



Machine Learning Prediction of Heat Capacity for Solid Inorganics

Steven K. Kauwe¹ · Jake Graser¹ · Antonio Vazquez¹ · Taylor D. Sparks¹

Received: 5 February 2018 / Accepted: 6 May 2018 / Published online: 29 May 2018
© The Minerals, Metals & Materials Society 2018

Abstract

Many thermodynamic calculations and engineering applications require the temperature-dependent heat capacity (C_p) of a material to be known a priori. First-principle calculations of heat capacities can stand in place of experimental information, but these calculations are costly and expensive. Here, we report on our creation of a high-throughput supervised machine learning-based tool to predict temperature-dependent heat capacity. We demonstrate that material heat capacity can be correlated to a number of elemental and atomic properties. The machine learning method predicts heat capacity for thousands of compounds in seconds, suggesting facile implementation into integrated computational materials engineering (ICME) processes. In this context, we consider its use to replace Neumann-Kopp predictions as a high-throughput screening tool to help identify new materials as candidates for engineering processes. Also promising is the enhanced speed and performance compared to cation/anion contribution methods at elevated temperatures as well as the ability to improve future predictions as more data are made available. This machine learning method only requires formula inputs when calculating heat capacity and can be completely automated. This is an improvement to common best-practice methods such as cation/anion contributions or mixed-oxide approaches which are limited in application to specific materials and require case-by-case considerations.

Keywords Neumann-Kopp · Heat capacity · Inorganic solids · Machine learning · Material design · High-throughput

Introduction

The accurate and careful measurement of heat flow is the basis of thermodynamics. In particular, calorimetry, the process of measuring heat into and out of systems, has led to remarkable scientific progress. For example, the thermodynamic concept of Gibbs free energy is expressed with the thermodynamic equation $G = H - TS$, where H is the enthalpy, T is the temperature, and S is the entropy of the system. The absolute values of these state variables are impossible to obtain for practical applications. Instead, scientific and engineering fields obtain the critical thermodynamic information by determining the relative change in Gibbs energy. This is accomplished by measuring changes

in entropy and enthalpy which are fundamentally tied to heat transfer and heat capacity. Under most conditions, the change in a system is expressed by the equation $\Delta G = \Delta H - T \Delta S$. Determining the heat capacity is critical to find changes in entropy and enthalpy, as calorimetry is based on the relationship $\Delta H = \int_{T_i}^{T_f} C_p(T) dT$ for the enthalpy and $\Delta S = \int_{T_i}^{T_f} \frac{C_p(T)}{T} dT$ for entropy, where C_p is the temperature-dependent heat capacity at constant pressure. For convenience, heat capacity values are typically fit by constants: a, b, c, d, e to a power series such as $C_p = a + bT + cT^{-2} + dT^{-0.5} + eT^2$ [20]. In addition to thermodynamic uses, measuring heat flow into and out of a system as a function of temperature is an attractive way to probe phase transitions. First-order transitions are easily characterized by latent heating. Higher-order phenomena can also be monitored [14]; examples include the characterization of glassy transition temperatures in polymers and applications for determining magnetic and ferroelectric ordering.

In addition to studying fundamental science, knowing the value of heat capacity as a function of temperature is valuable from a design perspective. Considerations of heat capacity as a physical property are especially useful when

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s40192-018-0108-9>) contains supplementary material, which is available to authorized users.

✉ Taylor D. Sparks
sparks@eng.utah.edu

¹ Department of Materials Science and Engineering, University of Utah, Salt Lake City, USA

considering design parameters for thermal engineering applications. The foundational importance of heat capacity is further emphasized by the inclusion of heat capacity tables in almost all introductory engineering textbooks. Moreover, heat capacity also directly influences other properties such as thermal conductivity. Thermal conductivity measurements are rare due to the difficulty associated with measuring actual heat flux and properly handling loss from phenomena such as radiation. Thermal conductivity (κ) is instead calculated using thermal diffusivity (α), heat capacity (C_p), and density (ρ) with the equation $\kappa = \alpha\rho C_p$.

Despite the utility of heat capacity for engineers and scientists, careful measurements as a function of temperature are not published for all materials of interest. Even in well established repositories of thermochemical data, such as the NIST: JANAF tables, there are only several hundred crystalline compounds reported. Considering the scope of other crystal structure databases, such as the inorganic crystal structures database (ICSD) hosting over 188,000 inorganic entries, there is a lack of corresponding heat capacity data. Heat capacity measurements require equipment and resources that many labs do not have, which contributes to this information imbalance. A common method for determining heat capacity is differential scanning calorimetry. Proper use of this technique requires multiple runs (baseline, standard, and sample) which can take hours or days. The measurements are also sensitive to error from oxidation, poor thermal contact, sintering, volatilization, and others [7].

