Supporting Information

Micro-SORS and machine learning for the non-invasive study of subsurface pigment degradation

A. Lux[1,2,3], J. Schlanz[4], A. Botteon[1], L. Monico[5], P. Matousek[3], R. Dima[4], C. Conti[1] , P. Strobbia[4]

[1]*Institute of Heritage Science, National Research Council (CNR-ISPC), Via Cozzi 53, 20125, Milan, Italy;*

[2]*Sapienza University of Rome, Faculty of Literature, Department of Classics, Piazzale Aldo Moro 5, 00185, Rome, Italy;*

[3]*Central Laser Facility, Research Complex at Harwell, STFC Rutherford Appleton Laboratory, UKRI, Harwell Oxford, OX11 0QX, United Kingdom;*

[4]*Department of Chemistry, University of Cincinnati, 201 Crosley Tower, Cincinnati, United States;*

[5]*Institute of Chemical Sciences and Technologies "Giulio Natta", National Research Council, (CNR-SCITEC), Via Elce di Sotto 8, 06123, Perugia, Italy.*

- Results and discussion
  - Red lead
    - *Figure S7…Reference spectra of the red lead sample;*
    - *Figure S8…Example spectra of the aged sample, both raw (red one) and baselined (blue one), and of the unaged sample after baseline (green one);*
    - *Figure S9…Full sets of spectra of the a) 75 μm, b) 150 μm, c) 225 μm and d) 300 μm defocusing steps of red lead;*
    - *Figure S10…Full sets of a) non-subtracted and b) subtracted spectra of the surface level of red lead;*
    - *Figure S11…Raman maps of the unaged (left panels) and aged (right panels) red lead. The values of the colorbars indicate the maximum and minimum of each individual level with respect to the general scale;*
    - *Figure S12…The average red lead unaged and aged spectra with the three selected features denoted as yellow stars;*
    - *Figure S13…Summary of the red lead unaged vs aged Extra Trees model results. (a) SHAP beeswarm plot showing the selected three features. A positive SHAP value explains whether features with intensities that increase (magenta) or decrease (blue) are more aged. (b) The confusion matrix of the ten unseen unaged and ten unseen aged spectra;*
    - *Figure S14…a) The average spectra of the eleven clusters using the red lead unaged and aged samples with the three selected features; b) The distribution of the unaged (0) and aged (1) red lead samples into eleven clusters. Orange represents a high number of spectra in a cluster, and yellow represents a low number of spectra in a cluster;*
    - *Figure S15…The percent degradation of the unaged and aged red lead samples using feature 1052. A more degraded area of the sample is represented by white, and a more intact area of the sample is represented by dark red.*
    - *Figure S16…The percent degradation of the unaged and aged red lead samples using features 147 (a) and 561 (b). A more degraded area of the sample is represented by white, and a more intact area is represented by dark red.*
    - *Figure S17…Summary of the clusters using the red lead unaged and aged samples with the three selected features. (a) Average spectrum of cluster 1, which represents the majority of the unaged spectra. (b) Average spectrum of cluster 6, which contains the unaged spectra in the 7x25 and 11x19 positions in the feature 1052 unaged heatmap. (c) Average spectrum of cluster 7, which contains the unaged spectrum in the 3x10 position in the feature 1052 unaged heatmap. (d-f) Unaged spectra that were not placed into cluster 1. The yellow star denotes the position of feature 1052;*
    - *Figure S18…Raman spectra of the selected three red (intact) and three white (degraded) cells from the feature 1051.89 aged heatmap. (a-c) Raman spectra of the intact cells. (d-f) Raman spectra of the degraded cells;*
    - *Figure S19…Comparison of the experimental and machine learning unaged and aged red lead heatmaps. Areas that agree between the two heatmaps are represented by dark red, and areas that differ are white;*
    - *Figure S20…The average aged defocusing steps for the five defocusing steps (classes 0-4) with the 3 selected features denoted as yellow stars. Class 0*

corresponds to the surface and class 4 corresponds to the deepest defocusing step;

- *Figure S21…SHAP summary plot of the red lead defocusing steps using the Linear Regression model;*
- *Figure S22…The red lead defocusing steps Linear Regression model outputs using the ten spectra from each class of the holdout set. Each red line represents how far away the true and predicted values are from each other;*
- *Figure S23…The distribution of the five red lead defocusing steps into twelve clusters. Orange represents a high number of spectra in a cluster, and yellow represents a low number of spectra in a cluster;*
- *Figure S24…Summary of the unsupervised clusters using the red lead defocusing steps with the three selected figures. (a) Average spectra of each of the 12 clusters. (b) Cluster 1 showing the high intensity on average for feature 1050. (c) Cluster 7 showing the low intensity on average for feature 1050;*
- *Figure S25… The percent degradation of the red lead defocusing steps using feature 1050. A more degraded area of the sample is represented by white, and a more intact area is represented by dark red;*
- *Figure S26…a) The percent degradation of the red lead defocusing steps using feature 550. A more degraded area of the sample is represented by white, and a more intact area is represented by dark red; b) The overall percent degradation of the red lead defocusing steps using features 550 and 1050. A more degraded area of the sample is represented by white, and a more intact area is represented by dark red;*
- *Figure S27…Raman spectra of two highly degraded (white) cells in the class 0 red lead defocusing step heatmap showing how the spectra change as the defocusing steps (classes) increase. (a-e) The Raman spectra in each defocusing step corresponding to the 0x37 cell in the heatmap. (f-j) The Raman spectra in each defocusing step corresponding to the 9x33 cell in the heatmap. The yellow star denotes the position of feature 1050;*
- *Figure S28…Summary of the selected cells in the red lead defocusing step heatmaps to show the intensity (and degradation) patterns. (a-b) The selected degraded (white) cells in the heatmaps. (c-d) The selected intact (red) cells in the heatmaps;*
- *Figure S29…Raman spectra of two highly intact (red) cells in the class 0 red lead defocusing step heatmap showing how the spectra change as the defocusing steps (classes) increase. (a-e) The Raman spectra in each defocusing step corresponding to the 7x6 cell in the heatmap. (f-j) The Raman spectra in each defocusing step corresponding to the 1x0 cell in the heatmap. The yellow star denotes the position of feature 1050;*
- *Figure S30… Comparison of the experimental and machine learning red lead defocusing steps heatmaps. Areas that agree between the two heatmaps are represented by dark red, and areas that differ are white.*

- Orpiment
  - *Figure S31…Full sets of spectra of the a) 100 µm, b) 200 µm, c) 300 µm and d) 400 µm defocusing steps of orpiment;*
  - *Figure S32…Raman maps of the unaged (left panels) and aged (right panels) orpiment, plotting pararealgar vs orpiment. The values of the colorbars indicate*

the maximum and minimum of each individual level with respect to the general scale;

- *Figure S33…Raman maps of the unaged (left panels) and aged (right panels) orpiment, plotting realgar vs orpiment. The values of the colorbars indicate the maximum and minimum of each individual level with respect to the general scale;*
- *Figure S34…Superimposition of the pararealgar and arsenolite Raman maps of orpiment. Red pixels show only pararealgar presence, green pixels indicate arsenolite, brown the presence of both, and yellow complete absence;*
- *Figure S35…Summary of the orpiment unaged vs aged Extra Trees model results. (a) SHAP beeswarm plot showing the selected five features. A positive SHAP value explains whether features with intensities that increase (magenta) or decrease (blue) are more aged. (b) The confusion matrix of the ten unseen unaged and ten unseen aged spectra;*
- *Figure S36…a) The average spectra of the nine clusters using the unaged and aged orpiment samples with the five selected features; b) The distribution of the unaged and aged orpiment lead samples into nine clusters. Orange represents a high number of spectra in a cluster, and yellow represents a low number of spectra in a cluster.*
- *Figure S37…Summary of the orpiment unaged vs aged unsupervised clusters that only contain unaged labelled spectra. (a) Spectra from cluster 4, which contains realgar. (b) Spectra from cluster 5, which contains realgar. (c) Spectra from cluster 8, which contains orpiment.*
- *Figure S38…Summary of cluster 9 from the orpiment unaged vs aged unsupervised learning. (a) Spectra from cluster 9, which contains aged and unaged labelled spectra. The spectra contain realgar and possibly traces of pararealgar but no arsenolite. (b) The aged labelled spectra from cluster 9 showing the possible traces of pararealgar;*
- *Figure S39…Summary of cluster 9 from the orpiment unaged vs aged unsupervised learning. (a) Spectra from cluster 9, which contains aged and unaged labelled spectra. The spectra contain realgar and possibly traces of pararealgar but no arsenolite. (b) The aged labelled spectra from cluster 9 showing the possible traces of pararealgar;*
- *Figure S40…Summary of cluster 7 from the orpiment unaged vs aged unsupervised learning. (a) Spectra from cluster 7, which contains unaged and aged spectra. The spectra contain orpiment and very small arsenolite peaks. (b) The unaged spectra from cluster 7 showing no presence of arsenolite. (c) The aged spectra from cluster 7 showing very small arsenolite peaks;*
- *Figure S41…The percent degradation of the unaged and aged orpiment samples using features 368 (a), 330 (b), 292 (c), and 192 (d). A more degraded area of the sample is represented by white, and a more intact area is represented by dark red;*
- *Figure S42…Raman spectra of the selected three red (intact) and three white (degraded) cells from the feature 269 aged orpiment heatmap. (a-c) Raman spectra of the intact cells. (d-f) Raman spectra of the degraded cells;*

- *Figure S43…Comparison of the experimental and machine learning unaged and aged orpiment heatmaps. Areas that agree between the two heatmaps are represented by dark red, and areas that differ are white;*
- *Figure S44…The distribution of the five defocusing steps into nine clusters. Orange represents a high number of spectra in a cluster, and yellow represents a low number of spectra in a cluster;*
- *Figure S45…Summary of the unsupervised clusters using the orpiment defocusing steps with the 21 selected features. (a) Average spectra of each of the nine clusters. (b) Cluster 5 showing the possible presence of pararealgar. (c) Cluster 9 showing the presence of realgar. (d) Cluster 8 showing spectra with possible traces of pararealgar and realgar;*
- *Figure S46…The percent degradation of the orpiment defocusing steps using features 155 (a) and 157 (b). A more degraded area of the sample is represented by white, and a more intact area is represented by dark red;*
- *Figure S47…The percent degradation of the orpiment defocusing steps using features 186 (a) and 204 (b). A more degraded area of the sample is represented by white, and a more intact area is represented by dark red;*
- *Figure S48…The percent degradation of the orpiment defocusing steps using features 232 (a) and 238 (b). A more degraded area of the sample is represented by white, and a more intact area is represented by dark red;*
- *Figure S49…The percent degradation of the orpiment defocusing steps using features 240 (a) and 269 (b). A more degraded area of the sample is represented by white, and a more intact area is represented by dark red;*
- *Figure S50…The percent degradation of the orpiment defocusing steps using features 270 (a) and 273 (b). A more degraded area of the sample is represented by white, and a more intact area is represented by dark red;*
- *Figure S51…The percent degradation of the orpiment defocusing steps using features 274 (a) and 278 (b). A more degraded area of the sample is represented by white, and a more intact area is represented by dark red;*
- *Figure S52…The percent degradation of the orpiment defocusing steps using features 294 (a) and 295 (b). A more degraded area of the sample is represented by white, and a more intact area is represented by dark red;*
- *Figure S53…The percent degradation of the orpiment defocusing steps using features 312 (a) and 313 (b). A more degraded area of the sample is represented by white, and a more intact area is represented by dark red;*
- *Figure S54…The percent degradation of the orpiment defocusing steps using features 368 (a) and 384 (b). A more degraded area of the sample is represented by white, and a more intact area is represented by dark red;*
- *Figure S55…The percent degradation of the orpiment defocusing steps using (a) feature 385 or (b) all features. A more degraded area of the sample is represented by white, and a more intact area is represented by dark red;*
- *Figure S56…Raman spectra of two degraded (white) cells in the class 0 orpiment defocusing step heatmap showing how the spectra change as the defocusing steps (classes) increase. (a-e) The Raman spectra in each defocusing step corresponding to the 9x12 cell in the heatmap. (f-j) The Raman spectra in each defocusing step corresponding to the 0x10 cell in the heatmap. The yellow star denotes the position of feature 268;*

- *Figure S57…Summary of the selected cells in the orpiment defocusing step heatmaps to show the intensity (and degradation) patterns. (a-b) The selected degraded (white) cells in the heatmaps. (c-d) The selected intact (red) cells in the heatmaps. The star denotes the position of feature 268;*
- *Figure S58…Raman spectra of two intact (dark red) cells in the class 0 orpiment defocusing step heatmap showing how the spectra change as the defocusing steps (classes) increase. (a-e) The Raman spectra in each defocusing step corresponding to the 6x10 cell in the heatmap. (f-j) The Raman spectra in each defocusing step corresponding to the 4x25 cell in the heatmap. The yellow star denotes the position of feature 268;*
- *Figure S59…Summary of the very white 6x36 cell in class 1 of the feature 238 heatmap. (a-e) The Raman spectra in each defocusing step corresponding to the 6x36 cell in the heatmap. (f) The Raman spectra for each class. The stars denote the position of feature 238;*
- *Figure S60…Comparison of the experimental and machine learning orpiment defocusing steps heatmaps. Areas that agree between the two heatmaps are represented by dark red, and areas that differ are white.*

Materials and methods

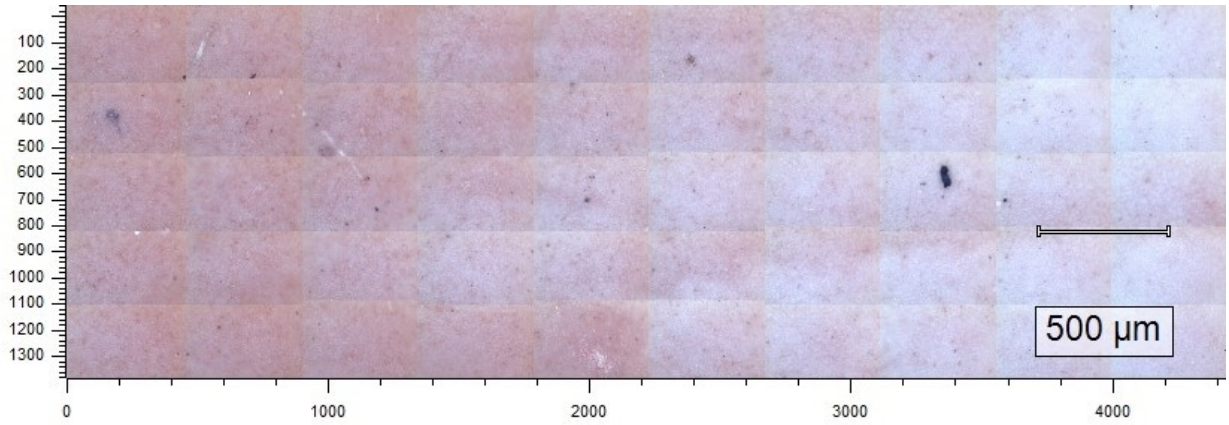


Figure 1. Optical image of the red lead sample.

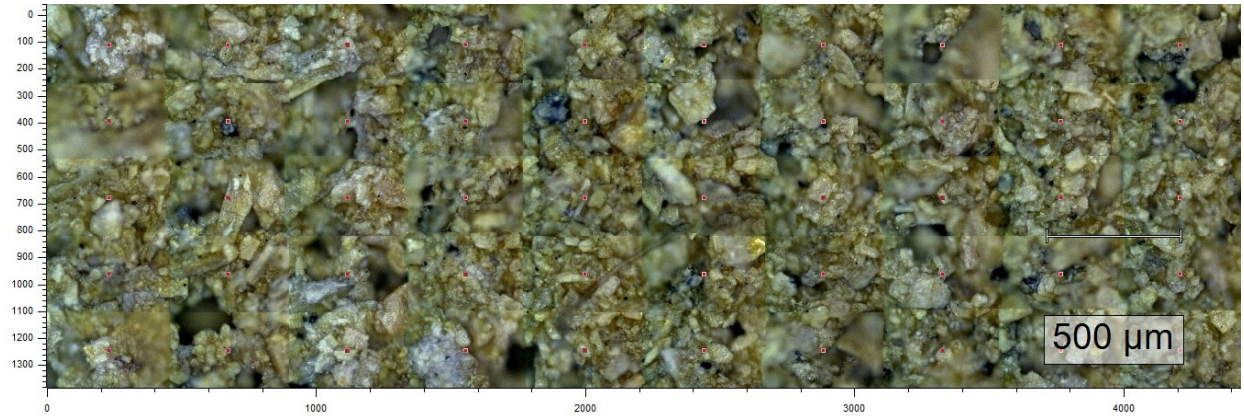


Figure 2. Optical image of the orpiment sample.

