# Thermoelectric Material Property Prediction Using Artificial Neural Networks

Luke Gibson[1], Luocheng Huang[2], Nathan Laurie[2], Ellen Murphy[2]
1. Department of Chemical Engineering 2. Department of Materials Science & Engineering
University of Washington, Seattle, WA

## Introduction & Background Information

The TEMANN (Thermoelectric Material Artificial Neural Network) package is a tool for researchers to probe the feasibility of novel materials for use in thermoelectric generators. Specifically, a user can input a chemical formula and TEMANN will provide a predicted Seebeck coefficient in uV/K. This package is built around an artificial neural network, which is a nonlinear supervised statistical data model.

Effective thermoelectric materials have low electrical resistivity and thermal conductivity. Thus, heat can be converted into electrical energy as carriers are excited near the heat source and diffuse to the lower energy space near the cold source (Figure 1).

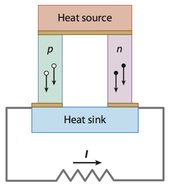



Figure 1: Cartoon of a thermoelectric generator.

## Dataset

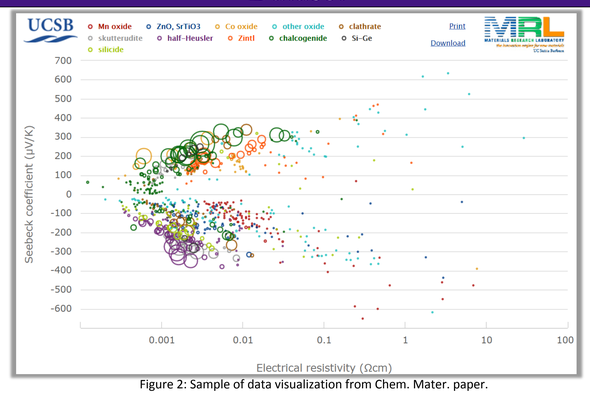



Figure 2: Sample of data visualization from Chem. Mater. paper.

The development of high-performance thermoelectric materials has eluded researchers for decades, as the underlying phenomena are not well understood. To advance this technology, the field increasingly relies on large datasets to tease out useful correlations to strong candidates. For this package we draw from one such of these datasets which was compiled at UC Santa Barbara to help researchers visualize various property spaces of potential thermoelectric materials. This primary dataset is augmented by data pulled from the Materials Project as well as the Atomic Simulation Environment.

## Artificial Neural Network

ANNs are a type of supervised learning model that takes a vector of input features and passes them through layers of nodes that transform the input signal and outputs the predicted quantity. ANNs are trained by adjusting the weights associated with every channel for every node.

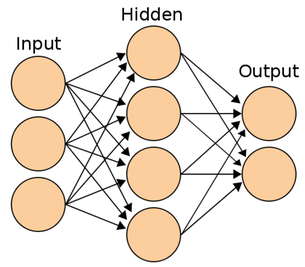



Figure 3: Graphical representation of an ANN with a single hidden layer

Prior to entering the input layer, the features in the feature vector are centered and scaled to have a mean of 0 and a standard deviation of 1 to ensure that all channels are treated equally.

Our feature vector is built with 3 inputs from the user:

1) Chemical formula
2) Space group of the material
3) Temperature of interest

Unique elements → elemental properties (80 channels)
Space group number → symmetry descriptors (15 channels)

## Training the ANN

Training of the ANN was done using sklearn's GridSearch class.

| Parameter Grid | |
|---|---|
| Number of Hidden Layers | 1, 3, 5 |
| Batch Size | 100, 200, 300 |
| Number of Epochs | 100, 200, 300 |

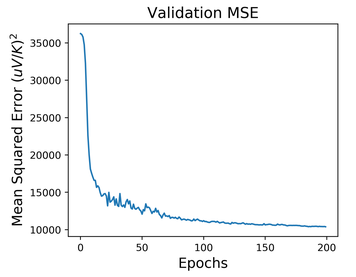
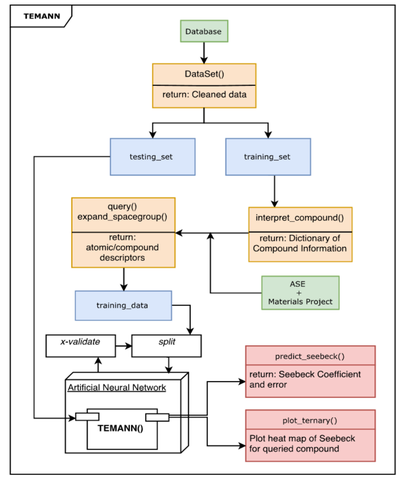



Figure 4: Validation loss (MSE) during training with the best parameters

**Best Parameters:**
Hidden layers: 1 Batch size: 100 Epochs: 200

## Design




Our code takes advantage of a number of packages available online such as pymatgen, Atomic Simulation Environment, SciKit-Learn, and Keras. In addition, we built custom functions to format data, link pre-built packages, and wrap all components into a single call.

## Visualization

As this package is made to be an exploratory tool we included a visualization, plot_ternarny(), for users to quickly examine a chemical space of their choosing.

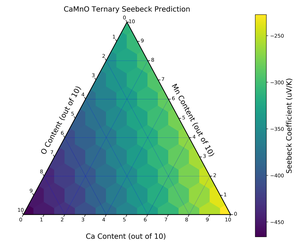



Figure 3: Ternary Prediction Graph for CaMnO

## Limitations

The functionality of this package would greatly benefit from a larger data set, more compound descriptors providing for more user flexibility, and a more rigorous data quality inspection. Furthermore, the model is less sensitive to the effects of stoichiometry and doping.

## References

**[1]** "Clemson Physicists' Research on thermoelectrics featured in Science.", http://newsstand.clemson.edu/mediarelations/clemson-physicists-research-on-thermoelectrics-featured-in-the-journal-science/
**[2]** Seshadri, R., Bonificio, et. al (2013). Data-Driven Review of Thermoelectric Materials: Performance and Resource Considerations. *ChemInform,44*(41)
**[3]**"artificalneutralnetwork", https://commons.wikimedia.org/wiki/File:Artifical_neural_network.svg
**[4]**Mark Harper et al (2015) python-ternary: Ternary Plots in Python. Zenodo. 10.5281/zenodo.34938
**[5]**pymatgen:*Python Materials Genomics (pymatgen) : A Robust, Open-Source Python Library for Materials Analysis.* Computational Materials Science, 2013, 68, 314–319.
**[6]**Keras: Challot, François, et al (2015) Github Repo. https://github.com/keras-team/keras

DIRECT
Data Intensive Research Enabling Clean Technologies

## Acknowledgements

Special thanks to Professor Ram Seshadri for graciously sharing his dataset and expertise as well as Dave Beck, Nick Montoni, Arushi Prakash, and Coco Mao for their guidance.

The Materials Project
UCSB
CHEMICAL ENGINEERING
MATERIALS SCIENCE & ENGINEERING