To overcome the challenges associated with measuring heat capacity, some researchers have turned to calculating heat capacity for compounds along side their measurements [17]. A common approach for predicting heat capacity uses Neumann-Kopp (N-K) estimations based on summing the heat capacity of the constituent elements. While this approach works surprisingly well at room temperature, it does poorly at high temperatures. Therefore, in an attempt to extend the utility of the Neumann-Kopp technique, Leitner et al. [9] used the concept of working with mixed-oxide constituents rather than individual elements. This method works better than the standard Neumann-Kopp approach at elevated temperatures; however, it relies on a case-by-case usage as there is no guarantee that such oxides and their thermodynamic data exist. Another approach uses contributions from cation and anion groups [10] to calculate heat capacity—referred to as cation/anion contribution (CAC). Implementing calculations using the CAC method is both easier and faster than the mixed-oxide method as they use identifiable cation/anion pairs and a temperature dependent power series. Finally, a more modern approach to predicting heat capacity relies on the computational determination of phonon frequencies [11] or other methods based around density functional theory (DFT). The modern

approach of simulation is limited in its use due to the speed of first-principle calculations.

In this work, we address a new approach for determining heat capacity based in the emerging field of materials informatics. This technique leverages machine learning (ML) algorithms and big data approaches to make statistically validated predictions without using physics-based calculations. For this work, we show that a model can be built by training with publicly available data in the NIST: JANAF tables. The results of this model allow us to compare the performance of machine learning to the performance of Neumann-Kopp calculations and group contribution methods such as CAC. In this context, we discuss the potential value of using this approach instead of, or in concert with, the traditional experimental approaches.

Methods

The machine learning method is fundamentally different from a first-principles approach. The models needed for machine learning prediction of heat capacity are built around statistical predictions rather than physics-based calculations.

Data Acquisition and Curating

Optimal performance requires high-quality data. To accurately predict heat capacity from chemical composition, temperature-dependent thermochemical information was gathered from the NIST: JANAF tables. The JANAF tables contain the chemical formula, phase, and thermochemical properties including: heat capacity, entropy, enthalpy, heats of formations, and equilibrium constants (all referenced at a standard state pressure of 0.1 MPa). Data was collected from the tables and sorted according to phase using in house code. All entries that included a crystal phase were extracted. Duplicate information was then discarded leaving information for crystal compounds and liquid compounds. The data was then sorted into solid and liquid entries—discarding the liquid phase and solid phases beyond the Curie point. The heat capacity and temperature were then isolated by removing extra thermochemical data such as enthalpy. This resulted in a dataset containing formula, temperature, and heat capacity information. A total of 263 formulae with discrete temperatures ranging from 298.15 to 3900 K were included in the training set totaling 3986 entries. Table 1 shows the format of the training data. Different allotropes were not explicitly considered. Instead the first entry to appear on the list of chemical formulae provided by NIST was selected.

Elemental heat capacity data are also an important input for Neumann-Kopp predictions. The THERMART: FREED

Table 1 Example of data format used in training set

Formula	Temperature (K)	Heat Capacity (J/mole K)
Al	250	12.997
...
Al	900	32.308
Al	933.45	32.959
Al ₂ O ₃	200	51.12
...
Al ₂ O ₃	900	122.662

database [20] provided such data in the form of heat capacity constants a , b , c , d , and e —used in the power series: $C_p(T) = a + bT + cT^{-2} + dT^{-0.5} + eT^2$. These calculated values compared well against JANAF values for elemental inputs.

Machine Learning: Feature Development and Algorithm Choice

In order to properly do machine learning, we need features which can be correlated to heat capacity. The chemical formula can provide features if the properties of its elemental and atomic constituents are considered. Employing formula-based feature development provides a universally applicable framework for heat capacity prediction that relies solely on the chemical formula. Conversely, first-principle calculations require structural information which in the status quo is difficult to obtain and not inherent to the chemical formula. A full list of elemental properties considered for feature development can be found in the supplemental information, Table S1. The formula, in conjunction with the elemental properties, provides features such as composition weighted average, stoichiometric sum, and variance. (See Fig. 1). The combination and comparison of elemental properties account for some of the physics involved in chemical bonding. Relations such as the differences in electronegativity, the shapes of neighboring orbitals, and the allowed valence states are just a number of the possible interactions that our feature development attempts to capture.