The number of features was reduced using the following feature selection methods for part 2: first, each sample's spectra were averaged to calculate the average width at half maximum of each peak using Scipy[1]. Secondly, the wavenumbers within the average width at half maximum of each peak were extracted to calculate Analysis of Variance (ANOVA) F-values. The feature, or wavenumber, from each peak that had the highest ANOVA value was extracted, reducing the dataset to 294 features for red lead and 48 for orpiment. After calculating the variance and dropping the features with variances less than 0.0018 for red lead and 0.00048 for orpiment, correlation and feature importance were calculated on the remaining 15 and 12 features for red lead and orpiment, respectively. The variance thresholds were selected to extract the features with relatively large variances. Only features that had high importance using the Extra Trees Classifier and that had low correlation coefficients (< 0.7) were kept, where the cut-off was chosen to reduce the number of features below six. The ANOVA, variance, correlation, and feature importance plots that were used to select the red lead and orpiment features can be found in Figure S3 and S4.

Using the selected features, 14 classifiers using stratified 10-fold cross-validation and a 70/30 train/test split were assessed by their accuracy and AUC scores (PyCarat package 3.3.2), which are common metrics used to evaluate the performance of a binary classification model. Accuracy represents the proportion of correctly labelled observations, and AUC explains how well the model is able to distinguish between the unaged and aged samples at all classification thresholds[2]. The hyperparameters of the top five classifiers based on AUC were tuned for AUC using RandomGridSearch with 100 iterations. If the tuned model had higher AUC than the original model, then it was chosen for comparison; if not, the original model was kept. The classifier with the highest AUC was chosen as the final model to distinguish between the unaged and aged spectra. The tuned and original Extra Trees Classifiers were chosen for samples A and B with an average AUC of 1.0000 and 1.0000 and an accuracy of 0.9971 and 0.9885, respectively.

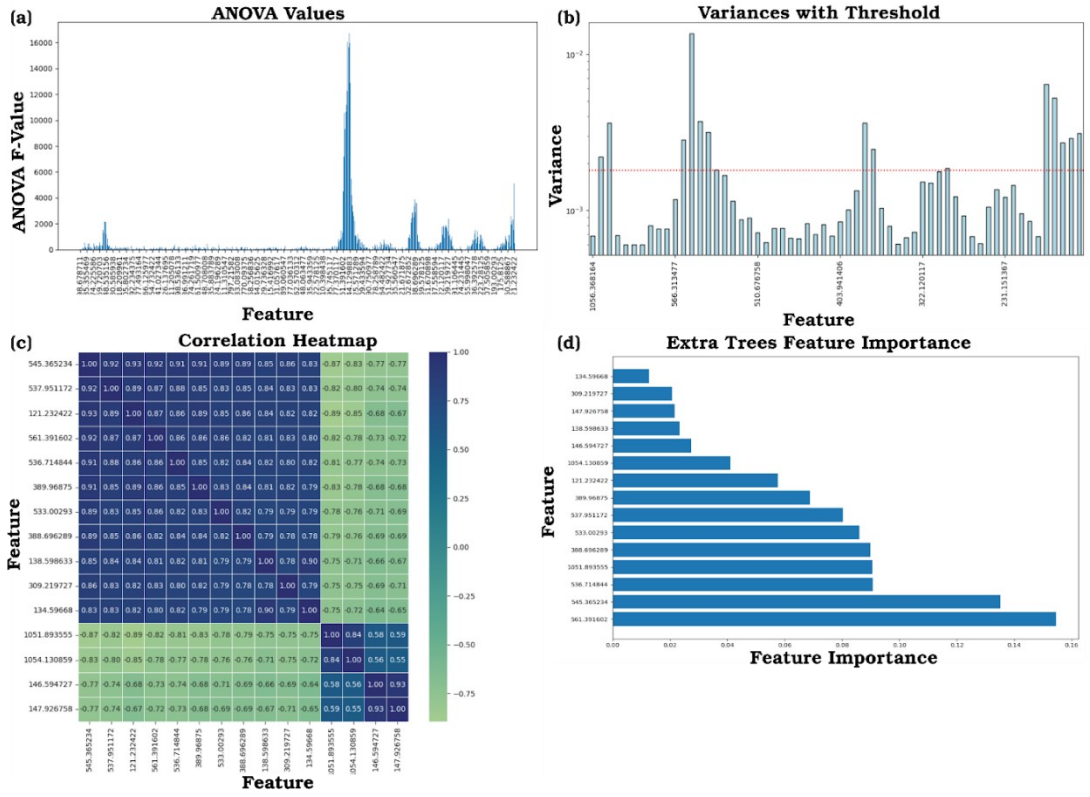


Figure S3. Summary of the feature selection techniques used to decrease the red lead data to three features to compare the unaged and aged samples. (a) The ANOVA F-values of the features found within the peak ranged at half maximum. (b) The variances of the features with the highest ANOVA value in each peak. The red line denotes the 0.0018 cut-off. (c) The correlation heatmap of the 15 features above the variance cut-off. (d) The Extra Trees feature importance of the 15 features.

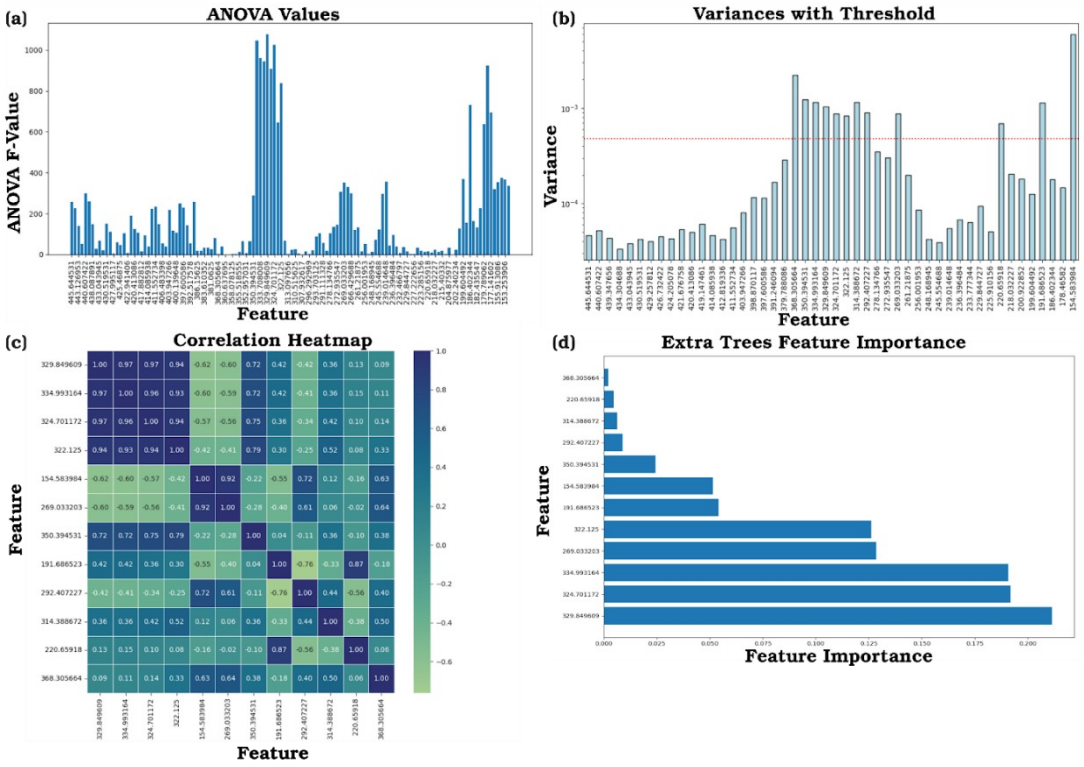


Figure S5. Summary of the feature selection techniques to decrease the orpiment data to five features to compare the unaged and aged samples. (a) The ANOVA F-values of the features found within the peak ranges at half maximum. (b) The variances of the features with the highest ANOVA value in each peak. The red line denotes the 0.00048 cut-off. (c) The correlation heatmap of the 12 features above the variance cut-off. (d) The Extra Trees feature importance of the 12 features.

The ANOVA and variance plots that were used to select the part 3 red lead and orpiment features can be

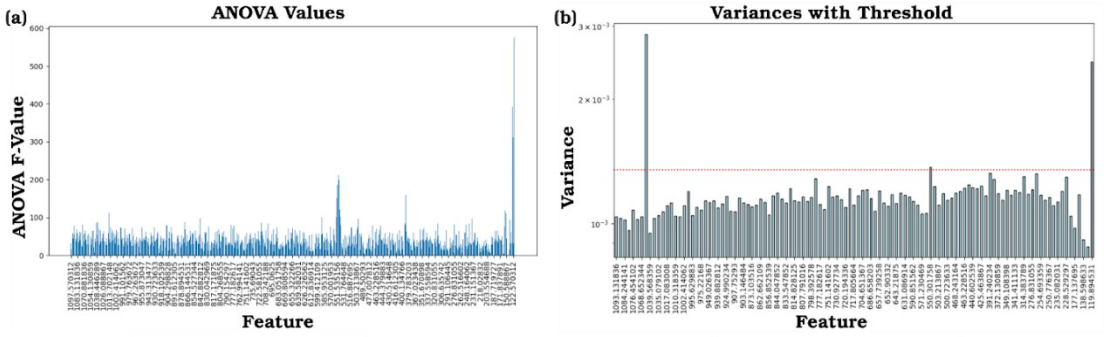


Figure S4. Summary of the feature selection techniques used to decrease the red lead data to three features to compare the defocusing steps. (a) The ANOVA F-values of the features found within the peak ranges at half maximum. (b) The variances of the features with the highest ANOVA value in each peak. The red line denotes the 0.00135 cut-off.

found in Figure S5 and S6. After calculating the variance for feature selection, the 3 features from red lead with variances below 0.00135 and the 21 features from orpiment with variances below 0.00044 were selected to create the model using stratified 10-fold cross-validation and a 70/30 train/test split. A smaller feature set was selected using orpiment, but the 21 features were chosen to prevent poor model

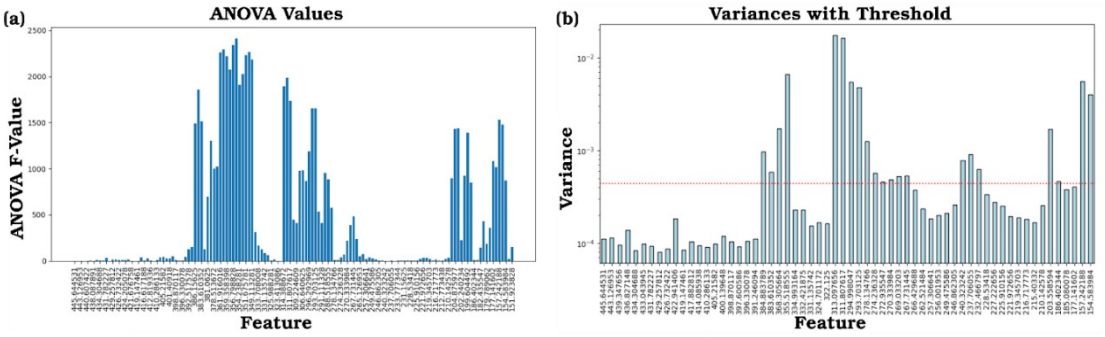


Figure S6. Summary of the feature selection techniques used to decrease the orpiment data to 21 features to compare the defocusing steps. (a) The ANOVA F-values of the features found within the peak ranges at half maximum. (b) The variances of the features with the highest ANOVA value in each peak. The red line denotes the 0.00044 cut-off.

performance. Less features needed to be used with red lead to eliminate features corresponding to fluorescent noise. The top five of 14 classifiers were tuned for accuracy and compared using accuracy and precision scores, which are common metrics used to evaluate the performance of a multiclassification model. Precision represents the fraction of positive results that were correctly predicted[3]. Tuned Linear Regression was chosen for red lead and tuned Light Gradient Boosting Machine classifier was chosen for orpiment since they outperformed the other models with an average accuracy of 0.3950 and 0.5440 and precision of 0.3786 and 0.5377, respectively.

To calculate the relative percent degradation for part 2, either Eq. (S1) or Eq. (S2) was used. If the intensity of the feature decreased in the unaged sample, then Eq. (S1) was used, while if the intensity of the feature increased in the unaged sample, then Eq. (S2) was used. By combining the percent of relative degradation for each feature, an overall relative percent degradation was calculated for each spectrum using Eq. (S3). The overall relative percent degradations were scaled between 0 and 1 using max-min normalization.

For part 3 unsupervised learning, Euclidean and ward were selected with 12 clusters for read lead and 9 clusters for orpiment. The features in the peak used to normalize the data were removed, which included feature 119.89 for red lead and feature 355.52 for orpiment. If the intensity of the feature decreased as the class number increased, then Eq. (S1) was used to calculate relative percent degradation, while if the intensity of the feature increased as the class number increased, then Eq. (S2) was used to calculate relative percent degradation. The overall relative percent degradation was calculated as discussed previously. To compare the experimental and machine learning heatmaps, the machine learning heatmaps were subtracted from the experimental heatmaps. The absolute value of the difference represents the similarity between the results.

Equation S1. The equation used to calculate the relative percent degradation of a feature that decreases in the unaged sample in part 2 or as the class number increases in part 3. Here, I_x is the intensity of feature x, $\min(\underline{I}_c)$ is the minimum average cluster intensity, and $\max(\underline{I}_c)$ is the maximum average cluster intensity.

$$\%relative\ degradation = \frac{I_x - \min\left(I_c\right)}{max\left(I_c\right) - \min\left(I_c\right)} \times 100$$

Equation S2. The equation used to calculate the relative percent degradation of a feature that increases in the unaged sample in part 2 or as the class number increases in part 3.

$$\%relative\ degradation = \frac{\max\left(I_c\right) - I_x}{max\left(I_c\right) - \min\left(I_c\right)} \times 100$$

Equation S3. The equation used to calculate the overall relative percent degradation for each spectrum in the sample matrix. Here, x is the total number of features and d_i is the degradation of value i:

$$overall\ relative\ \%\ degradation = \sum_{i=1}^{x} (d_i)$$

Results and discussion

Red lead

Red lead exhibits the most characteristic Raman bands at 550, 392, and 122 cm^{-1} (Figure S7)[4–9]. The low wavenumber band, around 122 cm^{-1}, corresponds to bending vibration of the Pb-O bond angle, whereas 392 and 550 cm^{-1} are both related to the symmetric stretching mode of the Pb(IV)-O bond. The effect of the aging process leads to the formation of white areas, easily recognizable in the whitish parts of the optical image of the sample in Figure S1. The compound is identified as hydrocerussite ($Pb_3(CO_3)_2(OH)_2$), or 'lead white', thanks to the 1050 cm^{-1} band (Figure S7)[10–13], stemming from the symmetric stretching vibration of the carbonate group. The chosen characteristic bands for mapping the distribution of these compounds are 122 cm^{-1} and 1050 cm^{-1} for red lead and lead white, respectively.

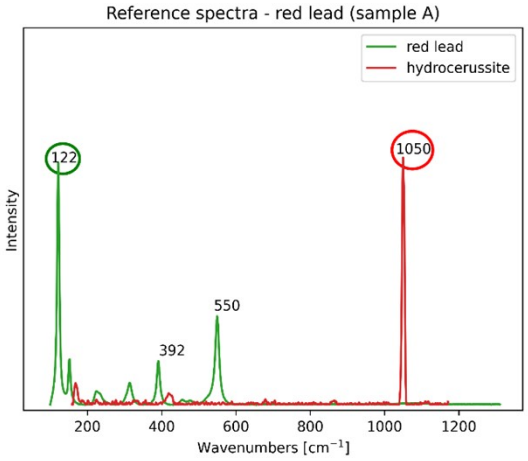


Figure S7. Reference spectra of the red lead sample.

An interesting aspect can be observed by looking at three random micro-SORS spectra of red lead in Figure S8 (full micro-SORS series is shown in Figure S9): the unaged spectra show an excellent signal-to-noise ratio, whereas the aged red lead sample is dramatically more fluorescent, indicating that the aging process had a huge impact on the sample. In fact, the aging process induced the formation of fluorescing species, most likely stemming from the aged linseed oil.

