Article

Probing anharmonic phonons in WS$_2$ van der Waals crystal by Raman spectroscopy and machine learning

Chisom Okeke,
Isaac Juma,
Antonio
Cobarrubia,
Nicholas Schottle,
Hisham Maddah,
Mansour
Mortazavi, Sanjay
K. Behura

sbehura@sdsu.edu

Highlights

Thermal-induced
anharmonic phonon
dynamics in WS$_2$ crystal is
demonstrated

First-order vibrational
modes followed a red shift
with an increase in
temperature

ML algorithms: multiple
linear regression, decision
tree, and random forest
are used

ML models developed to
train, test, and predict
frequencies at different
temperatures

Article

Probing anharmonic phonons in WS_2 van der Waals crystal by Raman spectroscopy and machine learning

Chisom Okeke,[1] Isaac Juma,[1] Antonio Cobarrubia,[2,3] Nicholas Schottle,[2] Hisham Maddah,[4] Mansour Mortazavi,[1] and Sanjay K. Behura[2,5,*]

SUMMARY

Understanding the optothermal physics of quantum materials will enable the efficient design of next-generation photonic and superconducting circuits. Anharmonic phonon dynamics is central to strongly interacting optothermal physics. This is because the pressure of a gas of anharmonic phonons is temperature dependent. Phonon-phonon and electron-phonon quantum interactions contribute to the anharmonic phonon effect. Here we have studied the optothermal properties of physically exfoliated WS_2 van der Waals crystal via temperature-dependent Raman spectroscopy and machine learning strategies. This fundamental investigation will lead to unveiling the dependence of temperature on in-plane and out-of-plane Raman shifts (Raman thermometry) of WS_2 to study the thermal conductivity, hot carrier diffusion coefficient, and thermal expansion coefficient.

INTRODUCTION

Temperature-dependent Raman line shift, Raman linewidth, and Raman intensity have been historically used to probe the anharmonic phonon dynamics in bulk as well as reduced-dimensional semiconducting and nonmetallic solids.[1–6] Understanding anharmonic phonons is critical to understand the thermal properties including thermal conductivity, thermal expansion, and specific heat of solid-state materials.[1,7] In case of reduced-dimensional sp^2-hybridized carbon materials such as graphene and carbon nanotubes, the temperature-dependent shifting of the Raman G phonon mode has been used as a probe to characterize the thermal transport properties.[8] This Raman-based thermometry approach has also been employed to study thermal transport properties of other reduced dimensional materials including transition metal dichalcogenides (TMDs).[9] TMDs (MX_2, where M = Mo, W, Nb, In, Ta, or Ti and X = S, Se or Te) are a class of layered crystals with weak interlayer van der Waals bonding and strong intralayer covalent bonding.[10,11] In the bulk form, TMDs are indirect bandgap semiconductors and can make a transition to direct bandgap semiconductors in their monolayer form. While single layers of TMD materials receive significant attention owing to their high on/off ratio and high quantum yield, multilayers of TMDs are also of special interest. This is because of the recent discovery of strongly interacting physics in few-layer WS_2 including phonon-assisted indirect bandgap transition lasing in ultra-thin WS_2 disc[12] and exciton-phonon quantum interference in few-layer WS_2.[13] Furthermore, few-layer TMDs such as few-layer WS_2 exhibit excellent light-matter interactions owing to the van Hove singularity induced enhancement of visible-frequency absorption.[14,15]

The central approach taken in this work is combining experimental study using temperature evolution of Raman spectroscopy and computational analysis using machine learning strategies as shown in Figures 1 and 2. Temperature-dependence of phonon modes is convenient to interpret anharmonicity between phonons and electron-phonon interactions.[16–20] These measurements can provide the fundamental parameters of the material properties such as temperature coefficient and thermal conductivity, which are important for device applications. Here, a thin film of WS_2 is exfoliated from a bulk crystal and transferred onto a 300 nm silicon dioxide (SiO_2)-coated silicon (Si) substrate followed by temperature-dependent Raman spectroscopic measurements in a temperature range of 80–460K. Since the goal of this study is to understand the physics of anharmonic phonon dynamics in WS_2 crystals, therefore, the substrate-supported films are used. We understand that to measure the thermal conductivity of WS_2 films, it may be accurate to use suspended films as the underneath substrate could play an important role in thermal dissipation. We observed that both the in-plane (E'_{2g}) and out-of-plane (A_{1g}) Raman vibrational frequencies of WS_2 films

[1]Department of Mathematics and Computer Science and Department of Chemistry and Physics, University of Arkansas at Pine Bluff, 1200 N. University Drive, Pine Bluff, AR 71601, United States

[2]Department of Physics, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182, United States

[3]Computational Science Research Center, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182, United States

[4]Department of Chemical Engineering, King Abdulaziz University, Rabigh 21911, Saudi Arabia

[5]Lead contact

*Correspondence:
sbehura@sdsu.edu
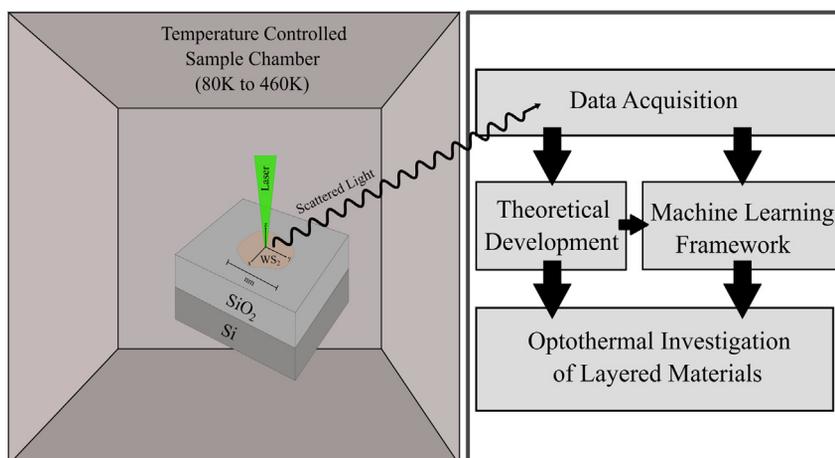
Figure 1. Schematic of the experimental setup
Process pathways taken in this work to probe the anharmonic phonons in bulk WS$_2$ crystal.

decrease with temperature rise. Data analysis and machine learning algorithms like multiple linear regression, decision trees, and random forest are employed to understand the effect of temperature on the phonon vibrational modes.