The elemental properties used for this model were collected from various online sources. The properties used represent an assortment of calculated and measured values. For the case of missing elemental properties—such as the nearest neighbor bond distance of a gas—a NaN value was assigned when attempting to create the feature. An example of formula creation is shown in Table 2. To simplify the machine learning process, we chose to replace NaN values with the median value of the associated feature. This treatment allows for a consistent feature set that can be compared against multiple algorithm types. When using a

For formula $A_aB_bC_c$, where elements A , B , C share a property i . Calculations for the average, sum, and variance features are shown.

weighted average:

$$avg_i = \frac{a}{a+b+c}A_i + \frac{b}{a+b+c}B_i + \frac{c}{a+b+c}C_i$$

stoichiometric sum:

$$sum_i = aA_i + bB_i + cC_i$$

variance:

$$var_i = \frac{(A - \bar{X})^2 + (B - \bar{X})^2 + (C - \bar{X})^2}{N}$$

$$\bar{X} = \frac{A_i + B_i + C_i}{N}, \quad N = \text{number of elements}$$

Fig. 1 Procedure for developing features using chemical formula

non-linear SVR kernel, features were scaled to have zero-mean and unit variance. Resulting feature vectors were also normalized.

Once the heat capacity data and features were assembled, three different machine learning algorithms were employed. The work in this paper quantifies the performance of linear regressors (LR), support vector regressors (SVR), and random forest regressors (RF) when predicting heat capacity. All three algorithms were created using the python 3 library, scikit-learn [15]. Five-fold cross-validation was used to generate training and validation sets which were used to quantify model performance. To ensure model integrity, a group cross-validation was performed. In other words, cross-validation splits were performed on the formula group and not individual entries to ensure that no formula in the training set would also be represented in the validation set at a different temperature.

In addition to predictions, the random forest algorithm also outputs feature importance, which is a metric indicating how heavily the model relied on a given feature in generating a prediction. A large number of the features included in the training data are insignificant in determining the heat capacity prediction. This was shown by training a RF model with all 171 features and comparing the feature's importance score to the score of a feature with randomly assigned integers ranging from -100 to 100 . Features scoring at or below a value of twice the random feature score were dropped from the training set. This set of refined features were used to generate three separate model types listed in the supplemental information Tables S2 and S3: (i) linear regression (ii) support vector regression (iii) random

Table 2 Creating six new features from Al_2O_3

Element	Atomic radius (pm)	Nearest neighbor distance (pm)
Aluminum	53	286
Oxygen	48	NaN
Feature		
Weighted avg.	76	$2(286) + 3(\text{NaN}) = \text{NaN}$
Sum	3.88	NaN
Variance	0.1225	NaN

Nearest neighbor distance illustrates the output when a property is not shared among all elements

forest models. We should note that Neumann-Kopp is inherently built into the feature vector as one of the features is the summation of room temperature heat capacity. Also, many of the remaining features are correlated to each other. Feature correlation can introduce bias in the importance rankings [1]; however, keeping them allows us to look at the general impact of all formula-based features. The correlation matrices for the significant features can be found in the supplemental information, Figures S1 and S2. The top five features for the room temperature model are shown in Table 3.

Overall model performance was determined by averaging performance metrics from each cross-validation set. Model parameters were selected using a grid-search technique to minimize the average of the root-mean-squared-errors in the model—Figures S3 and S4 show optimal parameters in the supplemental information. Root-mean-squared-error (RMSE), average percent error, and r-squared (R^2) values, were used to compare machine learning performance to the basic Neumann-Kopp model at room temperature, and the aforementioned CAC method at elevated temperatures. To quantify high temperature performance, we compared metrics associated with predictions using CAC to those of machine learning. We were able to calculate heat capacity from the Cation/Anion group contributions with 161 of the 263 total chemical formula. This was done over each formula's entire temperature range. Note, it is likely that many of the chemical formula used in this work were included in the original regression by Mostafa et al. [10] when creating the CAC coefficients. Because of this, it

Table 3 Most important model features at 298 K, not including Neumann-Kopp input

Feature	Importance score
Sum: covalent radii	17.4%
Sum: group	12.0%
Sum: Period	10.7%
Sum: Atomic Radius	10.4%
Sum: Martynov-Batsanov (MB) electronegativity	7.9%

is probable that our reported metrics overestimate CAC performance.