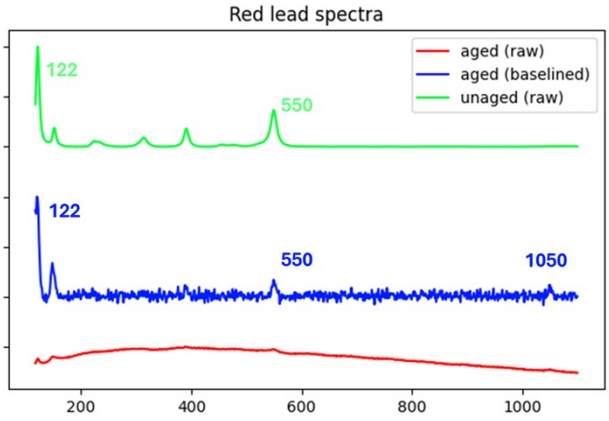


Figure S8. Example spectra of the aged sample, both raw (red one) and baselined (blue one), and of the unaged sample after baseline (green one).

Therefore, in order to be able to analyse these spectra, baseline correction was necessary (performed with

Python, exploiting the SNIP algorithm[14]), and it unavoidably retained noise in the final spectra, brought

about by fluorescence and photon shot noise, as can be seen from Figure S8. Moreover, the amount of

fluorescence varied from spectrum to spectrum, rendering the noise interference very heterogeneous

among the different acquisitions, with some spectra presenting extremely low signal-to-noise ratio. This

became an issue for the normalization step, since for the noisiest spectra the average value was not centred around zero and thus the algorithm would have picked up features as important even though they were just exaggerated by the noise (see Figure S10 for clarity). Therefore, we decided to subtract from

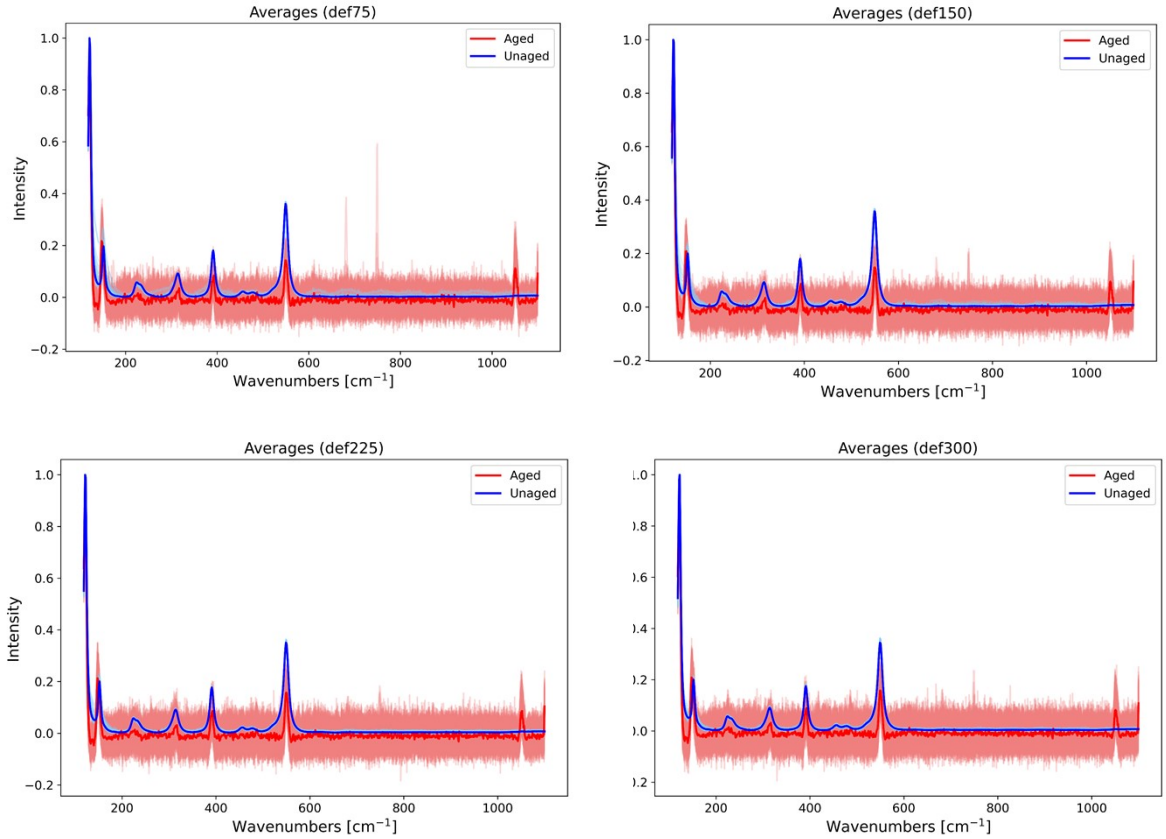


Figure S10. Full sets of spectra of the a) 75 µm, b) 150 µm, c) 225 µm and d) 300 µm defocusing steps of red lead.

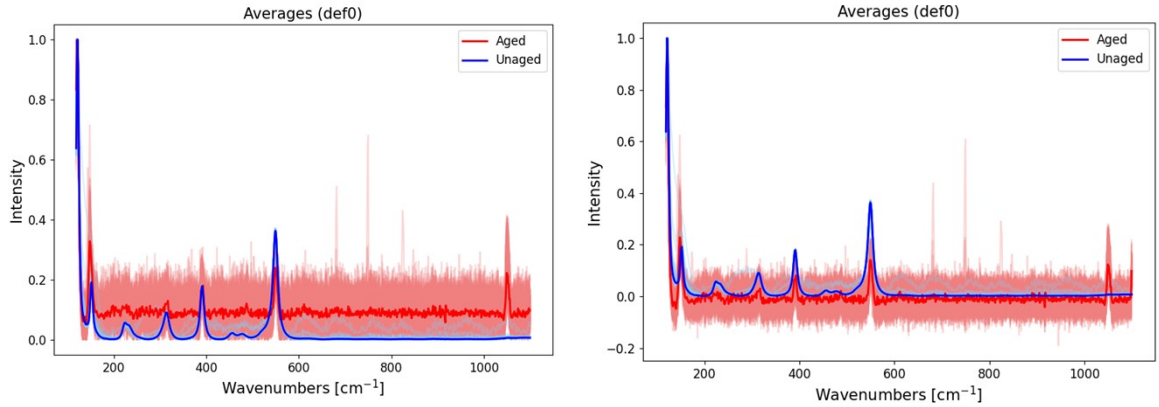


Figure S9. Full set of a) non-subtracted and b) subtracted spectra of the surface level of red lead.

each spectrum its own average intensity, in order to centre the baseline around zero and have comparable features with the unaged dataset. Even though this is not ideal, it was the most practical solution to be able to compare the aged and unaged datasets.

Red Lead - Part 1. Micro-SORS measurements and experimental maps

If on one hand the raw spectra of the aged red lead can be challenging to analyse at first due to the strong fluorescence background, the degradation process produces only one new compound, lead white. In Figure S11 we report the maps acquired on the unaged and aged samples at different defocusing steps. As mentioned above, the ratio between the two selected Raman bands of lead white and red lead is plotted.

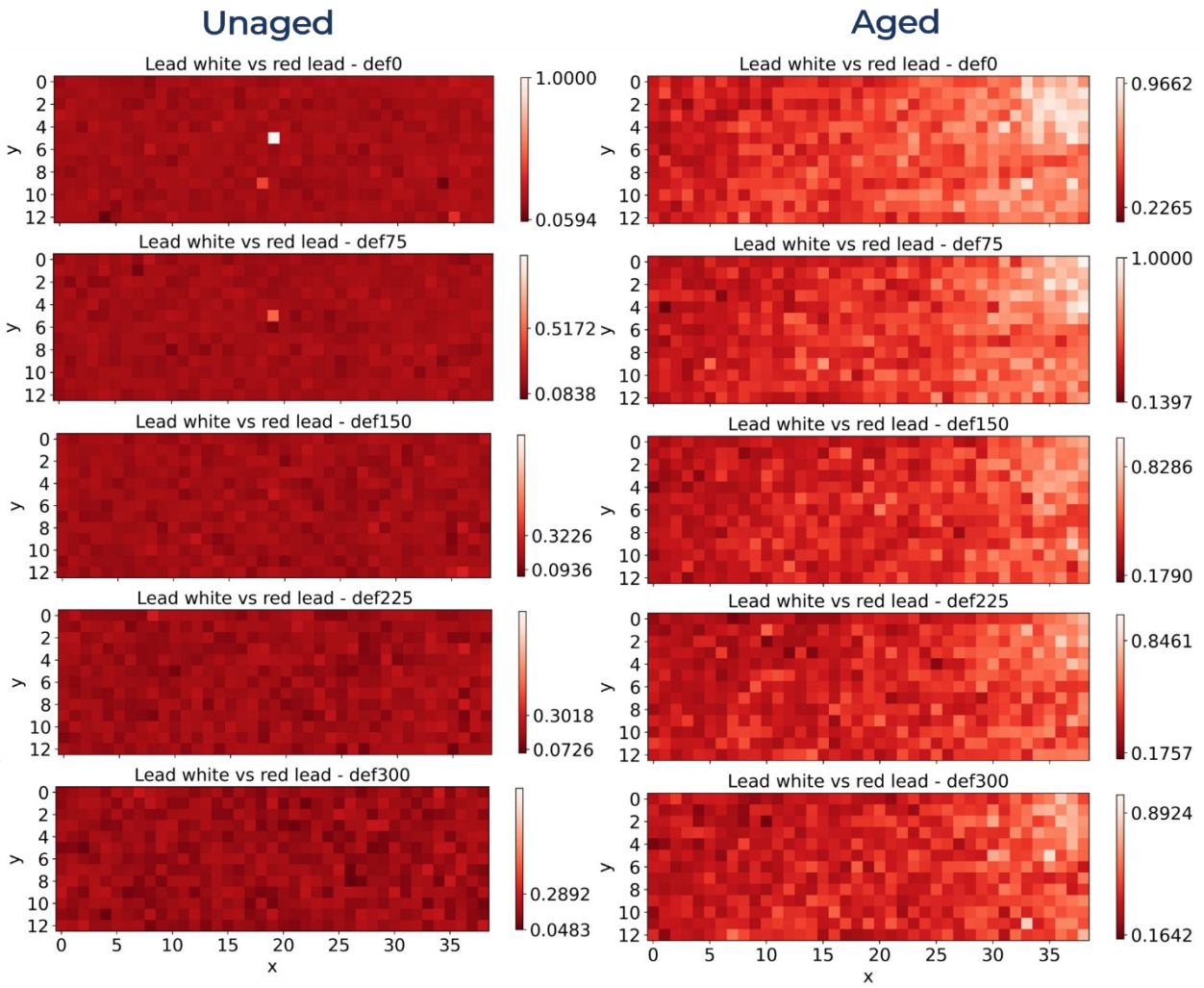


Figure S11. Raman maps of the unaged (left panels) and aged (right panels) red lead. The values of the colorbars indicate the maximum and minimum of each individual level with respect to the general scale.

By comparing Figure S1 and the right panels in Figure S11, it is straightforward to notice the similar pattern between the whitish portions in the optical image and the white pixels in the reconstructed map. It is worth noting that the unaged sample shows a substantial homogeneity in the distribution of the pigment, both on the plane and in depth. The intensity ratio values are very low, demonstrating a scarce or negligible amount of hydrocerussite, and these values decrease as the defocusing steps increase, confirming the presence of a weak natural degradation process. On the other hand, the aged maps show an

appreciable amount of lead white, demonstrating the effect of the aging process. In particular, the right part of this sample is the most degraded, indicating a possible difference in thickness or in exposure to the aging radiation that accelerated the degradation process on that side. Intensity ratios decrease as the defocusing increases, as expected: moreover, the evolution of the pattern of degradation is appreciable also below the surface. The very bright pixel on the surface of the unaged sample does not correspond to an actual band, it is an artefact created by the extremely poor signal-to-noise ratio presented in that specific spectrum (possibly occurred by the signal collection from an impurity).

Red Lead - Part 2. Unaged versus Aged Red Lead (Surface)

Machine Learning

The three features selected using machine learning to compare the unaged and aged spectra are displayed in Figure S12. Using the feature selection methods, feature 1052 is automatically selected and present in the only visible peak that represents lead white, which shows that the feature selection method is able to automatically capture the degradation peak. Additionally, the SHAP plot in Figure S13a shows that the lead white feature has the second greatest impact on the model. The other two features are on the edge of peaks found in both the intact and degraded samples, and they appear to be capturing peak width (feature 561) and peak position (feature 147) differences among the spectra. These differences appear to be present due to fluorescent noise and the samples being created separately. Using the three features, the Extra Trees model is able to correctly predict the 10 spectra from the unaged and aged samples in the holdout set as shown in Figure S13b. Therefore, feature selection with machine learning can be used to identify differences and degradation peaks among unaged and aged red lead spectra, even when fluorescent noise is present.

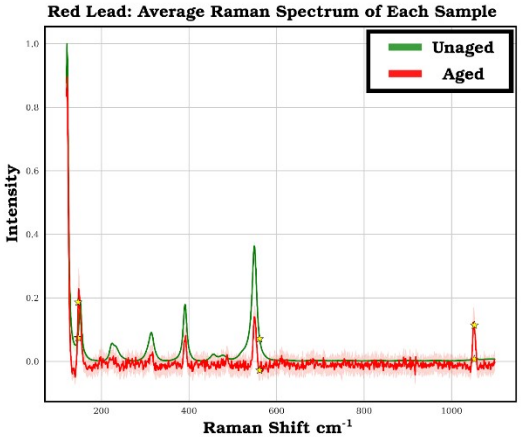


Figure S12. The average red lead unaged and aged spectra with the three selected features denoted as yellow stars.

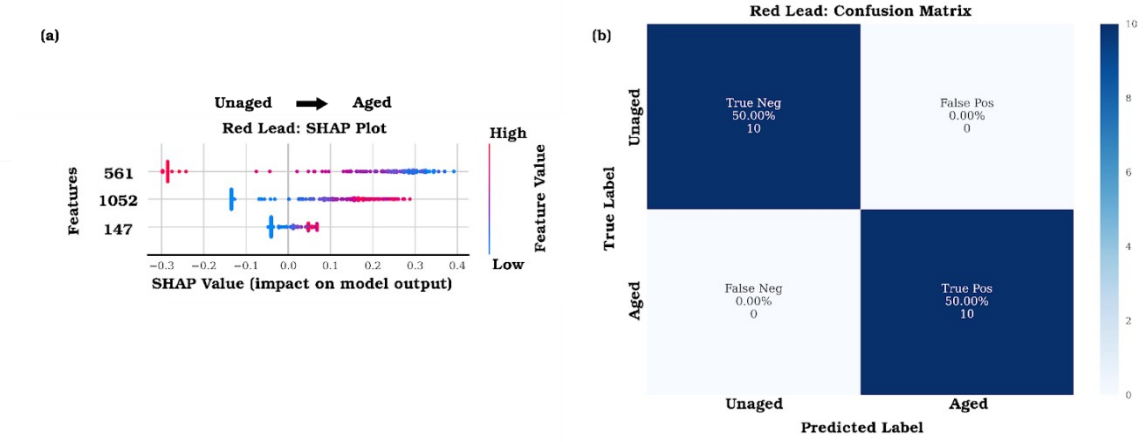


Figure S13. Summary of the red lead unaged vs aged Extra Trees model results. (a) SHAP beeswarm plot showing the selected three features. A positive SHAP value explains whether features with intensities that increase (magenta) or decrease (blue) are more aged. (b) The confusion matrix of the ten unseen unaged and ten unseen aged spectra.

Unsupervised Learning

Using the selected features, unsupervised learning clusters the data based on their intensities to quantitatively calculate the relative percent degradation across the surface of each sample, and it solves a problem when creating the relative percent degradation heatmaps. By clustering the spectra to obtain average minimum and maximum intensities for each feature in each cluster, the heatmaps are normalized and created using Equations S1-S2. The average intensities for each sample cannot be used to represent the degradation since each sample is heterogeneous and represents multiple intact and degraded pigments. Additionally, the minimum and maximum intensity values for each feature cannot be used since they may be high due to noise or low due to a peak shift. By using the average minimum and maximum intensities from the clusters that are grouped based on their intensities rather than on their defocusing step, the heatmaps are made with greater confidence. Figure S14a shows the average spectra for each cluster. Cluster 1 contains the majority of the unaged spectra as shown in Figure S14b, with only one aged spectrum being placed in cluster 1 and three unaged spectra being placed in other clusters.