RESULTS AND DISCUSSION

Room temperature Raman vibrational modes of few-layer WS$_2$

When a 532 nm laser beam is directed at an exfoliated but nonetheless bulk WS$_2$ crystal staged on a SiO$_2$/Si substrate as illustrated in Supporting Information Figure S1, the Raman spectrum reveals a number of first-order, second-order, and combinational first-order second-order phonon modes which we look at in the following with the help of Figure 3.[21] We start by identifying all vibrational modes present at room temperature. At about 171 cm^{-1} the first mode, LA(M) is recorded. This is the one longitudinal acoustic mode at the M point of the Brillouin Zone (BZ) which refers to the in-plane collective movement of the atoms.[22,23] At about 185 cm^{-1}, a first-order mode, LA(K) is noted. A$_{1g}$(M)-LA(M) is found at 225 cm^{-1} and is a combinational mode of both first-order and second-order vibrational modes. The E'$_{2g}$ mode is recorded at 346.401 cm^{-1} and refers to the in-plane vibration of each atom, while the A$_{1g}$ mode is at 415.44 cm^{-1} and represents the out-of-plane vibration of atoms as shown in Figure 4 inset. There are two more peaks; one at 297 cm^{-1} tagged as 2LA(M)-3E$^2_2$(M), and another one at 319.66 cm^{-1} identified as 2LA(M)-2E$^2_{2g}$(M). After the A$_{1g}$ mode, two more peaks are identified, one at 575 cm^{-1} tagged as LA(M)+A$_{1g}$(M),[24] and the last one at 700 cm^{-1}, the 4LA(M).[25] Silicon (substrate) peak is also identified at about 522 cm^{-1}. Table 1 is a presentation of all the Raman vibrational modes identified and tagged in the bulk WS$_2$.

As observed in Figure 3, we have two first-order in-plane (E'$_{2g}$) and out-plane (A$_{1g}$) vibrational modes with large intensity. These two modes are of particular interest and will be the focus of this study to understand the effect of temperature on phonon dynamics. This is because both the anharmonicities of phonon modes and crystal structure thermal expansion are temperature dependent. Therefore, it is critical to investigate the temperature evolutions of the phonon modes.

The number of layers of the exfoliated WS$_2$ on the SiO$_2$/Si substrate is estimated to be bulk with the number of layers more than 5. This is because the difference of peak positions (A$_{1g}$ - E'$_{2g}$) from the room temperature Raman spectrum is 69 cm^{-1} and the intensity ratio of 2LA and A$_{1g}$ Raman peaks is 0.43749 and the intensity ratio of A$_{1g}$ and E'$_{2g}$ is 0.425015, which corresponds to bulk crystal.[26,27] Calculations on peak differences (A$_{1g}$ - E'$_{2g}$) and the intensity ratios of 2LA and A$_{1g}$ (I$_{2LA}$/I$_{A1g}$) and A$_{1g}$ and E'$_{2g}$ (I$_{A1g}$/I$_{E'2g}$) are provided in Table 2.

Temperature evolutions of phonon vibrations (E'$_{2g}$ and A$_{1g}$) in few-layer WS$_2$

We studied the temperature dependence of Raman vibrational modes of the WS$_2$ layer exfoliated on SiO$_2$/Si substrate. The temperature dependence of the E'$_{2g}$ and A$_{1g}$ peak positions in WS$_2$ was measured
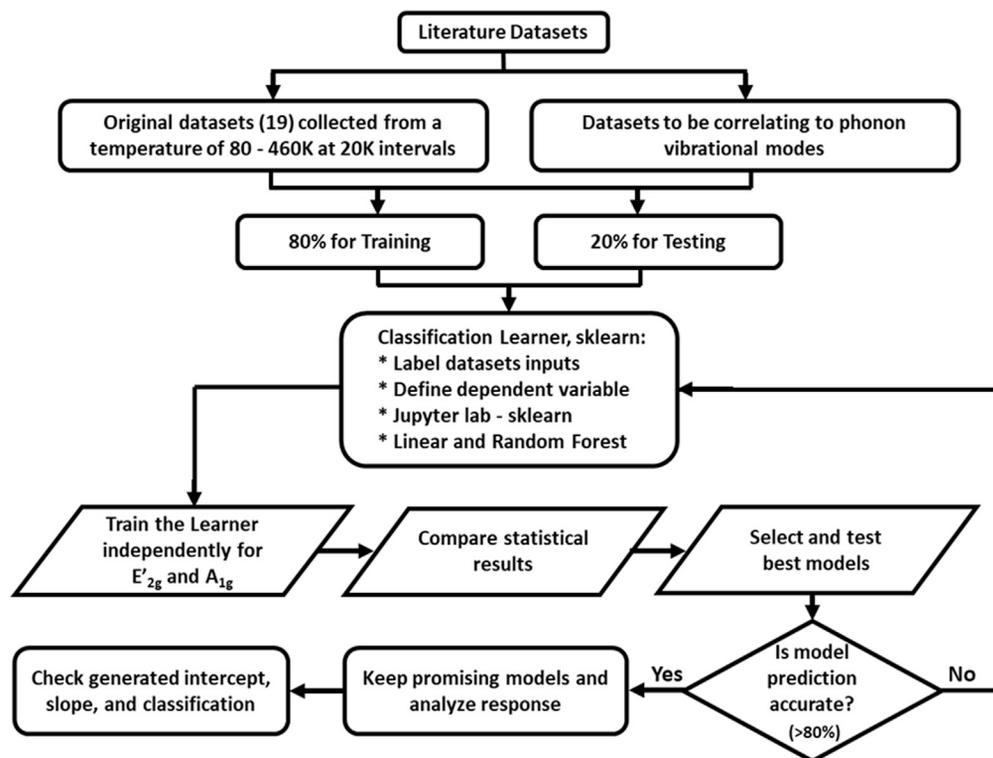
Figure 2. Framework of the machine learning analysis
The machine learning algorithms like multiple linear regression, decision trees, and random forest are employed to understand the effect of temperature on the phonon vibrational modes.

between 80 K and 460 K as shown in Figure 4. As illustrated in Figure 4 (inset), the A_{1g} vibrational mode is the vibration of the S atoms across the plane (shown by dotted lines), while the E'_{2g} is the in-plane vibration of both the W and S atoms.[26,28] These two vibrational modes are crucial in the structural integrity of layered semiconductor materials hence our focus on their opto-thermal characteristics. Lorentzian fitting of the Raman spectral intensity was used to plot and identify the peak positions at each temperature point. As indicated in Figure 4, both modes follow a systematic red shift with an increase in temperature. The full width at half maxima (FWHM) of the respective peaks exhibited broadening with increasing temperature (not shown) in line with general expectations. The 3D visualization of temperature evolutions of phonon vibrations (E'_{2g} and A_{1g}) in WS_2 is presented in the supplemental information Figure S2.