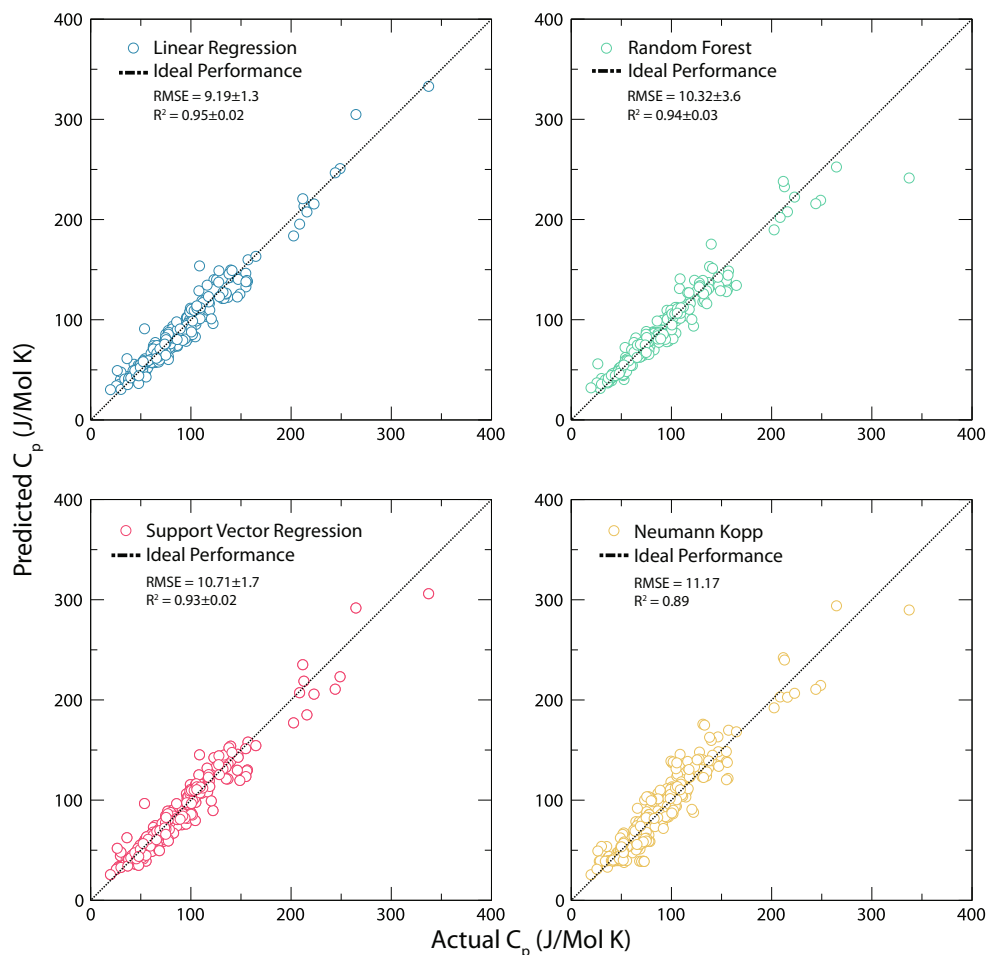
Results

Performance was determined for two different temperature regimes: (i) room temperature only, and (ii) from room temperature up to the maximum temperature for that compound's value in the training set. The linear regression, support vector regression, and random forest machine learning models all outperformed the Neumann-Kopp with lower RMSE values, lower percent error, and higher R^2 values. A graphical representation of performance in regime (i) is shown in Fig. 2, which contains data points representing the predicted heat capacity value vs the actual values for the basic Neumann-Kopp method and machine learning predictions at 298.15 K. (The black 45° line represents the ideal model performance.) For regime (i), the best machine learning predictions perform better than the traditional Neumann-Kopp prediction.

It is notable that one of the input features for the machine learning model is the room temperature Neumann-Kopp prediction. We removed this feature to test the importance of Neumann-Kopp prediction as an input to the machine learning model. Table 4 shows the error associated with the three ML predictions before and after the removal of Neumann-Kopp as a feature, as well as the standalone Neumann-Kopp predictions. Displayed values include the 95% confidence region wherein the true mean is expected to lie (e.g., $14.7 \pm (2 * \text{standarderror})$). This was calculated using the cross-validation metrics. Interestingly, the models perform nearly as well without the Neumann-Kopp features, by more heavily relying on the element-based features. For clarity, all figures were generated using Neumann-Kopp as a feature (Fig. 3).