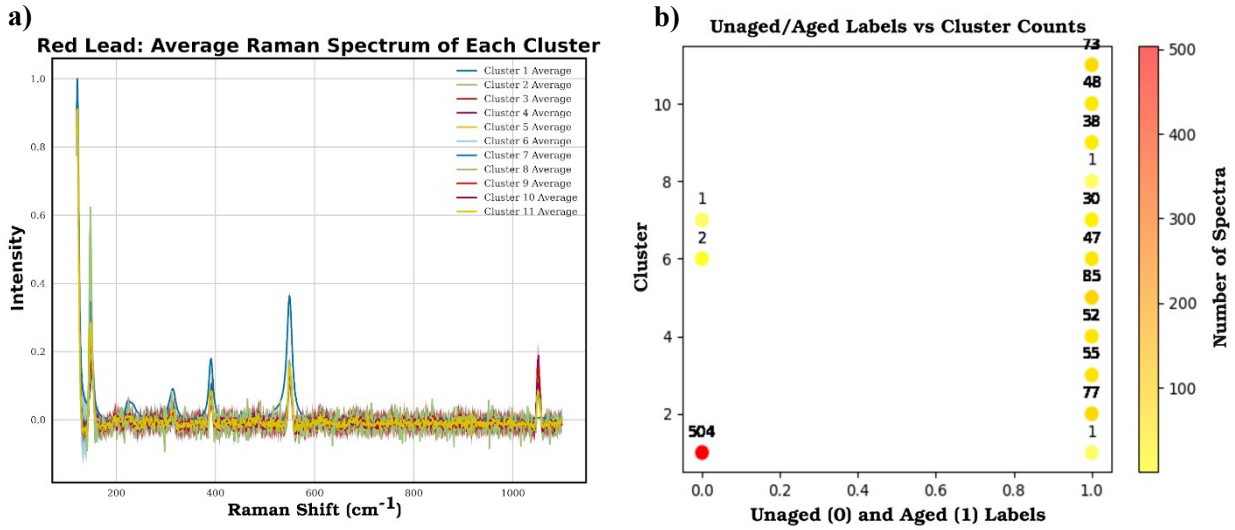


Figure S14. a) The average spectra of the eleven clusters using the red lead unaged and aged samples with the three selected features; b) The distribution of the unaged (0) and aged (1) red lead samples into eleven clusters. Orange represents a high number of spectra in a cluster, and yellow represents a low number of spectra in a cluster.

Calculating Degradation

Three unaged spectra being placed in aged clusters corresponds to what is seen in the relative percent degradation heatmap calculated using feature 1052 in Figure S15, which represents lead white. The quantitative heatmaps created using features 561 and 147 are displayed in Figure S16, but they are not used as a main reference for relative percent degradation since they do not correspond to an explicit red or lead white peak. Still, these features are required by the model to accurately predict the outcome. As observed in Part 1, the heatmap created using feature 1052 displays the heterogeneity in the aged sample and the uniformity in the unaged sample.

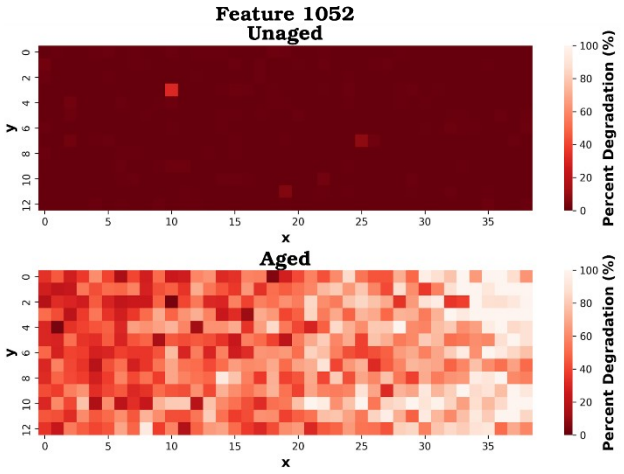


Figure S15. The percent degradation of the unaged and aged red lead samples using feature 1052. A more degraded area of the sample is represented by white, and a more intact area of the sample is represented by dark red.

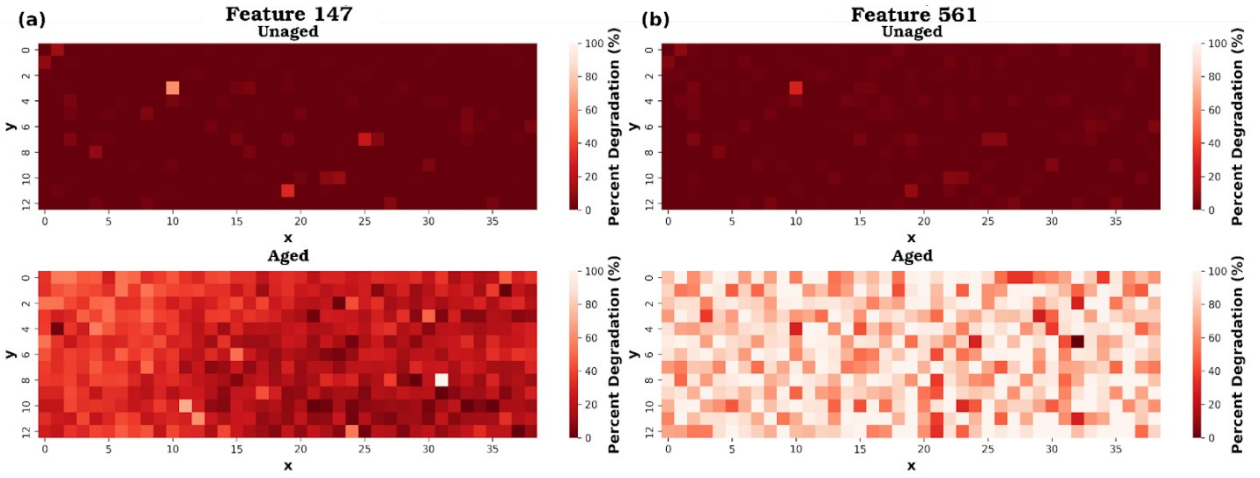


Figure S16. The percent degradation of the unaged and aged red lead samples using features 147 (a) and 561 (b). A more degraded area of the sample is represented by white, and a more intact area is represented by dark red.

As shown in Figure S15, there appears to be three cells in the unaged heatmap that are lighter than the others (3x10, 7x25, 11x19). These three light cells in the red lead unaged heatmap correspond to the three unaged spectra being placed in clusters 6 and 7 and not cluster 1 (Figure S14b). The spectra of the three cells are shown in Figure S17 along with the average spectra for clusters 1, 6, and 7, and they appear to be placed in clusters 6 and 7 due to the presence of noise giving the features higher intensities, as discussed in Part 1.

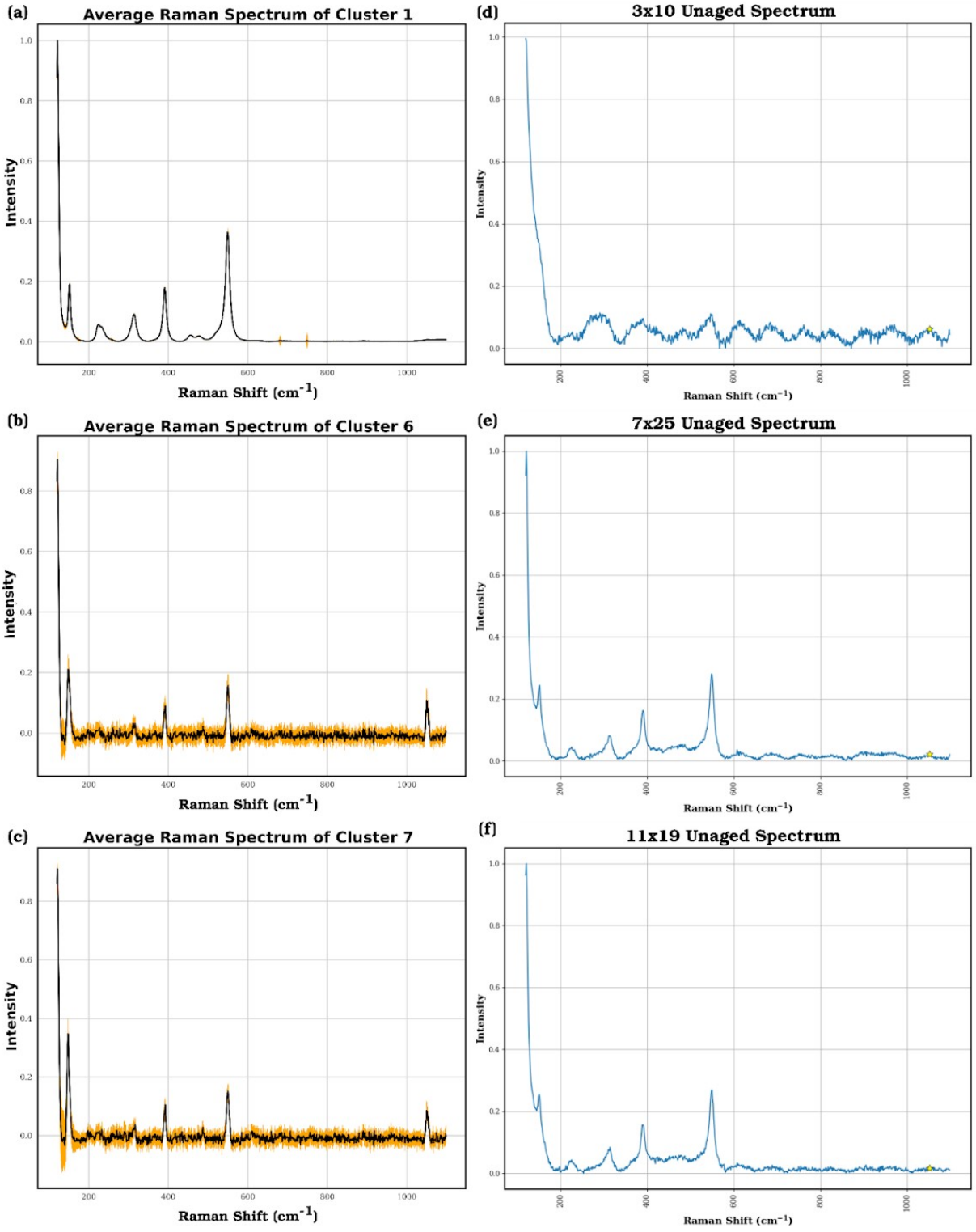


Figure S17. Summary of the clusters using the red lead unaged and aged samples with the three selected features. (a) Average spectrum of cluster 1, which represents the majority of the unaged spectra. (b) Average spectrum of cluster 6, which contains the unaged spectra in the 7x25 and 11x19 positions in the feature 1052 unaged heatmap. (c) Average spectrum of cluster 7, which contains the unaged spectrum in the 3x10 position in the feature 1052 unaged heatmap. (d-f) Unaged spectra that were not placed into cluster 1. The yellow star denotes the position of feature 1052.

The aged sample also appears to be more degraded in the upper right corner of the bottom heatmap in Figure S15, which corresponds with the sample thickness and radiation exposure discussed in the experimental section. To further ensure the method is correctly assigning cells with high and low degradation in the aged sample, three red (intact) cells (4x1, 2x10, and 0x18) and three white (degraded) cells (8x14, 10x37, and 2x36) from the aged heatmap were analysed. The three red (intact) cells (4x1, 2x10, and 0x18) and three white (degraded) cells (8x14, 10x37, and 2x36) from the aged heatmap were compared. As shown in Figure S18a-c, the Raman spectra for the white cells have low intensities (about 0.0) for the degradation peak except for cell 0x18. Cell 0x18 does have a slightly higher intensity (slightly below 0.2) due to feature 1051.89 not capturing the peak maximum. The red cells have high intensities (Figure S18d-f) for the degradation peak (around 0.2), which shows that most cells are correctly labelled as intact or degraded.

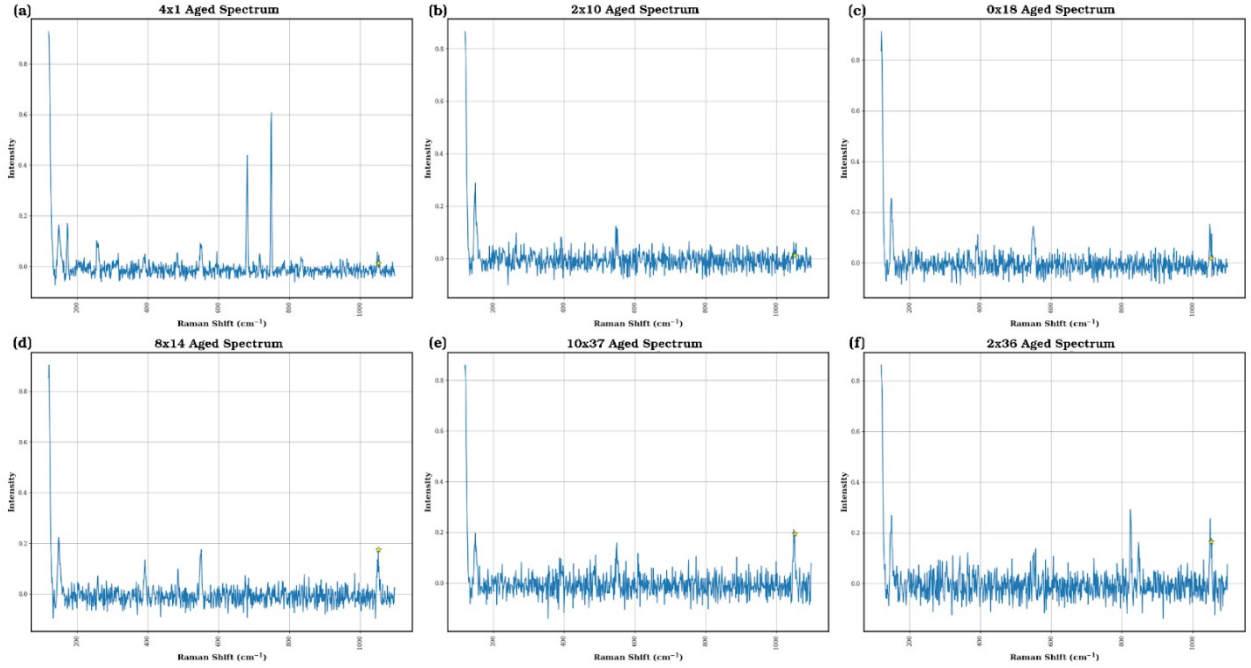


Figure S18. Raman spectra of the selected three red (intact) and three white (degraded) cells from the feature 1051.89 aged heatmap. (a-c) Raman spectra of the intact cells. (d-f) Raman spectra of the degraded cells.

The experimental heatmaps were also compared to the heatmaps in Figure S11 using feature 1052. As shown in Figure S19 the heatmaps appear very similar with only one cell (5x19) in the experimental unaged heatmap having a much higher intensity. Therefore, unsupervised learning using the selected features is successful in separating spectra based on their degradation without prior knowledge of unaged and aged labels, so that degradation can be mapped along the surface of each sample. However, it is important to note that noise and peak shifts can affect the feature selection and level of degradation calculated.

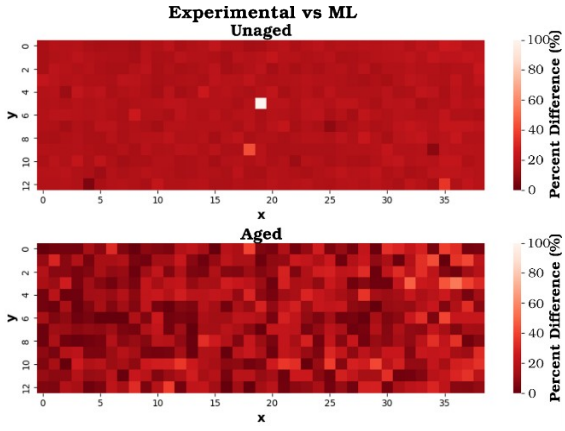


Figure S19. Comparison of the experimental and machine learning unaged and aged red lead heatmaps. Areas that agree between the two heatmaps are represented by dark red, and areas that differ are white.