We first analyze the data using linear approximation to understand the observed temperature dependence of E'_{2g} and A_{1g} phonon modes. The shift in E'_{2g} and A_{1g} peaks with change in temperature as shown in these experiments indicates the effect of temperature in layered WS_2 under study. It is observed that from 80 to 460K, a plot of the peak positions of both modes across the temperature range, shows a modal decrease with increase in temperature. The resulting graph exhibits a tendency to a linear fit. Here we use the known standard linear equation to explain the experimental observations:[2,3,21,29]

$$\omega = \omega_0 + \chi T \qquad \text{(Equation 1)}$$

where ω_0 is the extrapolated modal frequency (y-intercept) (at 0 K) and χ is the empirical temperature coefficient.

The temperature coefficient (χ) is the linear tendency (slope) characteristic of the graph as seen in Figure 5. With our experimental data, we were able to establish the χ value for the exfoliated layer of WS_2. However, an in-depth analysis of the linear regression (Figure S3) shows a pattern forming between the residuals and fitted values. This suggests that the Raman shift for E'_{2g} and A_{1g} modes cannot be purely represented by linearity in Equation 1. Non-linearity can be introduced into Equation 1 from pure volumetric effects and pure temperature anharmonicity.[29] From 80 to 280K for A_{1g} in WS_2 Raman mode, the Newtonian
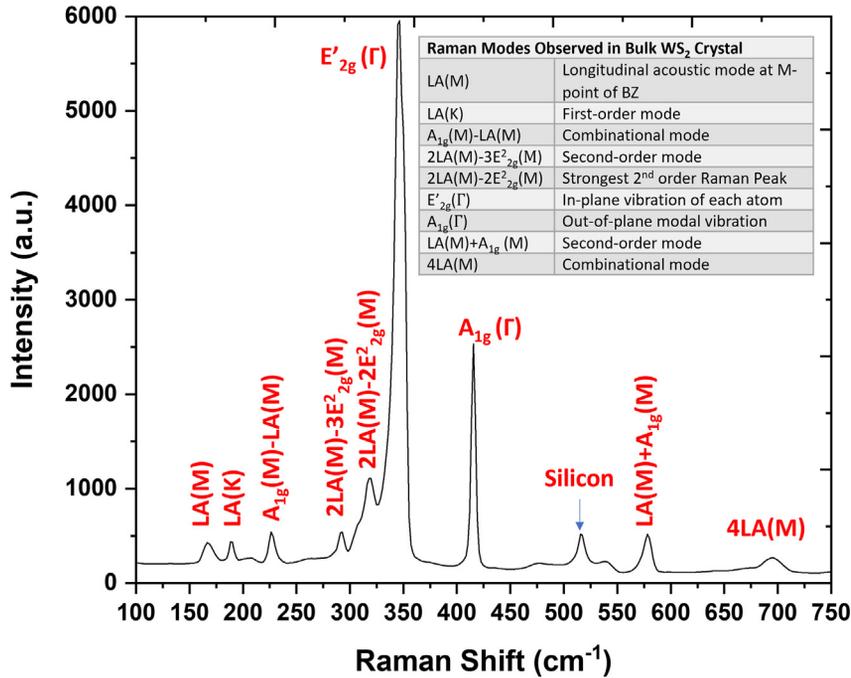
Figure 3. Room temperature Raman spectrum of bulk WS$_2$ crystal
The Raman spectrum of few-layer WS$_2$ at room temperature with an excitation of 532 nm. Inset shows various Raman modes.

peak-finding method for low-intensity measurements (<1000 a.u.) potentially fails due to the insufficient sampling rate of the Raman spectroscopy to capture small changes in the Raman shift. To best represent the true peak position for A$_{1g}$ from 80 to 280K, we approximate the peak position with a Gaussian distribution peak fitting method.

The variation in the peak position at low temperature while maintaining constant pressure ($\Delta\omega = \omega - \omega_0$) can be mathematically represented as follows:[29]

$$\Delta\omega = \left(\frac{d\omega}{dT}\right)_V \Delta T + \left(\frac{d\omega}{dV}\right)_T \Delta V$$

$$\Delta\omega = \chi_T \Delta T + \chi_V \Delta V \qquad \text{(Equation 2)}$$

where, χ_T = Self energy resulting from phonon-phonon interactions, the anharmonic effect and χ_V = Volume factor/contributions to the thermal expansion of the WS$_2$ lattice.

The assumption made in this experiment is that there is negligible expansion of the substrate used (SiO$_2$) to affect the overall thermal expansion of the subject matter. To explore further the volume factor and contribution of thermal expansion (χ_v), we use the Grüneisen constant model below:

$$\left(\frac{d\omega}{dV}\right)_T \Delta V = \omega_0 \left(exp\left(-3\gamma \int_0^T \alpha dT \right) - 1 \right) \qquad \text{(Equation 3)}$$

where, γ is the Grüneisen parameter, ω_0 is the frequency of the vibrational mode at zero Kelvin, and α the thermal expansion coefficient of a vibrational mode.

From the extrapolation of our experimental data through machine learning algorithms in Figure 5, we were able to establish frequency at zero Kelvin (0K) as below:

$$\omega_0\left(E'_{2g}\right) = 357.0 \pm 1.0 \text{ cm}^{-1}, \text{and}$$
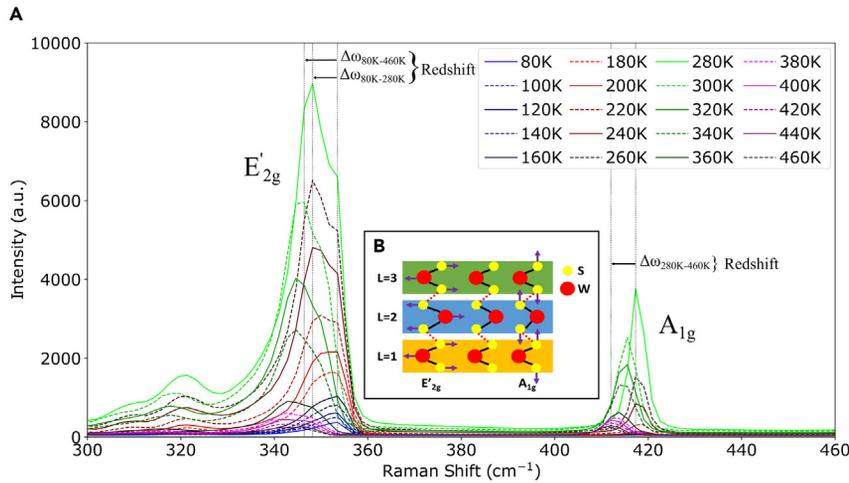
Figure 4. Temperature evolution of Raman spectra of bulk WS$_2$ crystal