At elevated temperatures (regime (ii)), Neumann-Kopp predictions become less reliable. Issues with Neumann-Kopp predictions beyond room temperature are primarily due to changes in heat capacity associated with the following phenomena: Curie temperature, structure transitions,

Fig. 2 Neumann-Kopp and machine learning predictions vs actual values at 298.15 K



melting, sublimation, or reaching the Debye limit for individual atoms. All of these phenomena drastically alter heat capacity [8]. For example, consider aluminum metal that melts at 933 K. The pure aluminum liquid has a constant heat capacity above this temperature. However, the aluminum-based compound, aluminum nitride, does not melt until 2473 K. Relying on the heat capacity of molten aluminum via Neumann-Kopp for these elevated temperatures will produce greater and greater errors with increased temperature.

Machine learning predictions in this work also lose accuracy when predicting over a large temperature range (regime *(ii)*); however, the error associated with machine learning is lower than CAC as seen in Fig. 4, and Table 5. The mixed-oxide contributions approach was also employed. However, after hours of work, we only identified 11 oxides which were composed of oxide constituents from the NIST JANAF tables. Clearly, comparing predictions for 11 compounds against 263 isn't valid so for the mixed oxide approach the error metrics are not included in the figures

Table 4 Performance at 298.15 K

Performance metrics	LR	SVR	RF	N-K
RMSE	9.19 ± 1.3	10.71 ± 1.7	10.32 ± 3.6	11.17
Percent error	8.92 ± 1.2	9.68 ± 0.9	7.96 ± 1.0	13.69
R^2	0.95 ± 0.02	0.93 ± 0.02	0.94 ± 0.03	0.89
Performance metrics w/o N-K feature				
RMSE	9.34 ± 1.3	10.53 ± 1.9	10.63 ± 3.6	11.17
Percent error	9.02 ± 1.3	9.77 ± 0.9	8.21 ± 1.0	13.69
R^2	0.95 ± 0.02	0.94 ± 0.02	0.94 ± 0.03	0.89

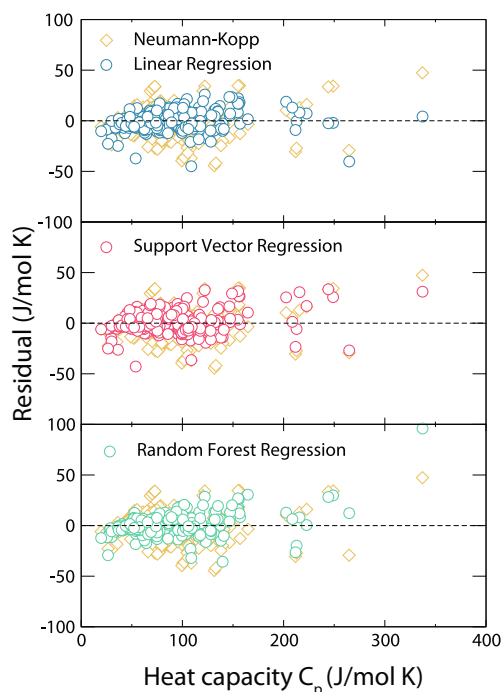


Fig. 3 Residual values (actual - predicted) of machine learning and Neumann-Kopp predictions at 298.15 K

or tables. The mixed-oxide approach of 11 compounds across 19 separate temperatures generated 194 unique heat capacity entries. Using these, we calculated an average RMSE of 10.4, average percent error of 5.70%, and an R^2 of 0.88. Table 6 shows the number of formulae considered with each method. For simplicity, the residual plot when including elevated temperatures is in the supplemental information, Figure S5.

Discussion

Machine learning, in itself, requires no knowledge of the underlying physical principles when predicting heat capacity. The process of converting the element-derived features into a prediction are handled by the algorithm using purely statistical methods. In the case of heat capacity, the origin and mechanism has been well known for many years. We expect the most important features to align with the physical mechanisms of heat capacity. In fact, the three most important features when predicting heat capacity at elevated temperatures, not including Neumann-Kopp calculations, are:

1. Sum: valence-shell s electrons
2. Sum: Unfilled f valence electrons
3. Sum: Outer shell electrons

These features are all properties typically associated with the bonding of the molecule. The traditional understanding of heat capacity, which utilizes the vibrational frequencies in the solid, relates well to this observation since bond strength and type will influence vibrational frequencies. Thus, the observed features reinforce the currently understood physical mechanism of heat capacity.