Red Lead - Part 3. Aged Red Lead Subsurface Study

Machine Learning

Next we investigated whether feature selection with machine learning could distinguish between the five different aged red lead defocusing steps, which vary in degradation. The three features used to make the Linear Regression model with the five defocusing steps are displayed in Figure S20. The features in the SHAP plot in Figure S21 correspond to either the red or lead white, and most features contribute to the prediction of each class. The two most important features—120 and 550—correspond to red lead, and the least important feature 1050 corresponds to the lead white. The feature selection process is still able to identify a lead white feature as in Part 2 even when more spectra and defocusing steps are analysed. As shown in the SHAP plot, feature 1050 does not contribute to the class 1 predictions, most likely due to the large variation of its intensity in class 1, which overlaps with the other classes. Feature 1050 also contributes very little to class 4 predictions since it represents the degraded peak, which has too little intensity variance in classes 3 and 4. Even though classes 3 and 4 are similar, the intact pigments help distinguish class 4 better than class 3, which causes feature 1050 to contribute more to class 3. Feature 120 contributes the most to the class 4 predictions since it varies the most among the five classes. However, since feature 120 is present in the pea k used to normalize the data, it is removed when calculating degradation.

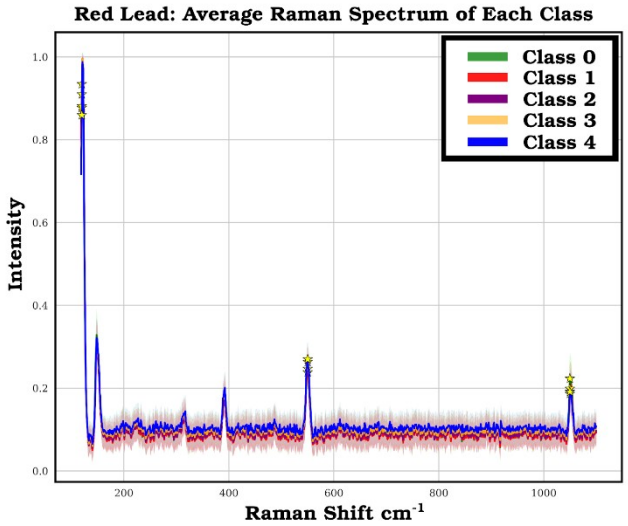


Figure S20. The average aged defocusing steps for the five defocusing steps (classes 0-4) with the 3 selected features denoted as yellow stars. Class 0 corresponds to the surface and class 4 corresponds to the deepest defocusing step.

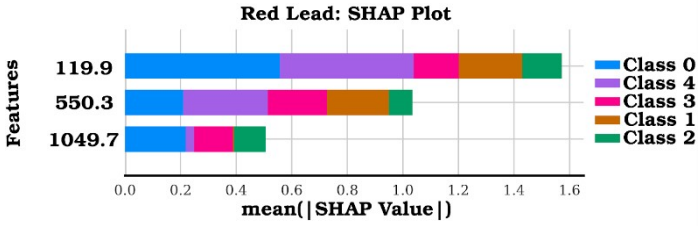


Figure S21. SHAP summary plot of the red lead defocusing steps using the Linear Regression model.

The model predicted the defocusing steps of the holdout dataset in Figure S22. Less than half of the spectra were labelled the same as their true label, with classes 1 and 2 having the most mislabelled spectra. Most of the class 2 spectra were

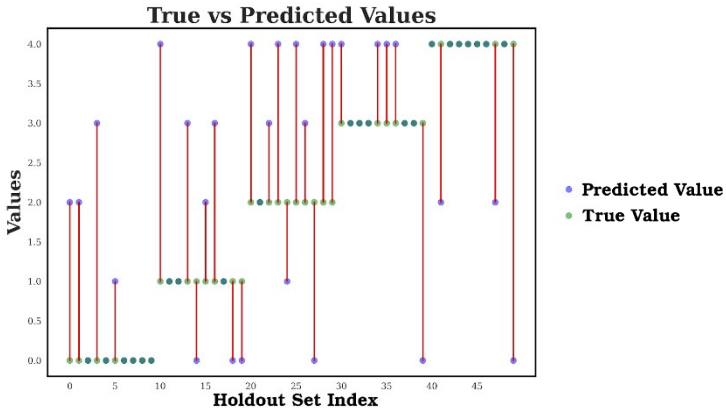


Figure S22. The red lead defocusing steps Linear Regression model outputs using the ten spectra from each class of the holdout set. Each red line represents how far away the true and predicted values are from each other.

predicted to be in class 3 or 4, which shows that classes 2-3 are similar to each other and the change in degradation may be starting to plateau in class 2. Furthermore, at least one spectrum in each class is predicted to be in class 0, with less being labelled as class 0 as the classes increase, which shows that there is still some degradation present in each class that may decrease with an increase in classes. Since the machine learning model predictions and the experimental Raman maps show great variability among each defocusing step, unsupervised learning was used to create a quantitative heatmap to gain a better understanding of how the red lead is degrading.

Unsupervised Learning

Agreeing with the Linear Regression model predictions in Figure S22, each of the 12 clusters contains spectra from each class (Figure S23), which we believe occurs due to the simplicity of the sample. Red lead only has one possible degradation product, which would cause the spectra in the different classes to have peaks that look very similar and only differ by their intensity values. The clustering confirms this as it separates the spectra based on the intensity values of the red and lead white features. The average spectrum for each of the 12 clusters is shown in Figure S24a. Cluster 1 contains spectra that have high intensity values for feature 1050 on average, while cluster 7 contains spectra that have low intensity values for feature 1050 on average, as shown in Figure S24b-c. Therefore, unsupervised learning using clustering can be combined with supervised machine learning to analyse the degradation of the defocusing steps of samples with fluorescent noise such as red lead.

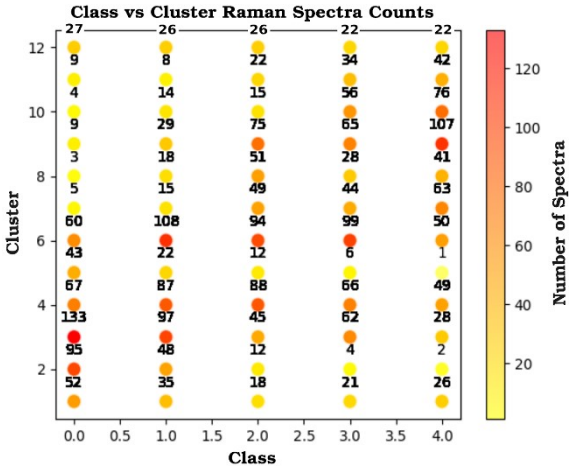


Figure S23. The distribution of the five red lead defocusing steps into twelve clusters. Orange represents a high number of spectra in a cluster, and yellow represents a low number of spectra in a cluster.

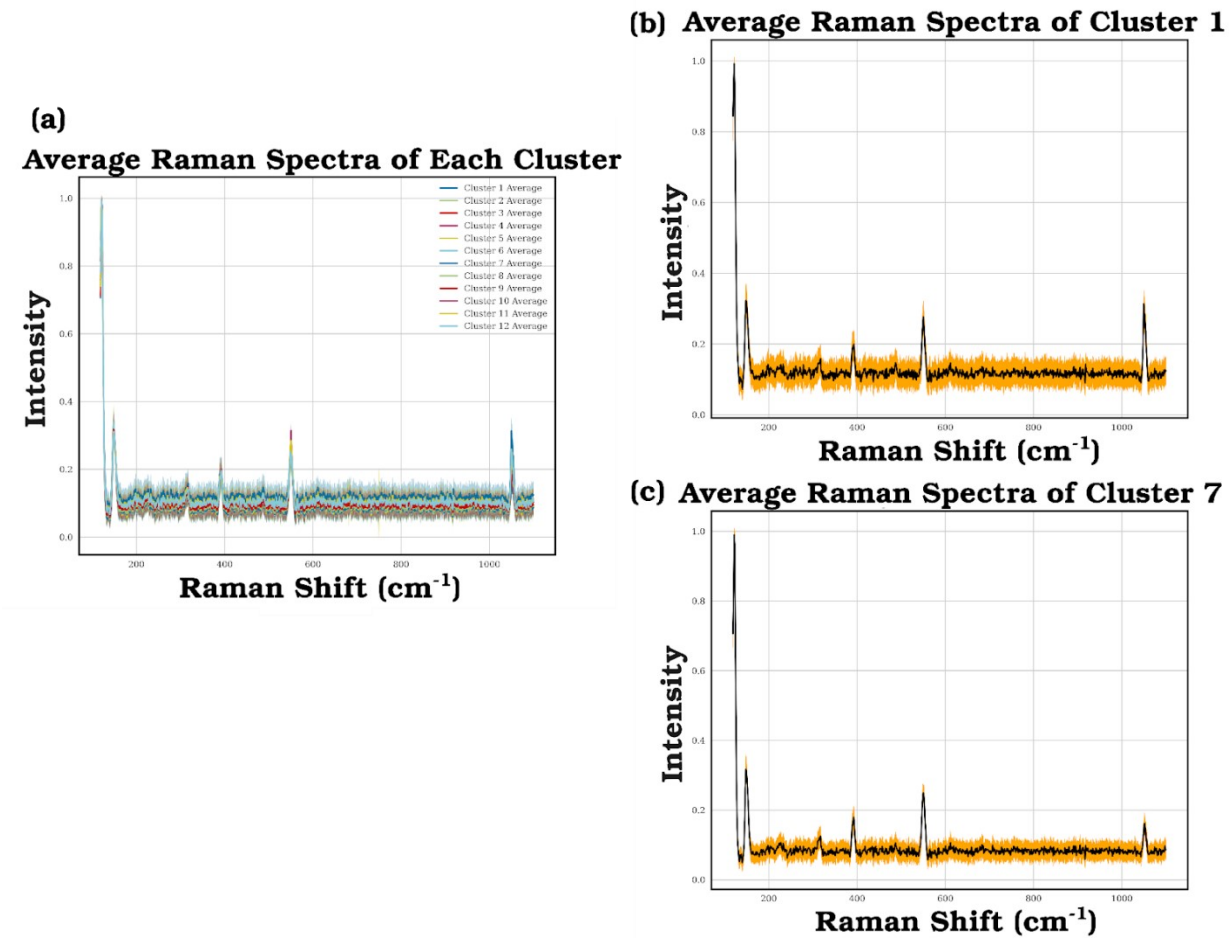


Figure S24. Summary of the unsupervised clusters using the red lead defocusing steps with the three selected figures. (a) Average spectra of each of the 12 clusters. (b) Cluster 1 showing the high intensity on average for feature 1050. (c) Cluster 7 showing the low intensity on average for feature 1050.

Calculating Degradation

The percent degradation was calculated using feature 1050 intensity values, which represents lead white, as shown in Figure S25. The heatmap created using feature 550 and the overall percent degradation are displayed in Figure S26. The degradation plot using feature 1050 captures the heterogeneity in each defocusing step, and it shows that the degradation decreases as the defocusing step increases, which confirms the Linear Regression model predictions. Similar to the aged sample in Part 2, the upper right corner of the surface of the sample (class 0) appears more degraded. The degradation is present in each class, which corresponds with the predictions of the Linear Regression model. As the defocusing steps get farther away from the light source, they should degrade less, which correlates with the degradation in the upper right corner decreasing with an increase in defocusing steps. Feature 1050 does have a slight increase in degradation in the last class compared to the first class.

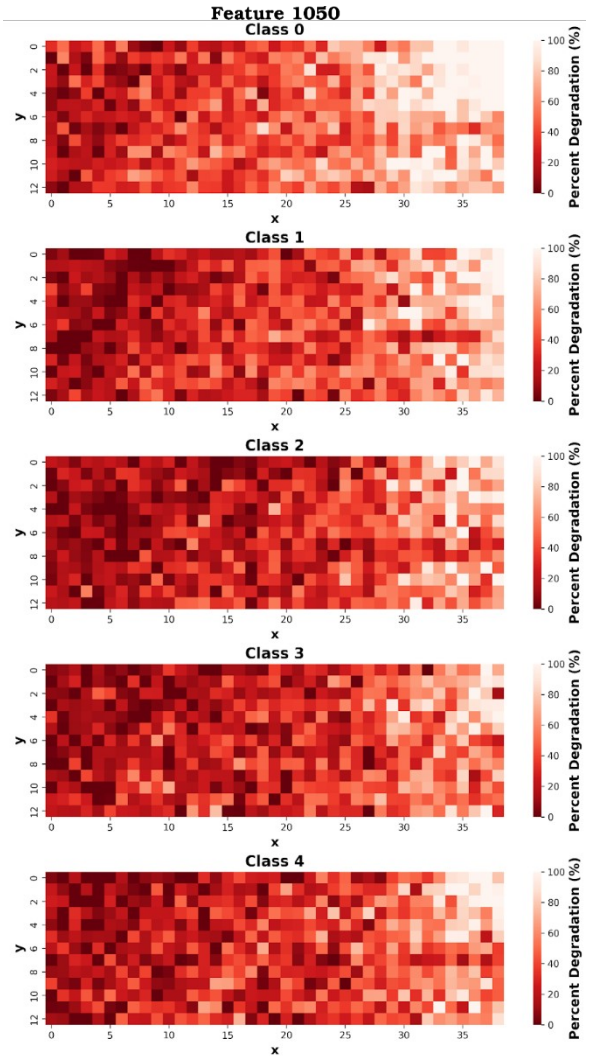


Figure S25. The percent degradation of the red lead defocusing steps using feature 1050. A more degraded area of the sample is represented by white, and a more intact area is represented by dark red.

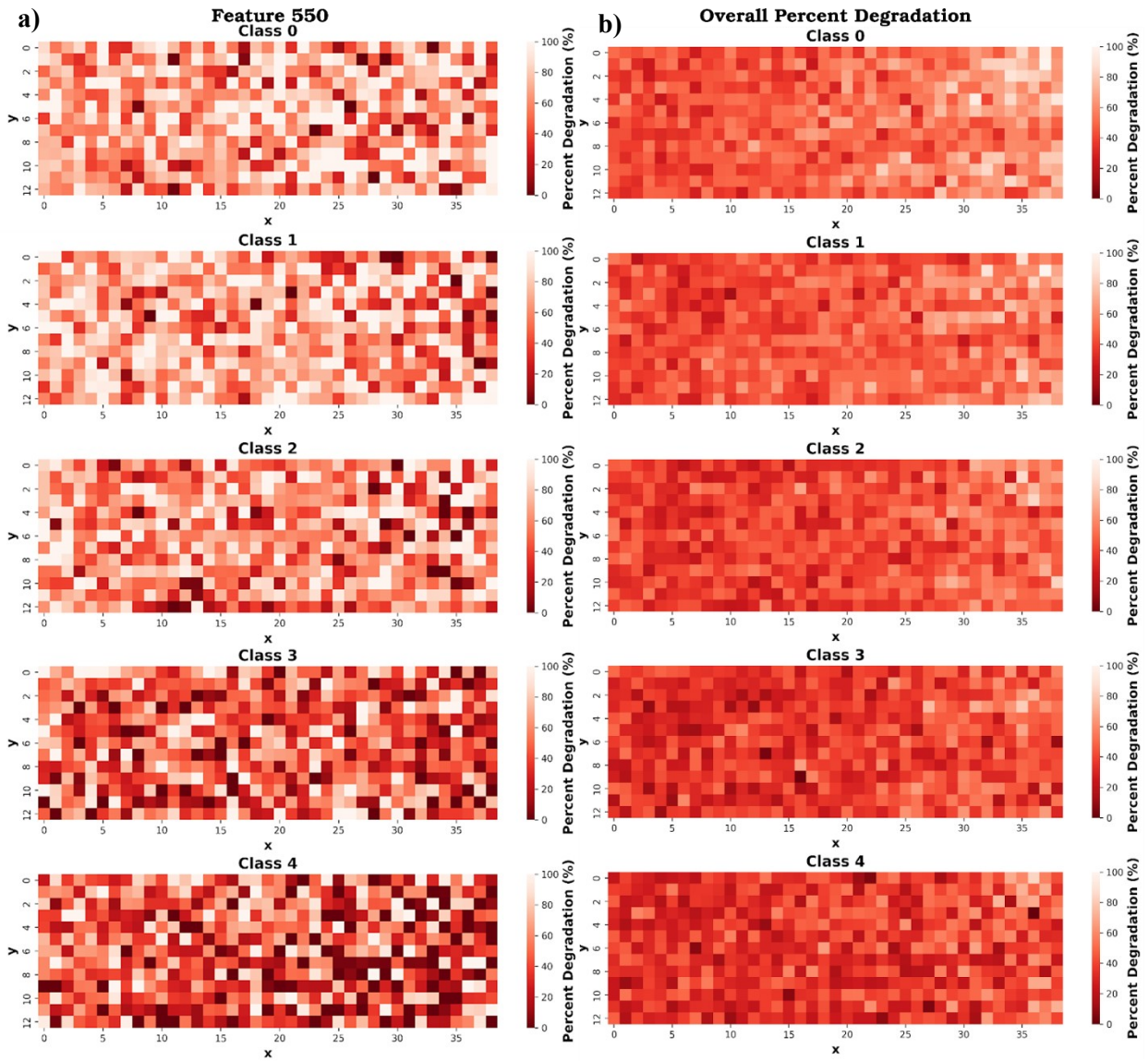


Figure S26. a) The percent degradation of the red lead defocusing steps using feature 550. A more degraded area of the sample is represented by white, and a more intact area is represented by dark red; b) The overall percent degradation of the red lead defocusing steps using features 550 and 1050. A more degraded area of the sample is represented by white, and a more intact area is represented by dark red.

