R^2 value over 2,000 predictions

Key Conclusions

- Ensemble neural networks outperform singular networks
- The neural network was more certain in its predictions on the right side of the line spectra than the left side.

Acknowledgements

- Support for the program has been provided by the National Science Foundation (NSF) through REU Award Number DMS – 2050754
- This project was proposed by the Nevada National Security Site