At room temperature, machine learning performs on par with Neumann-Kopp and has comparable performance to the CAC method up to 1000 K, as seen in Fig. 5. Beyond 1000, CAC error begins to rise due to the limitations with using a power series. All models have similar temperature behavior, with random forest exhibiting the lowest error. (A comparison of support vector regression and linear regression to CAC error with respect to temperature can be seen in Figures S6 and S7 in the supplementary information.) The Neumann-Kopp method of predicting heat capacity has traditionally been abandoned at elevated temperature values, forcing some to rely on the cumbersome Cation/Anion contribution method, or DFT calculations. These processes are time intensive, and in some instances, oxide and Cation/Anion values themselves are not well-known. The results in this work suggest that machine learning predictions can not only replace Neumann-Kopp calculations as a high-throughput method at room temperature, but can also be used at elevated temperatures with acceptable prediction accuracy. Machine learning offers a percent error at 7.96% for room temperature and 7.27% at elevated temperatures. These surpass Neumann-Kopp calculations of 13.69% and CAC errors of 17.30%.

As we ask the models to predict heat capacity at temperatures further and further above ambient room temperature, we can see the onset of error in the different approaches. The models developed clearly perform as well as Neumann-Kopp at room temperature and exhibit better accuracy with less effort compared to cation/anion contribution methods at high temperatures, but improvements can still be made. There are a number of additional techniques that could further improve this machine learning algorithm. Perhaps the most important change is to expand the data in the training set. This model is surprisingly accurate despite only being trained off of 263 formula entries. Hundreds of additional entries could be added if researchers consider thermochemical texts such as Barin's *Thermochemical Data of Pure Substances* [2]. Beyond data, features can also be expanded to improved performance. For example, new features could be added using a systematic approach to obtain cation/anion contribution and mixed-oxide calculations. The elemental features could be refined beyond the simple approach used in this work to capture more physical interactions.

Fig. 4 Cation/anion contribution and machine learning predictions vs actual values. Note consistent behavior in cation/anion contribution predictions resulting from broad temperature ranges