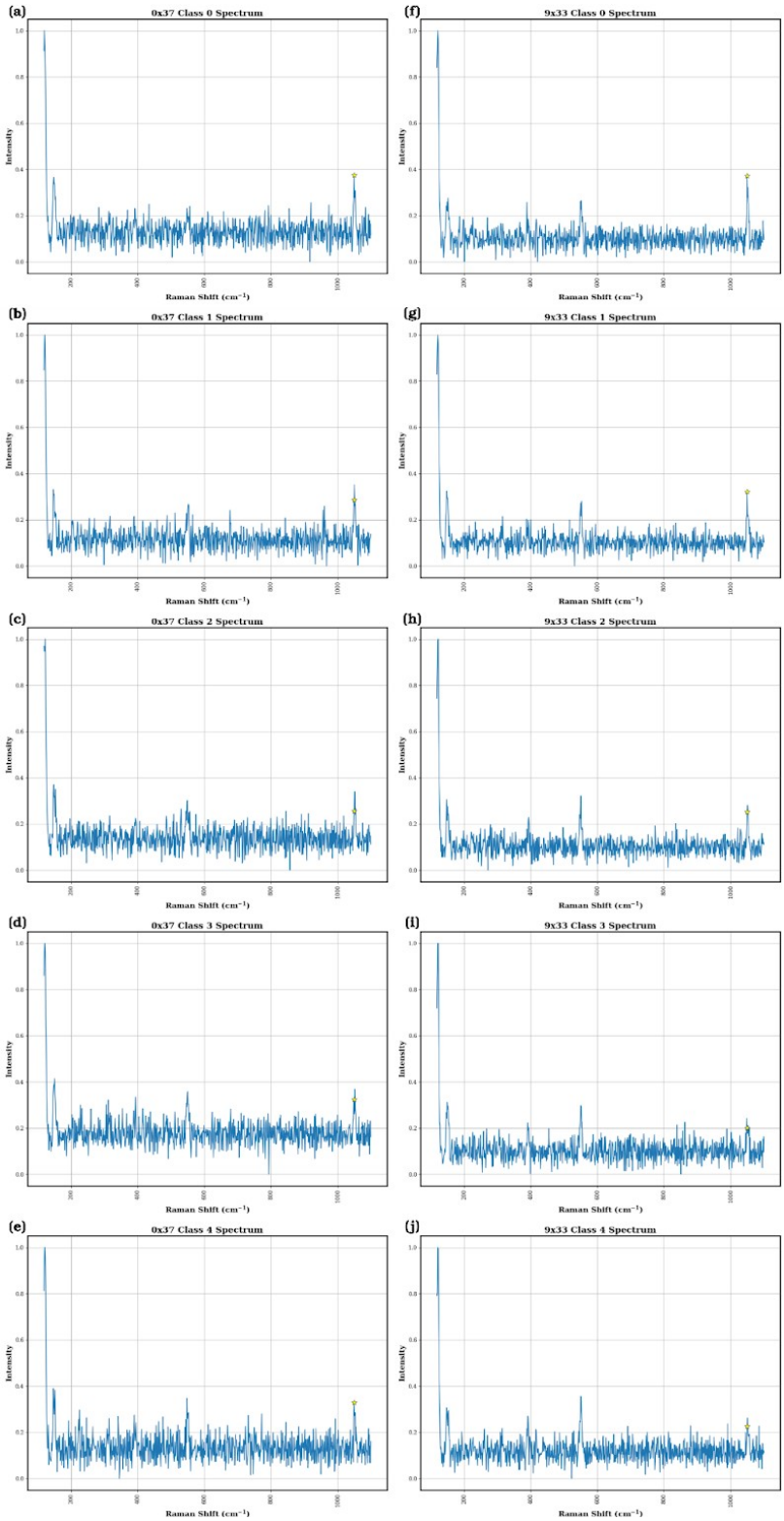


Figure S27. Raman spectra of two highly degraded (white) cells in the class 0 red lead defocusing step heatmap showing how the spectra change as the defocusing steps (classes) increase. (a-e) The Raman spectra in each defocusing step corresponding to the 0x37 cell in the heatmap. (f-j) The Raman spectra in each defocusing step corresponding to the 9x33 cell in the heatmap. The yellow star denotes the position of feature 1050.

To further investigate the degradation pattern, two very degraded (white) cells (0x37 and 9x33) in class 0 were selected. The Raman spectra of the cells (0x37 and 9x33) in each defocusing step are shown in Figure S27. Since these cells are highly degraded on the surface, they should decrease in degradation (and feature 1049.65 decrease in intensity) as the defocusing steps increase. Cell 9x33 shows this pattern as shown in Figure S28, but cell 0x37 shows classes 1 and 2 having lower intensities than classes 3 and 4. Due to noise and a slight shift in the peak maximum, feature 1049.65 is not able to represent the maximum intensity in the degradation peak in every defocusing step, which makes the higher defocusing steps appear more degraded. Even though the degradation may not steadily decrease as the defocusing steps increase for each cell, it is still important to note that the highly degraded (white) cells have almost double the intensity for the degradation peak compared to the very intact (red) cells. The degraded and intact cells are still being correctly labelled. Two intact (red) cells (7x6 and 1x0) in class 0 were also selected for further analysis. Since the 7x6 and 1x0 cells are intact, they should not change much as the defocusing steps increase. As shown in Figure S29, the intensity of feature 1049.65 is very close in all the classes, but some higher classes have higher intensities due to noise and a shift in the peak maximum.

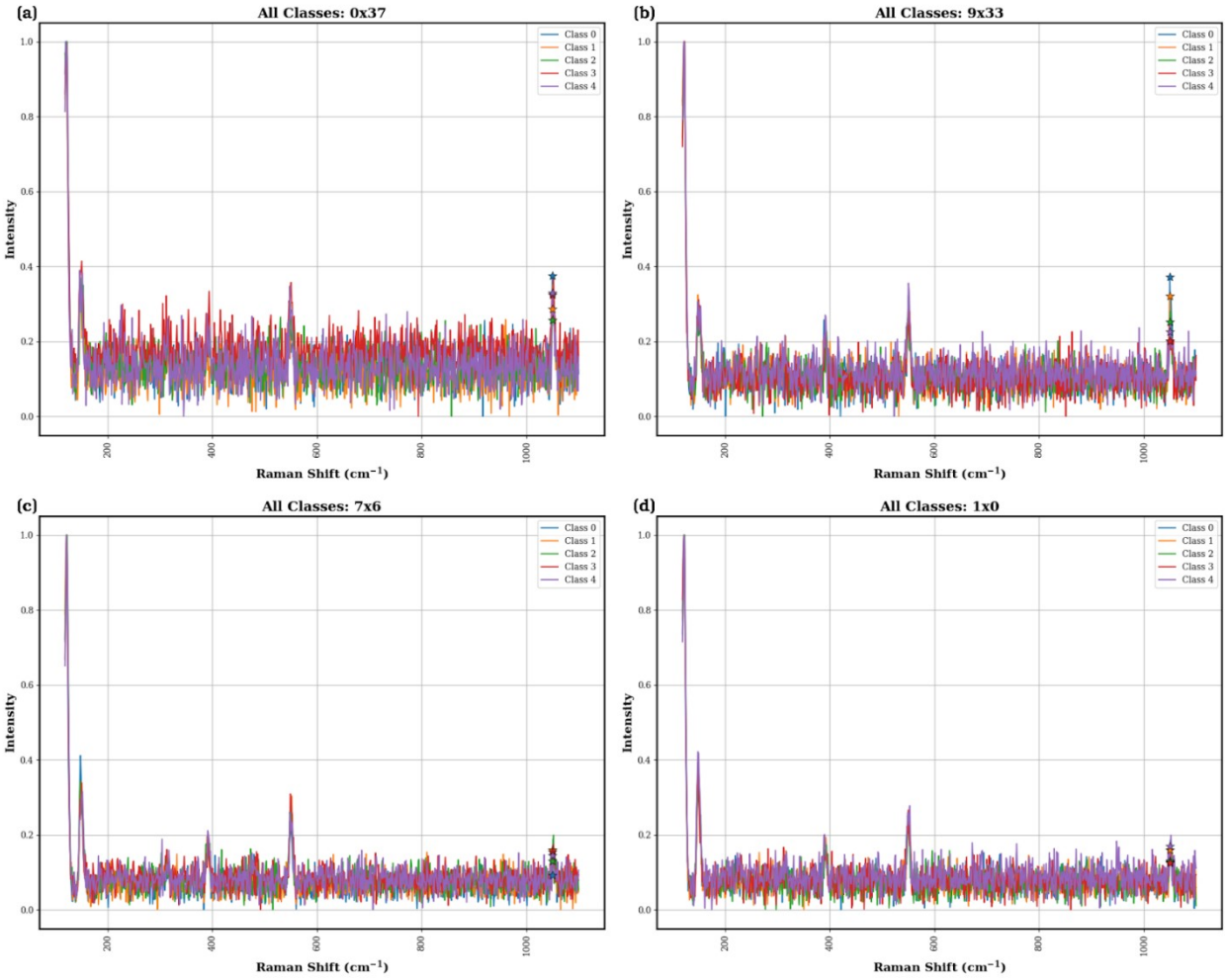


Figure S28. Summary of the selected cells in the red lead defocusing step heatmaps to show the intensity (and degradation) patterns. (a-b) The selected degraded (white) cells in the heatmaps. (c-d) The selected intact (red) cells in the heatmaps.

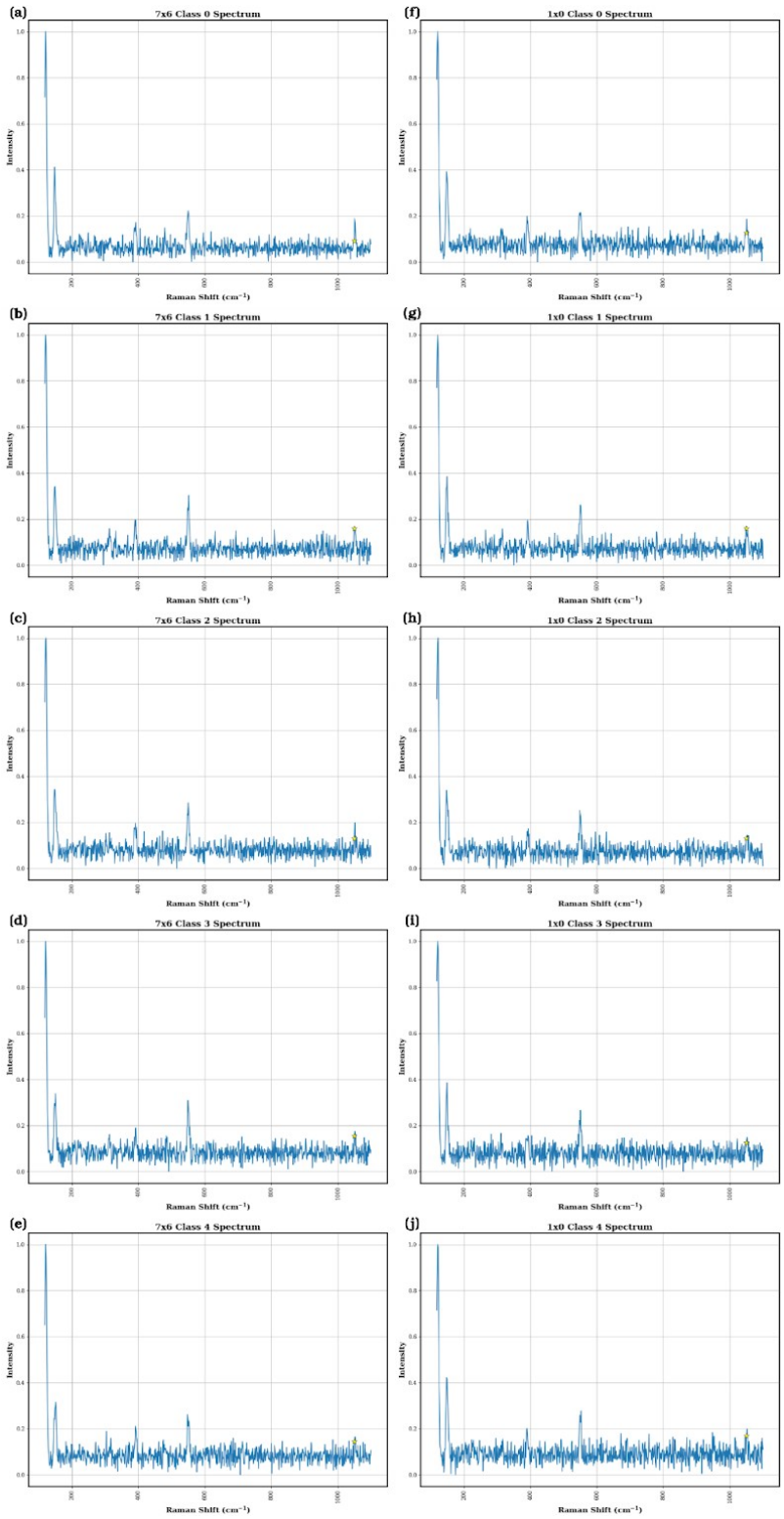


Figure S29. Raman spectra of two highly intact (red) cells in the class 0 red lead defocusing step heatmap showing how the spectra change as the defocusing steps (classes) increase. (a-e) The Raman spectra in each defocusing step corresponding to the 7x6 cell in the heatmap. (f-j) The Raman spectra in each defocusing step corresponding to the 1x0 cell in the heatmap. The yellow star denotes the position of feature 1050.

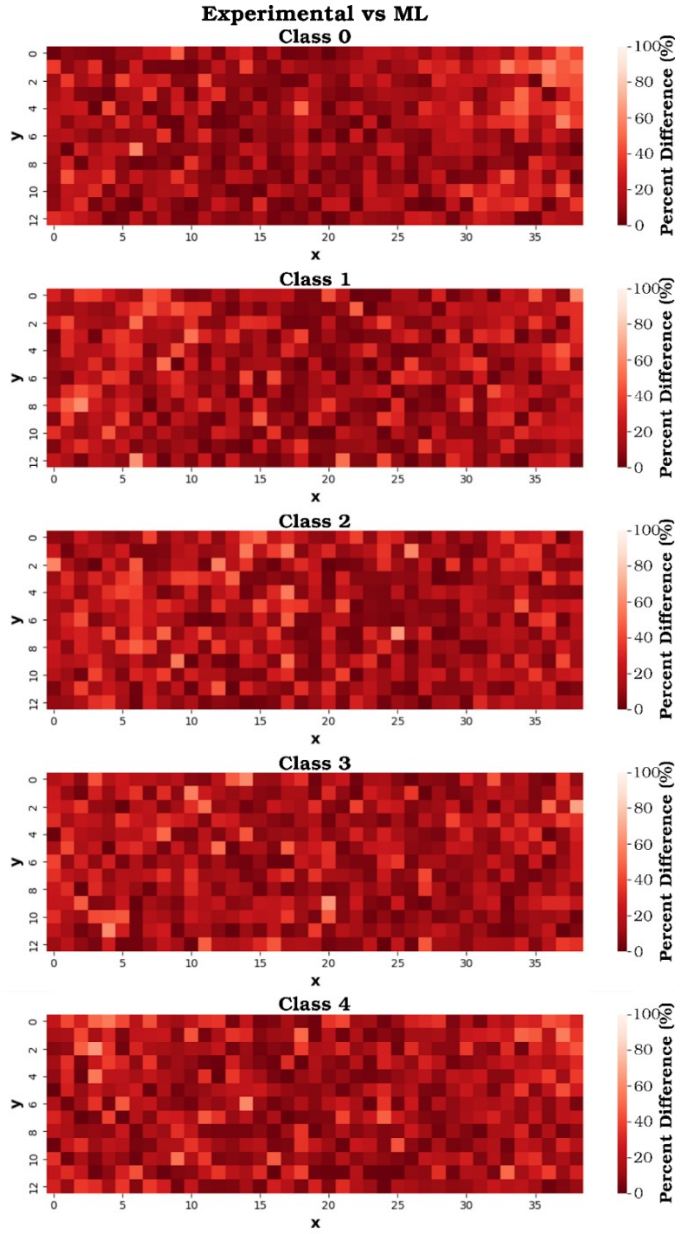


Figure S30. Comparison of the experimental and machine learning red lead defocusing steps heatmaps. Areas that agree between the two heatmaps are represented by dark red, and areas that differ are white.