(A) Raman spectra WS$_2$ film transferred on SiO$_2$/Si with a function of temperature (80–460K). Color and line style indicate the temperature from (80–460K) and vertical lines indicate the redshift from the peaks of Raman Shift 80/280K–280K and 460K for E'$_{2g}$ and A$_{1g}$. (B) Schematic of E'$_{2g}$ and A$_{1g}$ Raman vibrational modes.

$$\omega_0\left(A_{1g}\right) \ = \ 421.1 \pm 0.6 \ \text{cm}^{-1}$$

A 2019 report[30] provided the following thermal expansion coefficients for both vibrational modes of WS$_2$, which we will apply in the Grüneisen constant model:

$$\alpha\left(E'_{2g}\right) \ = \ 1.403 \times 10^{-5} \text{K}^{-1}$$

$$\alpha\left(A_{lg}\right) \ = \ 0.074 \times 10^{-5} \ \text{K}^{-1}$$

Finally, the Grüneisen constant for the two bulk WS$_2$ vibrational modes are presented as follows, courtesy of Ding et al. 2015.[31]

$$(\gamma)E'_{2g} \ = \ 0.92$$

$$(\gamma)A_{1g} \ = \ 2.17$$

Applying all the values into the model in the following:

$$\chi_V \Delta V = \omega_0 \left(\exp\left(-3\gamma \int_0^T \alpha dT \right) - 1 \right) \qquad \text{(Equation 4)}$$

We establish that the pure volume effects for each vibrational mode is as shown in the following:

$$\text{For } E'_{2g}: \quad \chi_V \Delta V = \omega_0 \left(e^{-3.87228 \times 10^{-5} \ K^{-1} \ T} - 1 \right)$$

$$\text{For } A_{1g}: \quad \chi_V \Delta V = \omega_0 \left(e^{-0.46028 \times 10^{-5} \ K^{-1} \ T} - 1 \right)$$

Figure 6 presents the variation of volume expansion coefficient with temperature derived from both E'$_{2g}$ (black) and A$_{1g}$ (blue) Raman vibrational modes.

Machine learning analysis

Machine learning analysis further proves we can dive deeper into analyzing the experimental data we have and furthermore, to see if it gives us any form of correlation to the experimental results.[32–35] This will invariably prove that predictions can be made from the data to increasingly solidify the case that phononic vibrations depend on volume expansion and temperature. Insights on how the data we obtained from our experiments can be used to further make future predictions and decisions moving forward are demonstrated.[36] We have employed here three machine learning models such as linear regression, decision

Table 1. Various Raman vibrational modes for bulk WS$_2$ at room temperature

Vibrational mode (cm^{-1})	Identity	Reason
171	LA(M)	Longitudinal acoustic mode at M-point of BZ
185	LA(K)	First-order mode
225	A$_{1g}$(M)-LA(M)	Combinational mode
297	2LA(M)-3E$^2_{2g}$(M)	Second-order mode
319.66	2LA(M)-2E$^2_{2g}$(M)	Strongest 2ND order Raman Peak
346.401	E$'_{2g}$(Γ)	In-plane vibration of each atom
415.44	A$_{1g}$(Γ)	Out-of-plane modal vibration
575	LA(M)+A$_{1g}$(M)	Second-order mode
700	4LA(M)	Combinational mode

tree regressor, and random forest regressor and compared their results to see which of these models provide accurate predictions. The detailed machine learning codes are provided in the supporting information section 4.

Linear regression

Linear regression is a machine learning algorithm based on supervised learning.[37] It performs a regression task. Regression models a target prediction value based on independent variables. Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (inputs) and y (output). A set of data consisting of n series of x and y values is given, where x is the independent variable which is the temperature in the range 80–460 K, and y is the dependent variable which is the variable that needs to be predicted (volume expansion). In other words, there is only one unique feature that is represented by x, and the target is represented by y. Considering a linear relation between variable x and variable y, we can write:

$$y = mx + b \qquad \text{(Equation 5)}$$

Here, m is the slope of the line, while b is the intercept at the y axis. In reality, because of noise or mismatch between data and model, there is an error €. Therefore, the modified linear equation is:

$$y = mx + b + € \qquad \text{(Equation 6)}$$

This model tries to minimize the sum of the mean squared errors (MSE) for the given datasets and this is the aim of linear regression. Since linear regression problem searches for a suitable model in the form of $y = mx + b + €$, a candidate slope, m and intercept, b are chosen first. For each recorded (x, y) pair, square of (y-mx-b) which is equal to square of €, is added to the total error. The line having the smallest total error is the best fit line. Figure 6 describes the linear regression-based predicted values of variation of volume expansion with temperature for both the E$'_{2g}$ (black) and A$_{1g}$ (blue) vibrational modes. Table 3 provides the R^2 values for the volume expansion predictions for the linear regression model.

Decision tree regressor

A decision tree is a machine learning technique. Typically, this method is used for classification tasks, but there is a possibility of applying it for a regression task. A decision tree method is an open and easy-to-understand method. For a given training vector x ∈ R^n (where *n* is a number of features) and a training label y ∈ R^l (*i* = 1,2,…l represents a number of labels) the regression tree algorithm recursively portions the features domain into smaller regions (separate classes). It is important to choose the correct metrics for the best data split and determining when a tree node should become a terminal.