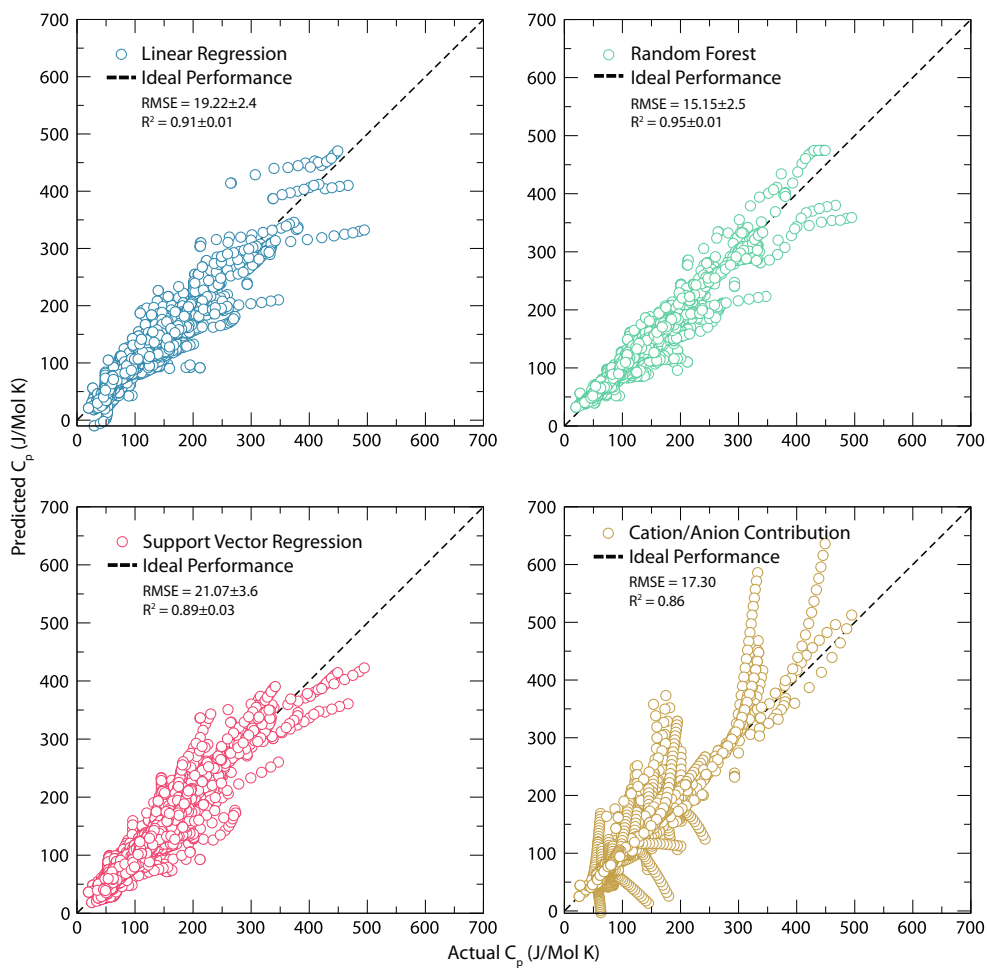


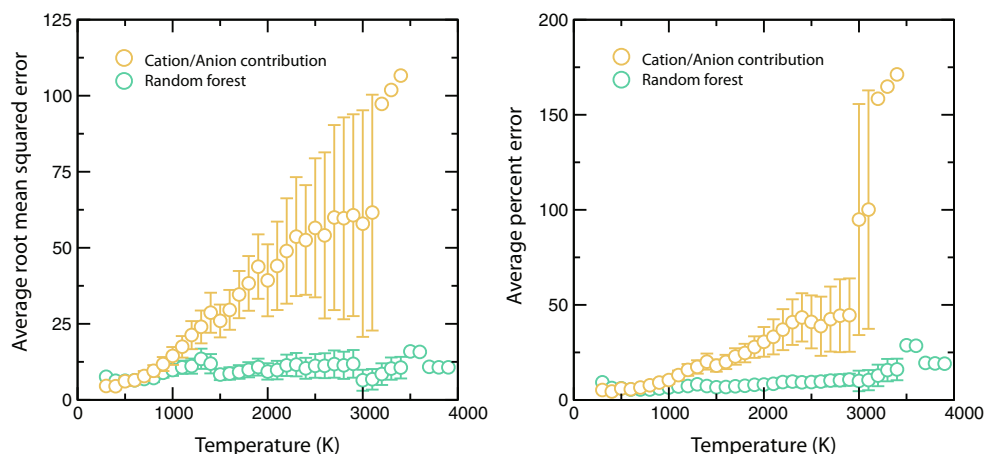
Table 5 Performance when including elevated temperatures

Performance metrics	LR	SVR	RF	CAC
RMSE	19.22 ± 2.4	21.07 ± 3.6	15.15 ± 2.5	17.30
Percent error	11.38 ± 0.7	13.40 ± 2.6	7.27 ± 0.1	13.29
R ²	0.91 ± 0.01	0.89 ± 0.03	0.95 ± 0.01	0.86
Performance metrics w/o N-K feature				
RMSE	19.42 ± 2.5	21.07 ± 3.6	14.40 ± 2.6	17.30
Percent error	11.52 ± 0.6	13.40 ± 2.6	7.31 ± 1.0	13.29
R ²	0.91 ± 0.01	0.89 ± 0.03	0.95 ± 0.02	0.86

Table 6 Number of valid entries per method

Method	Number of valid formulae	Cp data (all temperatures)
Machine learning	263	3986
Cation/anion contributions	161	2483
Mixed-oxide contributions	11	194

Fig. 5 Comparison of performance versus temperature. Heat capacities were grouped in 100 degree increments and performance scores were averaged. 95% confidence intervals are shown



Conclusion

Heat capacity is a fundamental thermodynamic property that is critical to many materials design processes. The ability to predict—and therefore design—new materials around heat capacity considerations has immense potential. Traditional methods for predicting heat capacity, such as DFT based calculations [6], Cation/anion contribution methods, and Neumann-Kopp estimations, are limited either by speed or accuracy.

The work presented here shows that machine learning predictions can effectively be used to rapidly predict heat capacity. In particular, the use of elemental information as inputs to machine learning models allow for accurate, fast, and broad predictive power. The models developed in this work are capable of predicting heat capacity for any inorganic formula with any temperature—in fractions of a second—at or above the accuracies exhibited by the Neumann-Kopp and Cation/Anion contribution methods. This is particularly promising for use in high-throughput material screening and could be valuable in influencing engineering decisions.

Two high temperature approaches of particular interest are the work of Leitner et al. [9] using mixed-oxide combinations, and techniques used by Mostafa et al. [10] which utilizes group contributions of Cation/Anion pairs—the high temperature comparison in this work. A comparison of the machine learning method to the mixed-oxide approach could possibly be expanded with considerable effort. Improvements to the machine learning method, beyond the models in this work, are likely to be found through improved feature development and expanded training data. New relationships between current elemental properties could be explored. Deployment of advanced feature selection techniques should also be used. These include cluster resolution feature selection and multidimensional principle component analysis which have

already been shown to be effective in improving material structure predictions [13].

The performance of the machine learning models in this work provides further confirmation that machine learning is playing a significant and growing role in predicting materials properties and developing new materials [3–5, 12, 16, 18, 19]. In the context of the Materials Genome Initiative which calls for greater collaboration between computational and experimental material scientists, this addition of artificial intelligence and machine learning stands to bring a new dynamic to materials discovery.