The experimental heatmap and the heatmap created using feature 1050 in Figure S25 are comparable with small differences, as shown in Figure S30. Therefore, even when using defocusing steps and increasing the number of spectra, a feature corresponding to the degradation peak is still automatically selected, which can be used to map the degradation in each defocusing step and determine which areas of the sample are more degraded. However, similar to Part 2, noise and peak shifts can affect the degradation calculation.

Orpiment

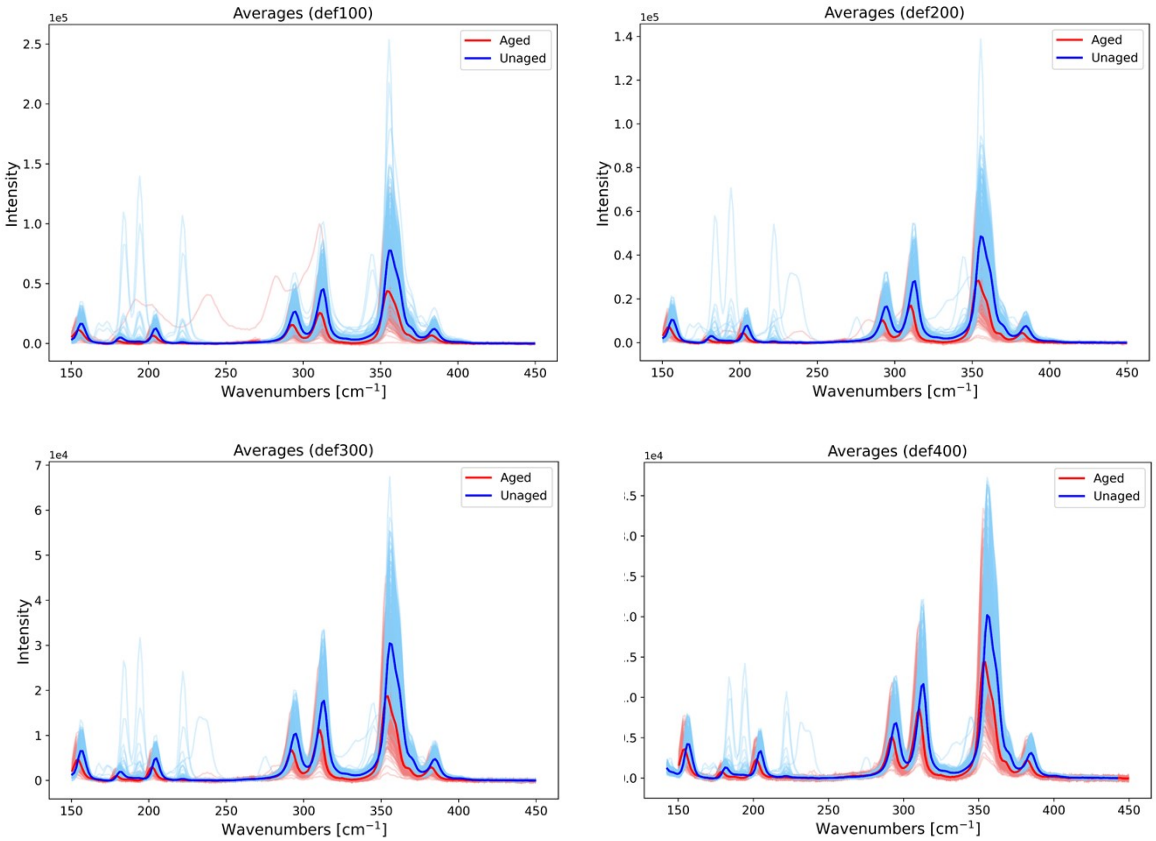


Figure S31. Full sets of spectra of the a) 100 µm, b) 200 µm, c) 300 µm and d) 400 µm defocusing steps of orpiment.

An important consideration can be made about these maps: when the signal is particularly low, like in the

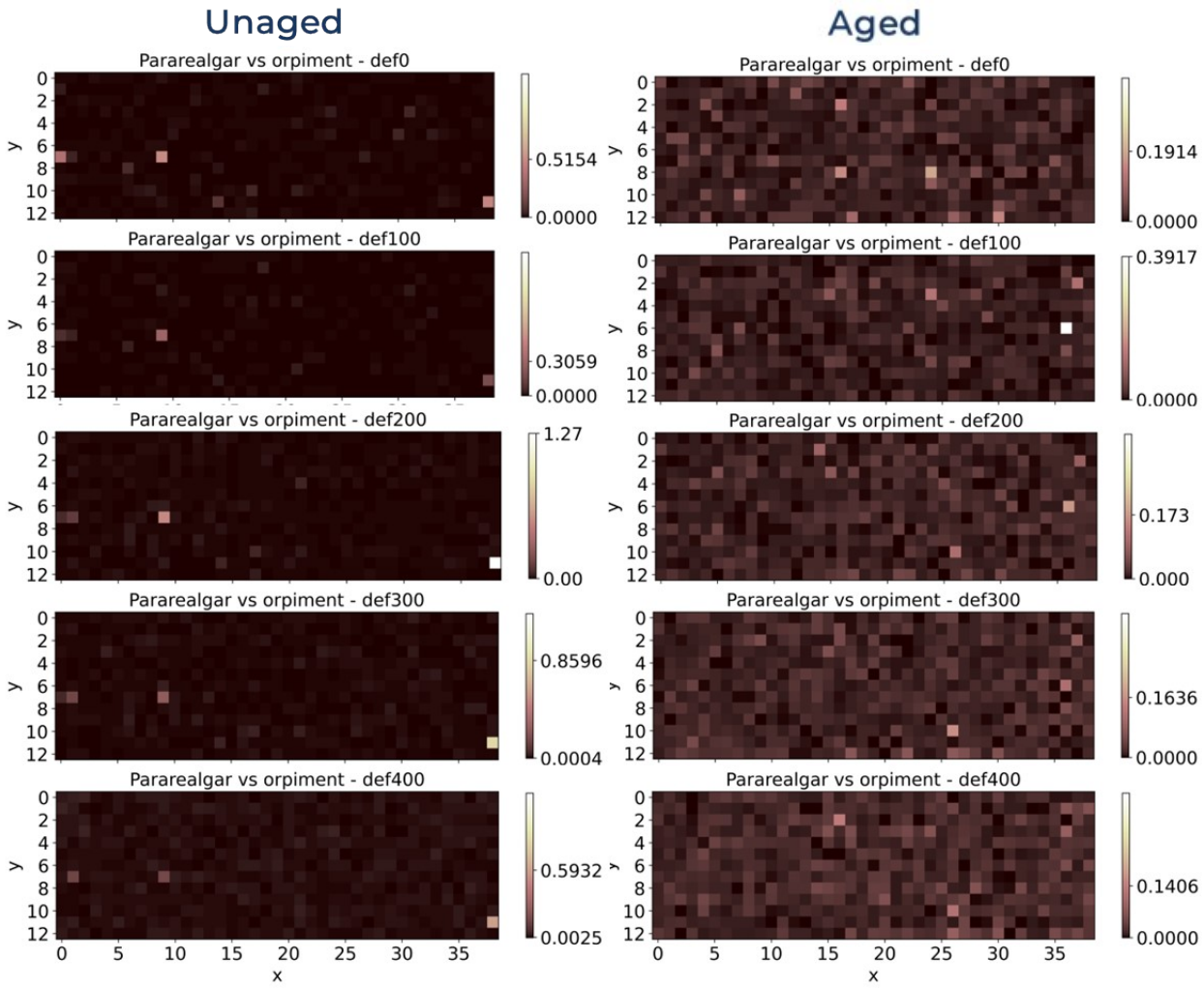


Figure S32. Raman maps of the unaged (left panels) and aged (right panels) orpiment, plotting pararealgar vs orpiment. The values of the colorbars indicate the maximum and minimum of each individual level with respect to the general scale.

case of pararealgar and realgar due to their very low amount in the sample (Figure S32 and S33), the

Raman maps introduce an instrumental effect when the defocusing steps increase. In fact, since the overall signal-to-noise ratio decreases at larger defocusing steps, the noise amount has a greater influence over the intensity ratios since it gets closer in the order of magnitude. That means that when bands are not

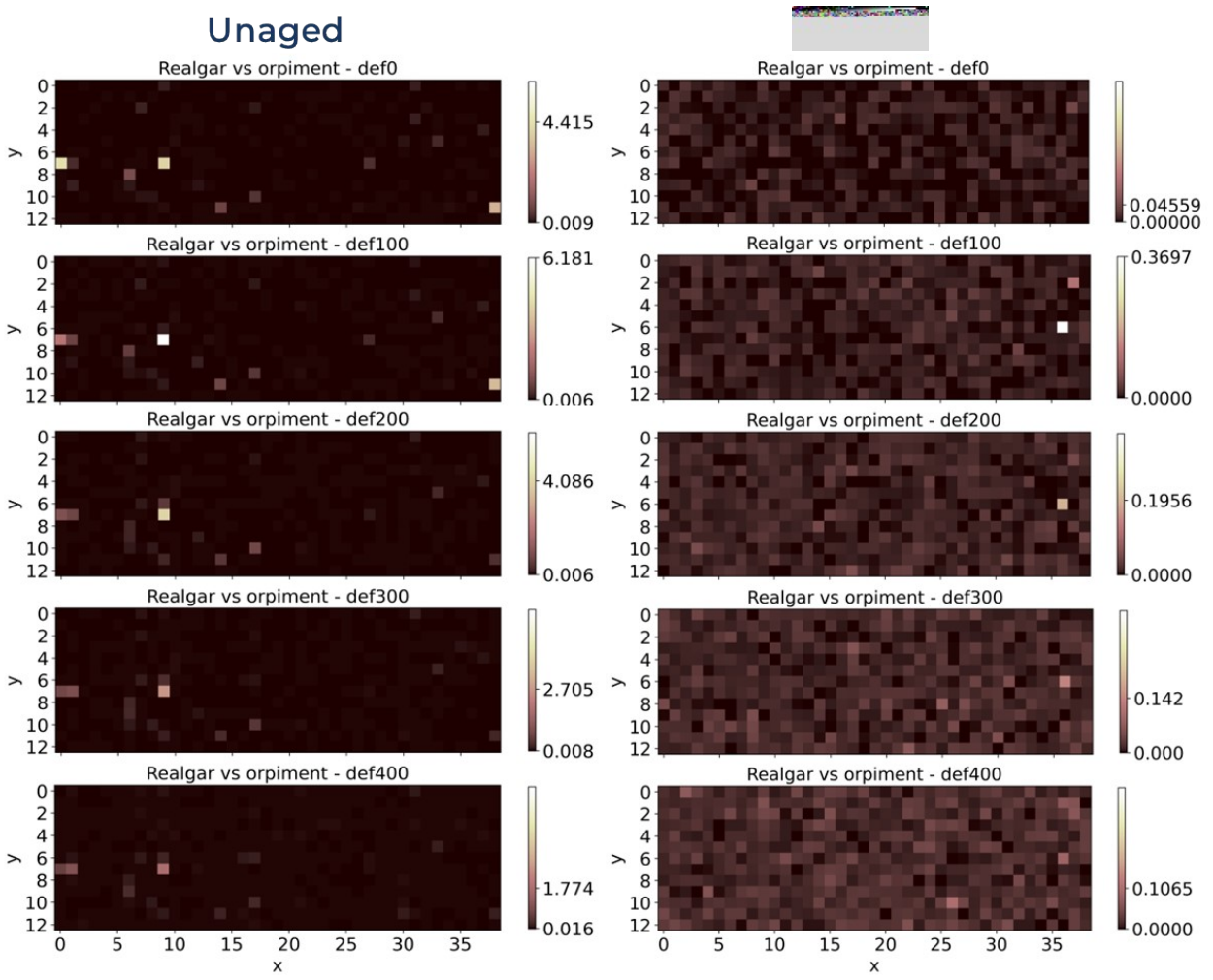


Figure S33. Raman maps of the unaged (left panels) and aged (right panels) orpiment, plotting realgar vs orpiment. The values of the colorbars indicate the maximum and minimum of each individual level with respect to the general scale.

strong enough to be easily distinguished from noise, a lighter background can be observed at higher defocusing steps, as it can be seen in the figures below showing the aged Raman maps of orpiment, plotting pararealgar vs orpiment. Moreover, we can again observe the outlier pixel at the second defocusing step, belonging to the pararealgar signal, already discussed above. The other bright pixel present at all defocusing steps is not related to pararealgar, as it can be observed by a closer look at the specific spectrum, but to the realgar pigment, whose band at 221 cm^{-1} influences the value found by the algorithm to the assigned pararealgar band (231 cm^{-1}). Since removing it would mean showing an almost completely dark map, we decided to keep it as an example of how cross interpretation is key when dealing with Raman maps of compounds possessing multiple degradation products. Another interesting information arises here: pararealgar is basically absent in the unaged sample, and realgar is limited to very small portions, whereas the aged sample shows only traces of pararealgar when compared to orpiment. Of course, as mentioned above, the spatial distribution of the aged and unaged samples cannot be compared, but this fact still indicates a probability that pararealgar does not form only as a degradation product of realgar but of orpiment as well.

As expected, the realgar vs orpiment map (Figure S33) did not have relevant evidence of the realgar pigment, as it is not a degradation product of orpiment (although traces can be found). The average brighter colour in the last defocusing steps is due to the decreasing orpiment signal, which becomes more comparable to the average noise. Given the particularity of the sudden bright pixel in Figure S32, we investigated further the specific spectrum belonging to that pixel, and we found out that such intensity ratio was generated by the script selecting the rising slope of pararealgar secondary band (274 cm^{-1}) rather than the arsenolite band. Since any manipulation of the data (narrower range selection for the arsenolite band or map subtraction) would've only partially fixed the issue, we decided that the best visualization would be a superimposition of the two maps, in order to differentiate whether a bright pixel is due to arsenolite or pararealgar. The result can be observed in Figure S34. Thanks to this visualization, it is much simpler to evaluate the contribution of each degradation product: if a pixel is green or one of its shades, it means that arsenolite presence is greater, and vice versa for red pixels; then, there could be cases where both compounds are basically absent (yellow pixels) or present (brown pixels). In the case of the unaged sample, it is interesting to notice that on the surface there is not much superimposition of the two signals, whereas starting from the second defocusing step we have one spectrum that possesses intense bands of both compounds. On the other hand, the aged sample seems to have a more homogeneous distribution of the degradation products, with little superimposition.