Table 2. Peak differences and intensity ratios analysis for selected vibration modes

2LA (cm^{-1})	E$'_{2g}$ (cm^{-1})	A$_{1g}$ (cm^{-1})	A$_{1g}$-E$'_{2g}$ (cm^{-1})	I$_{2LA}$/I$_{A1g}$	I$_{A1g}$/I$_{E'2g}$
319.66	346.401	415.44	69.143	0.43749	0.425015
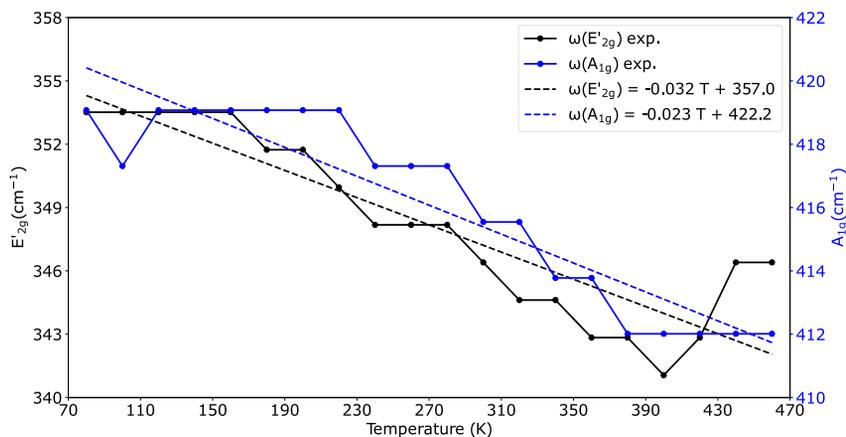
Figure 5. Dependence of phonon frequencies on temperature
Variation of Raman vibrational frequencies (E'_{2g} and A_{1g}) with temperature. Color indicates the Raman shift for E'_{2g} (black, left y axis) and A_{1g} (blue, right y axis). Dashed lines indicate the linear regression.

Since in this work a decision tree algorithm is used for a regression task, then the target is a continuous or quantitative value. Thus, for node m, which represents a region R_m with N_m observations, MSE or mean absolute error (MAE) are possible regression criteria to minimize impurity function (H). As for determining locations for future data splits, a test size of test_size = 0.2, and a random state random_state = 11 were implemented on the model. Minimization of an error can be done by using mean values of the terminal nodes for MSE and we can focus on getting a high R^2 or accuracy score to confirm the efficiency of the model in prediction. Here, the decisions made are based on the nodes with the lowest MSE values. Figure 7 highlights how the data is split according to the sample size and MSE value. Table 3 provides the R^2 values for the volume expansion predictions for decision tree regressor model.

Random forest regressor
Random Forest is a supervised learning algorithm. It builds an ensemble of decision trees, usually trained with the bagging method. The main idea of the bagging method is that a combination of learning models increases the overall accuracy of the desired outcome and result. Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.[38] A good advantage of random forest is that it can be used for both classification and regression problems, which form
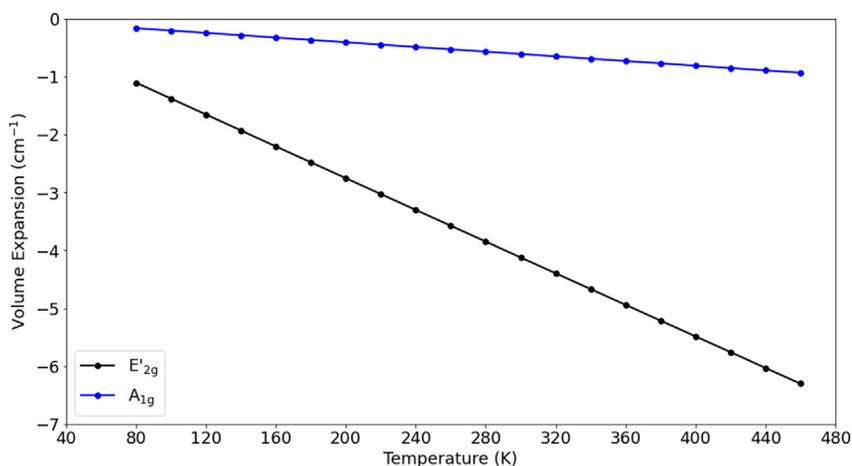
Figure 6. Dependence of volume expansion coefficient on temperature
Variation of volume expansion coefficient with temperature derived from both E'_{2g} (black) and A_{1g} (blue) Raman vibrational modes.

Table 3. Accuracy scores for each of the machine learning models employed in this study

RFR Model		DT Model		Linear Regression					
R^2		R^2		R^2		MSE		RMSE	
E'_{2g}Cv	A_{1g}	E'_{2g}Cv	A_{1g}	E'_{2g}Cv	A_{1g}	E'_{2g}Cv	A_{1g}	E'_{2g}Cv	A_{1g}
98.44	98.98	93.08	93.11	99.98	99.98	0.00000646	0.00000000197	0.00245	0.0000444

the majority of current machine learning systems. Here, we focus more on a regression problem that seeks to predict a discreet and quantitative result from a set of given inputs. Random forests seek to affect such correlation reduction by a further injection of randomness. Instead of determining the optimal split of a given node of a (constituent) tree by evaluating all allowable splits on all covariates, as is done with single tree methods (decision trees) or bagging, a subset of the covariates drawn at random, is employed.

A random forest is a collection of tree predictors h ($\mathbf{x}; \theta_k$), k = 1 …. .K where $\mathbf{x}$ represents the observed input (covariate) vector of length p with associated random vector X and the θ_k are independent and identically distributed random vectors. Here, we try to focus on the regression setting for which we have a numerical quantitative outcome Y but make some points of contact with classification (categorical outcome) problems. The observed (training) data are assumed to be independently drawn from the joint distribution of (X, Y) and comprise n (p +1) tuples (x_1, y_1),… … …,(x_n, y_n).

Here our $\mathbf{x}$, which is the observed input, represents our independent variable (temperature range 80 - 460K) and our dependent variable which is the desired prediction outcome. Random forest regressor with a max_depth of "4" was used in this computation. A test size of test_size = 0.2, and a random state random_state = 11. One may suspect that the computational complexity of an ensemble of, say, 100 trees, is probably 100 times that of a single tree. This is actually not true. The random forest algorithm can be very efficient, especially when the number of descriptors is very large. The efficiency of the algorithm, compared to that of growing a single decision tree, comes from two differences between the two algorithms. Table 3 shows the accuracy score of the RFR model. Cross-validation is a technique in machine learning used to assess a model and test its performance and accuracy. It involves reserving a specific sample of a dataset on which the model is not trained. Later on, the model is tested on this sample to evaluate it. The original data sample is randomly divided into several subsets. The machine learning model trains on all subsets, except one. After training, the model is tested by making predictions on the remaining subset.