Acknowledgements The authors gratefully acknowledge support from the NSF CAREER Award DMR 1651668.

References

1. Altmann A, Tološi L, Sander O, Lengauer T (2010) Permutation importance: a corrected feature importance measure. *Bioinformatics* 26(10):1340–1347. <https://doi.org/10.1093/bioinformatics/btq134>
2. Barin I, Platzki G (1989) Thermochemical data of pure substances. vol 304, Wiley Online Library
3. Gaultois MW, Olynyk A, Mar A, Sparks TD, Mulholland GJ, Meredith B (2016) Perspective: Web-based machine learning models for real-time screening of thermoelectric materials properties. *APL Mater* 4(5):053213. <https://doi.org/10.1063/1.4952607>
4. Gaultois MW, Sparks TD, Borg CK, Seshadri R, Bonificio WD, Clarke DR (2013) Data-driven review of thermoelectric materials: performance and resource considerations. *Chem Mater* 25(15):2911–2920. <https://doi.org/10.1021/cm400893e>
5. Ghadbeigi L, Harada JK, Lettiere BR, Sparks TD (2015) Performance and resource considerations of li-ion battery electrode materials. *Energy Environ Sci* 8(6):1640–1650. <https://doi.org/10.1039/C5EE00685F>
6. Graser J, Kauwe SK, Sparks TD (2017) Machine learning and energy minimization approaches for crystal structure predictions: A review and new horizons, *Chemistry of Materials*, in press

7. III LMC, Taylor RE (1975) Radiation loss in the flash method for thermal diffusivity. *J Appl Phys* 46(2):714–719. <https://doi.org/10.1063/1.321635>
8. Kittel C (2005) *Introduction to solid state physics*. Wiley, Hoboken
9. Leitner J, Voka P, Sedmidubský D, Svoboda P (2010) Application of Neumann–Kopp rule for the estimation of heat capacity of mixed oxides. *Thermochimica Acta* 497(1):7–13. <https://doi.org/10.1016/j.tca.2009.08.002>
10. Mostafa ATMG, Eakman JM, Montoya MM, Yarbrow SL (1996) Prediction of heat capacities of solid inorganic salts from group contributions. *Ind Eng Chem Res* 35(1):343–348. <https://doi.org/10.1021/ie9501485>
11. Narasimhan S, de Gironcoli S (2002) Ab initio. *Phys Rev B* 65:064302. <https://doi.org/10.1103/PhysRevB.65.064302>
12. Oliynyk A, Antono E, Sparks TD, Ghadbeigi L, Gaultois MW, Meredig B, Mar A (2016) High-throughput machine-learning-driven synthesis of full-Heusler compounds. *Chem Mater* 28(20):7324–7331. <https://doi.org/10.1021/acs.chemmater.6b02724>
13. Oliynyk A, Mar A Discovery of intermetallic compounds from traditional to machine-learning approaches. *Accounts of chemical research*. <https://doi.org/10.1021/acs.accounts.7b00490>
14. Parker WJ, Jenkins RJ, Butler CP, Abbott GL (1961) Flash method of determining thermal diffusivity, heat capacity, and thermal conductivity. *J Appl Phys* 32(9):1679–1684. <https://doi.org/10.1063/1.1728417>
15. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: Machine learning in Python. *J Mach Learn Res* 12:2825–2830
16. Seshadri R, Sparks TD (2016) Perspective: Interactive material property databases through aggregation of literature data. *APL Mater* 4(5):053206. <https://doi.org/10.1063/1.4944682>
17. Sparks TD, Fuierer PA, Clarke DR (2010) Anisotropic thermal diffusivity and conductivity of la-doped strontium niobate Sr₂Nb₂O₇. *J Amer Ceram Soc* 93(4):1136–1141. <https://doi.org/10.1111/j.1551-2916.2009.03533.x>
18. Sparks TD, Gaultois MW, Oliynyk A, Brgoch J, Meredig B (2016) Data mining our way to the next generation of thermoelectrics. *Scr Mater* 111:10–15. <https://doi.org/10.1016/j.scriptamat.2015.04.026>
19. Tehrani AM, Ghadbeigi L, Brgoch J, Sparks TD (2017) Balancing mechanical properties and sustainability in the search for superhard materials. *Integr Mater Manuf Innov* 6(1):1–8. <https://doi.org/10.1007/s40192-017-0085-4>
20. Thermart: Freed-thermodynamic database (2017). <http://www.thermart.net/freed-thermodynamic-database/>