We can see from Figures 8A and 8B how the Training score and Cross Validation vary with increasing training size. Here we notice in Figure 8A, at a training size of 2, the Training score and Cross validation are at 1 and -40, respectively. Moving further forward, we can see that as the training size increase to 6 and more, no other significant change occurs with the two scores. The same thing can be said of Figure 8B. At a training size of 10 and more, no further changes can be seen within the scores. Table 3 shows the R^2 accuracy score for all the models, the mean square error and the root-mean-square error for the linear model.
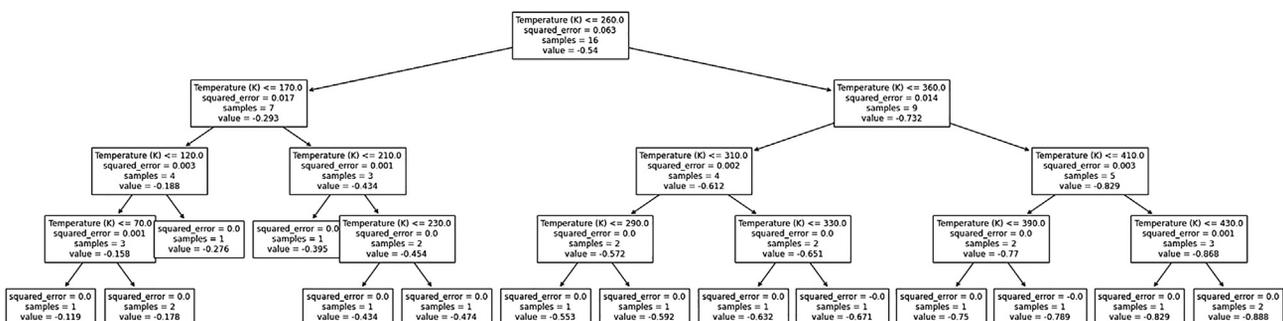
Figure 7. Decision tree classification
Decisions made are based on the nodes with the lowest MSE values.
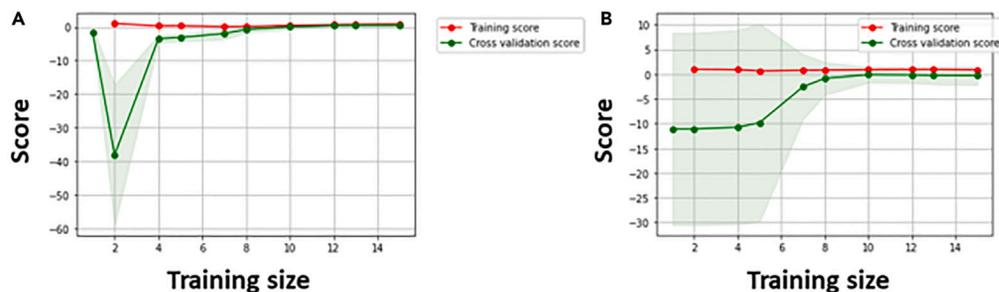
Figure 8. Random forest training and cross validation

(A) Training and Cross Validation for A_{1g} and (B) Training and Crossing Validation score for E'_{2g}.

Conclusion

In this study, we have demonstrated thermal-induced anharmonic phonon dynamics in WS_2 van der Waals crystal. We found that both the first-order phonon vibrational modes E'_{2g} and A_{1g} follow a systematic red shift with an increase in temperature and the FWHM of the respective peaks exhibited broadening with increasing temperature. We developed a linear regression ML model to train, test, and predict these frequencies at different temperatures within the range of our experiment. We can conclude that the RFR model with the highest accuracy score is the best fit to make predictions within the nature of the dataset we have acquired. By taking peak intensity and frequency information and temperature into consideration at the same time, a high accuracy rate is obtained. The method developed in this work can also be used for other 2D materials and can provide a valuable reference for material characterization in several fields. This fundamental study will lead (i) to unveil the dependence of temperature and laser power on in-plane and out-of-plane (E'_{2g} and A_{1g}) Raman shifts of TMDs in order to study the thermal conductivity, hot carrier diffusion coefficient, and thermal expansion coefficient and (ii) to understand the functionalities of TMD-based optoelectronic devices at extreme conditions.

Limitations of the study

In our study, the sampling rate of the Raman spectroscopy limits the resolution of the Raman shift at low temperature. We have employed Gaussian peak fitting to approximate the peak, but finer-tuned statistical and optimization procedures could improve the Raman peak positions at low temperatures.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Temperature-dependent Raman spectroscopy
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Machine learning

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.107174.

AUTHOR CONTRIBUTIONS

S.K.B. conceived the idea. C.O., I.J., and S.K.B designed and carried out the present work and drafted the manuscript. A.C., N.S., and H.M. analyzed the experimental and machine learning data. M.M. reviewed the work. All authors revised and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests in this work.

REFERENCES

1. Lin, J., Guo, L., Huang, Q., Jia, Y., Li, K., Lai, X., and Chen, X. (2011). Anharmonic phonon effects in Raman spectra of unsupported vertical graphene sheets. Phys. Rev. B *83*, 125430. https://doi.org/10.1103/PhysRevB.83.125430.

2. Late, D.J., Shirodkar, S.N., Waghmare, U.V., Dravid, V.P., and Rao, C.N.R. (2014). Thermal Expansion, Anharmonicity and Temperature-Dependent Raman Spectra of Single- and Few-Layer MoSe2 and WSe2. ChemPhysChem *15*, 1592–1598. https://doi.org/10.1002/cphc.201400020.

3. Najmaei, S., Ajayan, P.M., and Lou, J. (2013). Quantitative analysis of the temperature dependency in Raman active vibrational modes of molybdenum disulfide atomic layers. Nanoscale *5*, 9758–9763. https://doi.org/10.1039/C3NR02567E.

4. Tang, H., and Herman, I.P. (1991). Raman microprobe scattering of solid silicon and germanium at the melting temperature. Phys. Rev. B *43*, 2299–2304. https://doi.org/10.1103/PhysRevB.43.2299.

5. Zobeiri, H., Xu, S., Yue, Y., Zhang, Q., Xie, Y., and Wang, X. (2020). Effect of temperature on Raman intensity of nm-thick WS2: combined effects of resonance Raman, optical properties, and interface optical interference. Nanoscale *12*, 6064–6078. https://doi.org/10.1039/C9NR10186A.

6. Zouboulis, E.S., and Grimsditch, M. (1991). Raman scattering in diamond up to 1900 K. Phys. Rev. B *43*, 12490–12493. https://doi.org/10.1103/PhysRevB.43.12490.

7. Piscanec, S., Lazzeri, M., Mauri, F., Ferrari, A.C., and Robertson, J. (2004). Kohn Anomalies and Electron-Phonon Interactions in Graphite. Phys. Rev. Lett. *93*, 185503. https://doi.org/10.1103/PhysRevLett.93.185503.

8. Balandin, A.A., and Nika, D.L. (2012). Phononics in low-dimensional materials. Mater. Today *15*, 266–275. https://doi.org/10.1016/S1369-7021(12)70117-7.

9. Liu, X., and Zhang, Y.-W. (2018). Thermal properties of transition-metal dichalcogenide. Chin. Phys. B *27*, 034402. https://doi.org/10.1088/1674-1056/27/3/034402.

10. Duan, X., Wang, C., Pan, A., Yu, R., and Duan, X. (2015). Two-dimensional transition metal dichalcogenides as atomically thin semiconductors: opportunities and challenges. Chem. Soc. Rev. *44*, 8859–8876. https://doi.org/10.1039/C5CS00507H.

11. Chhowalla, M., Shin, H.S., Eda, G., Li, L.-J., Loh, K.P., and Zhang, H. (2013). The chemistry of two-dimensional layered transition metal dichalcogenide nanosheets. Nat. Chem. *5*, 263–275. https://doi.org/10.1038/nchem.1589.

12. Sung, J., Shin, D., Cho, H., Lee, S.W., Park, S., Kim, Y.D., Moon, J.S., Kim, J.-H., and Gong, S.-H. (2022). Room-temperature continuous-wave indirect-bandgap transition lasing in an ultra-thin WS2 disk. Nat. Photonics *16*, 792–797. https://doi.org/10.1038/s41566-022-01085-w.

13. Tan, Q.-H., Li, Y.-M., Lai, J.-M., Sun, Y.-J., Zhang, Z., Song, F., Robert, C., Marie, X., Gao, W., Tan, P.-H., and Zhang, J. (2023). Quantum interference between dark-excitons and zone-edged acoustic phonons in few-layer WS2. Nat. Commun. *14*, 88. https://doi.org/10.1038/s41467-022-35714-3.

14. Behura, S., Chang, K.C., Wen, Y., Debbarma, R., Nguyen, P., Che, S., Deng, S., Seacrist, M.R., and Berry, V. (2017). WS2/Silicon Heterojunction Solar Cells: A CVD Process for the Fabrication of WS2 Films on p-Si Substrates for Photovoltaic and Spectral Responses. IEEE Nanotechnol. Mag. *11*, 33–38. https://doi.org/10.1109/MNANO.2017.2676184.

15. Debbarma, R., Behura, S.K., Wen, Y., Che, S., and Berry, V. (2018). WS2-induced enhanced optical absorption and efficiency in graphene/silicon heterojunction photovoltaic cells. Nanoscale *10*, 20218–20225. https://doi.org/10.1039/C8NR03194K.

16. Taube, A., Judek, J., Jastrzębski, C., Duzynska, A., Świtkowski, K., and Zdrojek, M. (2014). Temperature-Dependent Nonlinear Phonon Shifts in a Supported MoS2 Monolayer. ACS Appl. Mater. Interfaces *6*, 8959–8963. https://doi.org/10.1021/am502359k.

17. Yang, T., Huang, X., Zhou, H., Wu, G., and Lai, T. (2016). Excitation mechanism of A1g mode and origin of nonlinear temperature dependence of Raman shift of CVD-grown mono- and few-layer MoS2 films. Opt Express *24*, 12281–12292. https://doi.org/10.1364/OE.24.012281.

18. Huang, X., Gao, Y., Yang, T., Ren, W., Cheng, H.-M., and Lai, T. (2016). Quantitative Analysis of Temperature Dependence of Raman shift of monolayer WS2. Sci. Rep. *6*, 32236. https://doi.org/10.1038/srep32236.

19. M, T., and Late, D.J. (2014). Temperature Dependent Phonon Shifts in Single-Layer WS2. ACS Appl. Mater. Interfaces *6*, 1158–1163. https://doi.org/10.1021/am404847d.

20. Lanzillo, N.A., Glen Birdwell, A., Amani, M., Crowne, F.J., Shah, P.B., Najmaei, S., Liu, Z., Ajayan, P.M., Lou, J., Dubey, M., et al. (2013). Temperature-dependent phonon shifts in monolayer MoS2. Appl. Phys. Lett. *103*, 093102. https://doi.org/10.1063/1.4819337.

21. Sahoo, S., Gaur, A.P.S., Ahmadi, M., Guinel, M.J.F., and Katiyar, R.S. (2013). Temperature-Dependent Raman Studies and Thermal Conductivity of Few-Layer MoS2. J. Phys. Chem. C *117*, 9042–9047. https://doi.org/10.1021/jp402509w.

22. Molas, M.R., Nogajewski, K., Potemski, M., and Babiński, A. (2017). Raman scattering excitation spectroscopy of monolayer WS2. Sci. Rep. *7*, 5036. https://doi.org/10.1038/s41598-017-05367-0.

23. Shuai, Q., Hang, Y., Zongqi, B., Gang, P., and Xueao, Z. (2017). Identifying the Number of WS2 Layers via Raman and Photoluminescence Spectrum (Atlantis Press), pp. 1408–1413.

24. Gaur, A.P.S., Sahoo, S., Scott, J.F., and Katiyar, R.S. (2015). Electron–Phonon Interaction and Double-Resonance Raman Studies in Monolayer WS2. J. Phys. Chem. C *119*, 5146–5151. https://doi.org/10.1021/jp512540u.

25. Peimyoo, N., Shang, J., Yang, W., Wang, Y., Cong, C., and Yu, T. (2015). Thermal conductivity determination of suspended mono- and bilayer WS2 by Raman spectroscopy. Nano Res. *8*, 1210–1221. https://doi.org/10.1007/s12274-014-0602-0.

26. Berkdemir, A., Gutiérrez, H.R., Botello-Méndez, A.R., Perea-López, N., Elías, A.L., Chia, C.-I., Wang, B., Crespi, V.H., López-Urías, F., Charlier, J.-C., et al. (2013). Identification of individual and few layers of WS2 using Raman Spectroscopy. Sci. Rep. *3*, 1755. https://doi.org/10.1038/srep01755.

27. Bhattacharya, S., Veluthandath, A.V., Huang, C.C., Murugan, G.S., and Bisht, P.B. (2020). Effect of coating few-layer WS2 on the Raman spectra and whispering gallery modes of a microbottle resonator. J. Opt. *22*, 105003. https://doi.org/10.1088/2040-8986/abad50.

28. Ding, Y., Zheng, W., Lu, X., Liang, Y., Zhu, Y., Jin, M., and Huang, F. (2020). Raman Tensor of Layered SnS2. J. Phys. Chem. Lett. *11*, 10094–10099. https://doi.org/10.1021/acs.jpclett.0c03024.

29. Sinha, S., Sathe, V., and Arora, S.K. (2019). Temperature dependent Raman investigations of few-layered WS2 nanosheets. Solid State Commun. *298*, 113626. https://doi.org/10.1016/j.ssc.2019.04.013.

30. Kumar, D., Singh, B., Kumar, P., Balakrishnan, V., and Kumar, P. (2019). Thermal expansion coefficient and phonon dynamics in coexisting allotropes of monolayer WS2 probed by Raman scattering. J. Phys. Condens. Matter *31*, 505403. https://doi.org/10.1088/1361-648X/ab3fec.

31. Ding, Y., and Xiao, B. (2015). Thermal expansion tensors, Grüneisen parameters and phonon velocities of bulk MT2 (M = W and Mo; T = S and Se) from first principles calculations. RSC Adv. *5*, 18391–18400. https://doi.org/10.1039/C4RA16966B.

32. Wang, X., Han, D., Hong, Y., Sun, H., Zhang, J., and Zhang, J. (2019). Machine Learning Enabled Prediction of Mechanical Properties of Tungsten Disulfide Monolayer. ACS Omega *4*, 10121–10128. https://doi.org/10.1021/acsomega.9b01087.

33. Madden, M., and Ryder, A. (2003). Machine Learning Methods for Quantitative Analysis of Raman Spectroscopy Data (SPIE).

34. Hart, G.L.W., Mueller, T., Toher, C., and Curtarolo, S. (2021). Machine learning for alloys. Nat. Rev. Mater. *6*, 730–755. https://doi.org/10.1038/s41578-021-00340-w.

35. Mao, Y., Dong, N., Wang, L., Chen, X., Wang, H., Wang, Z., Kislyakov, I.M., and Wang, J. (2020). Machine Learning Analysis of Raman Spectra of MoS2. Nanomaterials *10*, 2223.

36. Maddah, H.A., Berry, V., and Behura, S.K. (2020). Cuboctahedral stability in Titanium halide perovskites via machine learning. Comput. Mater. Sci. *173*, 109415. https://doi.org/10.1016/j.commatsci.2019.109415.

37. Ari, B., and Güvenir, H. (2002). Clustered linear regression. Knowl. Base Syst. *15*, 169–175. https://doi.org/10.1016/S0950-7051(01)00154-X.

38. Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., and Feuston, B.P. (2003). Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. J. Chem. Inf. Comput. Sci. *43*, 1947–1958. https://doi.org/10.1021/ci034160g.

39. Seabold, S., and Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. In Proceedings of the 9th Python in Science Conference.

40. Fabian Pedregosa, G.V., Gramfort, A., Michel, V., Bertrand, T., Grisel, O., Blondel, M., Peter, P., Weiss, R., Vincent, D., Vanderplas, J., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Other		
WS$_2$	Milipore Sigma	808806-1EA
Software and algorithms		
Linear regression	statsmodels: Econometric and statistical modeling with python[39]	https://www.statsmodels.org/stable/api.html
Decision tree regressor	Scikit-learn: Machine Learning in Python[40]	https://scikit-learn.org/stable/index.html
Random forest regressor	Scikit-learn: Machine Learning in Python[40]	https://scikit-learn.org/stable/index.html

RESOURCE AVAILABILITY

Lead contact

Further information and reasonable request for resources will be handled by our lead contact Sanjay Behura (sbehura@sdsu.edu).

Materials availability

This study did not generate new unique materials. All chemicals were obtained from commercial resources and used as received.

Data and code availability

- All data reported in this paper will be shared by lead contact upon reasonable request.

- This paper does not report original code.

- Any additional information from this paper can be requested from the lead contact.

METHOD DETAILS

Temperature-dependent Raman spectroscopy

Thin layers of WS$_2$ are exfoliated from high-quality layered bulk crystal using mechanical exfoliation methods and then transferred onto SiO$_2$-coated Si substrates. The transferred WS$_2$ layer on SiO$_2$/Si is then studied using temperature-dependent Raman spectroscopy. The temperature from 80K to 460K at a 20K interval was varied to study the phonon vibrational modes using 532 nm laser excitation as presented in Figure 1. The laser intensity and laser power are ∼6000 a.u. [E$'_{2g}$] and ∼2500 a.u. [A$_{1g}$], and 10 mW, respectively. The temperature of the mounted samples was controlled by a liquid nitrogen cooled stage. The use of inelastic scattering-based Raman spectroscopy is conventionally accepted in the structural, optical, and electronic characterization of layered materials. Furthermore, the temperature dependent phonon vibrational mode investigations provide an additional dimension to the Raman spectroscopy tool as the pressure of a gas of anharmonic phonons is temperature dependent. As shown in Figure 1, the scattered light from the temperature-controlled sample chamber is acquired to model the temperature evolutions of phonon vibrations followed by building a machine learning framework to study the optothermal properties of few-layer WS$_2$.

QUANTIFICATION AND STATISTICAL ANALYSIS

Machine learning

The data gained from the experimental procedures was saved in .csv format and analyzed using the Linear regression model. It was noticed that at a certain point, there was a linear relationship between the frequencies and the temperature. For these experiments, 19 data points were collected from a temperature range of 80 — 460K at 20K intervals. Pandas, NumPy, and seaborn libraries from Jupyter lab was used. Frequencies (E$'_{2g}$ and A$_{1g}$) were seen as the dependent variables which are the same variable that need to be predicted. Here, Y = "E$'_{2g}$ and A$_{1g}$" and X = "Temperature". Model-selection module was imported from

sklearn for training and testing data. For the training and test data, a test size of 0.2 was used and a random state of 11 was used as well. The linear regression model was imported from linear_model library on Jupyter lab. The model was fit into the variables y and x as the Frequency and Temperature, respectively. This model was implemented independently of E'_{2g} and A_{1g}. The data sets were split into testing and training sets with a test_size of 0.2. Two more machine learning models were employed to determine which model will provide the best accuracy score for our data set. Random forest regressor with a max depth of '5' and Decision Tree regressor with a max depth of '4'. The libraries were imported to Jupyter lab, and the modules were also imported to implement this. Our data set was imported to python and a linear regression model was run with data values of temperature and frequencies separately using python library statsmodel.api. The summary generated an intercept and slope value. For E'_{2g} the intercept and slope are 422.2 and -0.023, respectively with adjusted R^2 value of 0.844. For A_{1g} the intercept and slope are 357.0and -0.032, respectively with adjusted R^2 value of 0.798. For the volume expansion on vibrational mode A_{1g}, intercept and slope are -0.000235 and -0.001973, respectively. Similarly, for vibrational mode E'_{2g} we have -0.01348 and -0.01366, respectively.