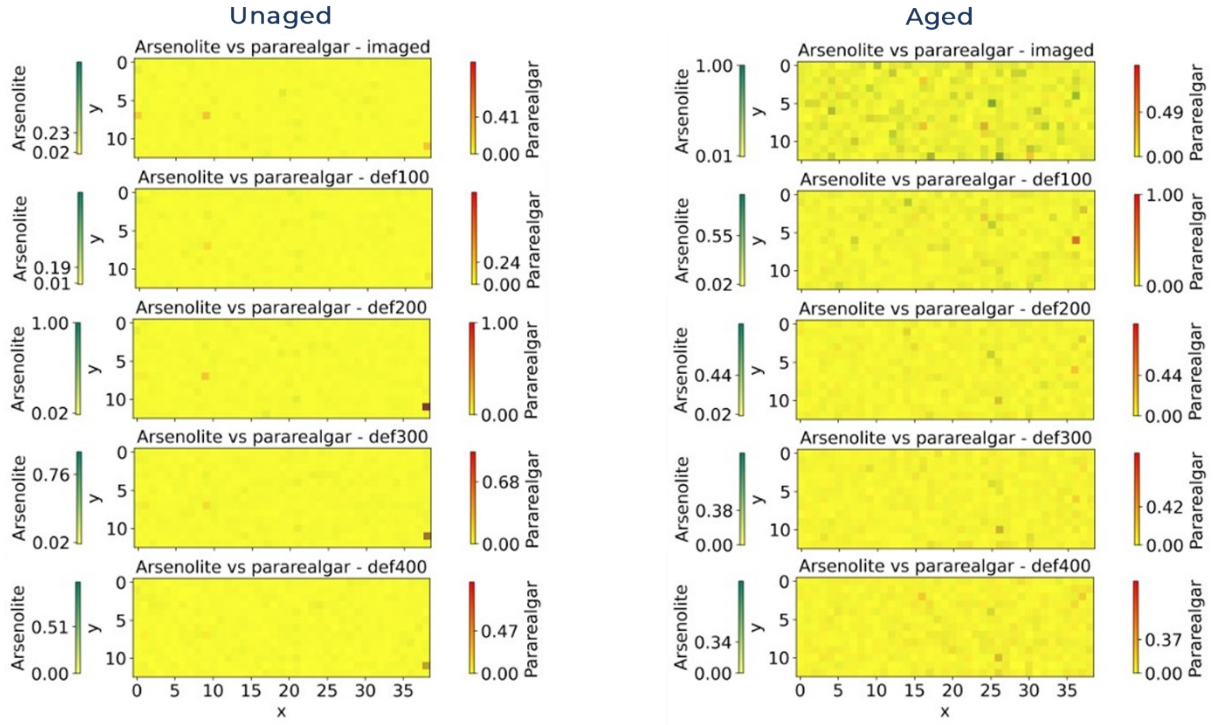


Figure S34. Superimposition of the pararealgar and arsenolite Raman maps of orpiment. Red pixels show only pararealgar presence, green pixels indicate arsenolite, brown the presence of both, and yellow complete absence.

Orpiment - Part 2. Unaged vs Aged Orpiment (surface)

Supervised Machine Learning: Unaged vs Aged Orpiment Spectra

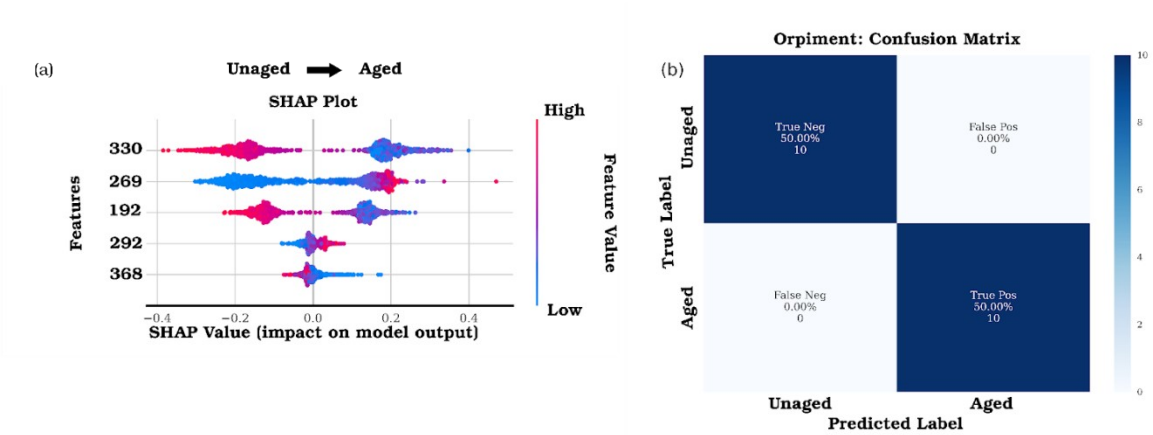


Figure S35. Summary of the orpiment unaged vs aged Extra Trees model results. (a) SHAP beeswarm plot showing the selected five features. A positive SHAP value explains whether features with intensities that increase (magenta) or decrease (blue) are more aged. (b) The confusion matrix of the ten unseen unaged and ten unseen aged spectra.

Unsupervised Learning: Unaged vs Aged Orpiment Spectra

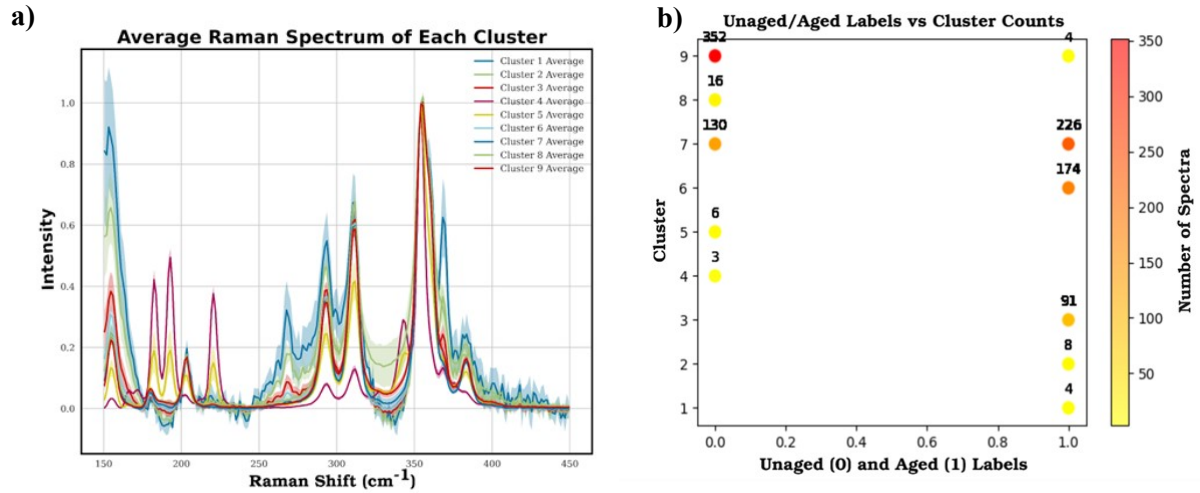


Figure S36. a) The average spectra of the nine clusters using the unaged and aged orpiment samples with the five selected features; b) The distribution of the unaged and aged orpiment lead samples into nine clusters. Orange represents a high number of spectra in a cluster, and yellow represents a low number of spectra in a cluster.

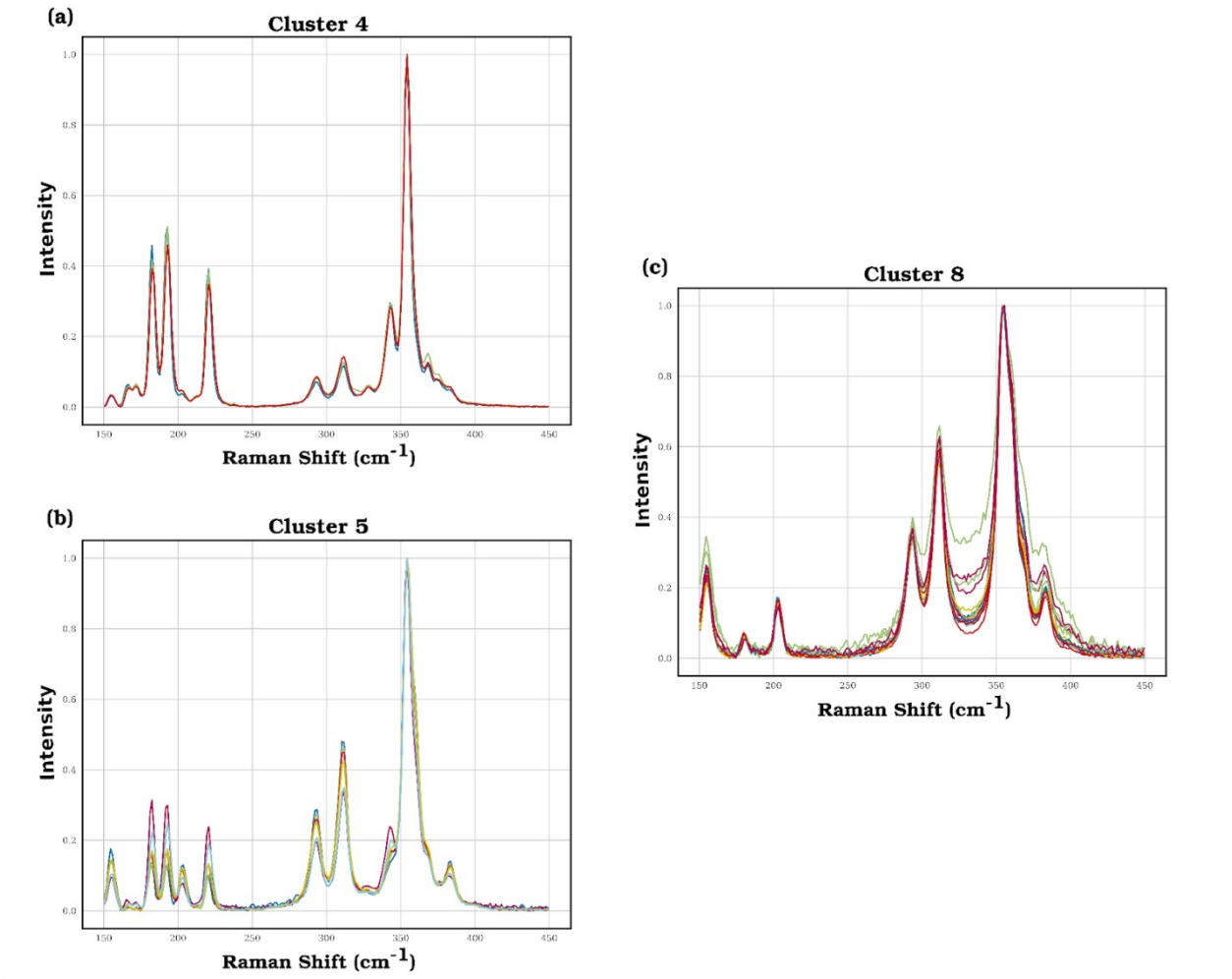


Figure S37. Summary of the orpiment unaged vs aged unsupervised clusters that only contain unaged labelled spectra. (a) Spectra from cluster 4, which contains realgar. (b) Spectra from cluster 5, which contains realgar. (c) Spectra from cluster 8, which contains orpiment.

Clusters 4, 5, and 8 (Figure S37a-c) only contain unaged spectra. The spectra in cluster 8 are noisy, but they appear to only contain orpiment. Clusters 4 and 5 contain realgar, with cluster 4 having higher realgar peaks.

Cluster 9 contains the most unaged spectra and four aged spectra. As shown in Figure S38a, cluster 9 appears to contain little to no arsenolite but many spectra with realgar present and possibly traces of pararealgar. The four aged spectra in cluster 9 are plotted in Figure S38b, which shows that the aged

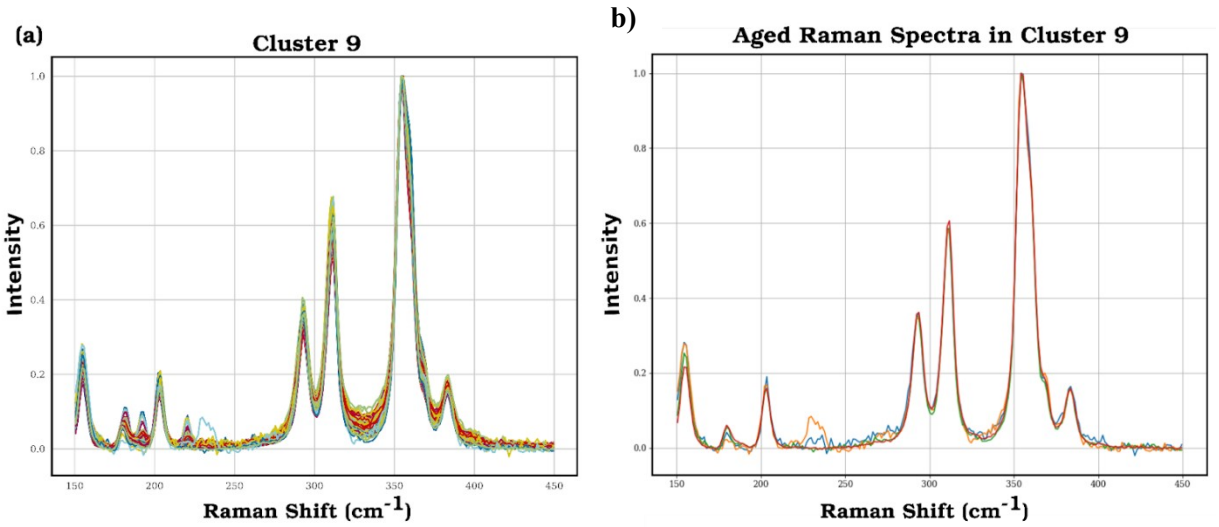


Figure S38. Summary of cluster 9 from the orpiment unaged vs aged unsupervised learning. (a) Spectra from cluster 9, which contains aged and unaged labelled spectra. The spectra contain realgar and possibly traces of pararealgar but no arsenolite. (b) The aged labelled spectra from cluster 9 showing the possible traces of pararealgar.

spectra may contain pararealgar but no realgar.

Clusters 1, 2, 3, and 6 (Figure S39a-d) only contain aged spectra. Clusters 1 and 2 are noisy, but they appear to only contain orpiment and arsenolite. Clusters 3 and 6 mostly contain orpiment and arsenolite with a few spectra containing realgar and possibly traces of pararealgar.

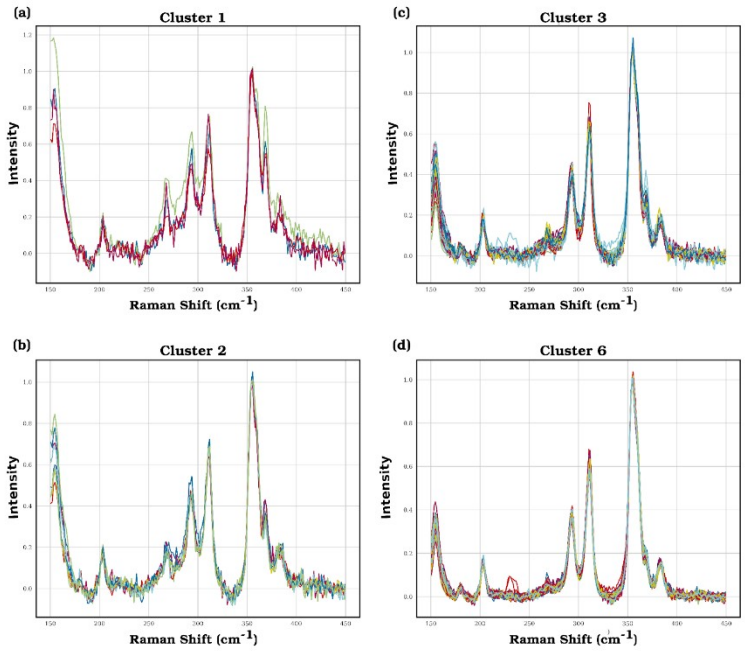


Figure S39. Summary of the orpiment unaged vs aged unsupervised clusters that only contain aged labelled spectra. (a) Spectra from cluster 1, which contains arsenolite. (b) Spectra from cluster 2, which contains arsenolite. (c) Spectra from cluster 3, which contains arsenolite and realgar and/or pararealgar. (d) Spectra from cluster 6, which contains arsenolite and few realgar and/or pararealgar.

Cluster 7 (Figure S40a) contains both unaged and aged spectra. Cluster 7 appears to contain mostly orpiment with very little arsenolite peaks, which could be a result of both samples (unaged and aged) being placed in this cluster. The aged spectra have such little arsenolite that they are grouped with the unaged orpiment spectra. In Figure S40b-c, the unaged and aged spectra in cluster 7 are plotted separately, which confirms the lack of arsenolite in the unaged spectra and the small presence of arsenolite in the aged spectra.

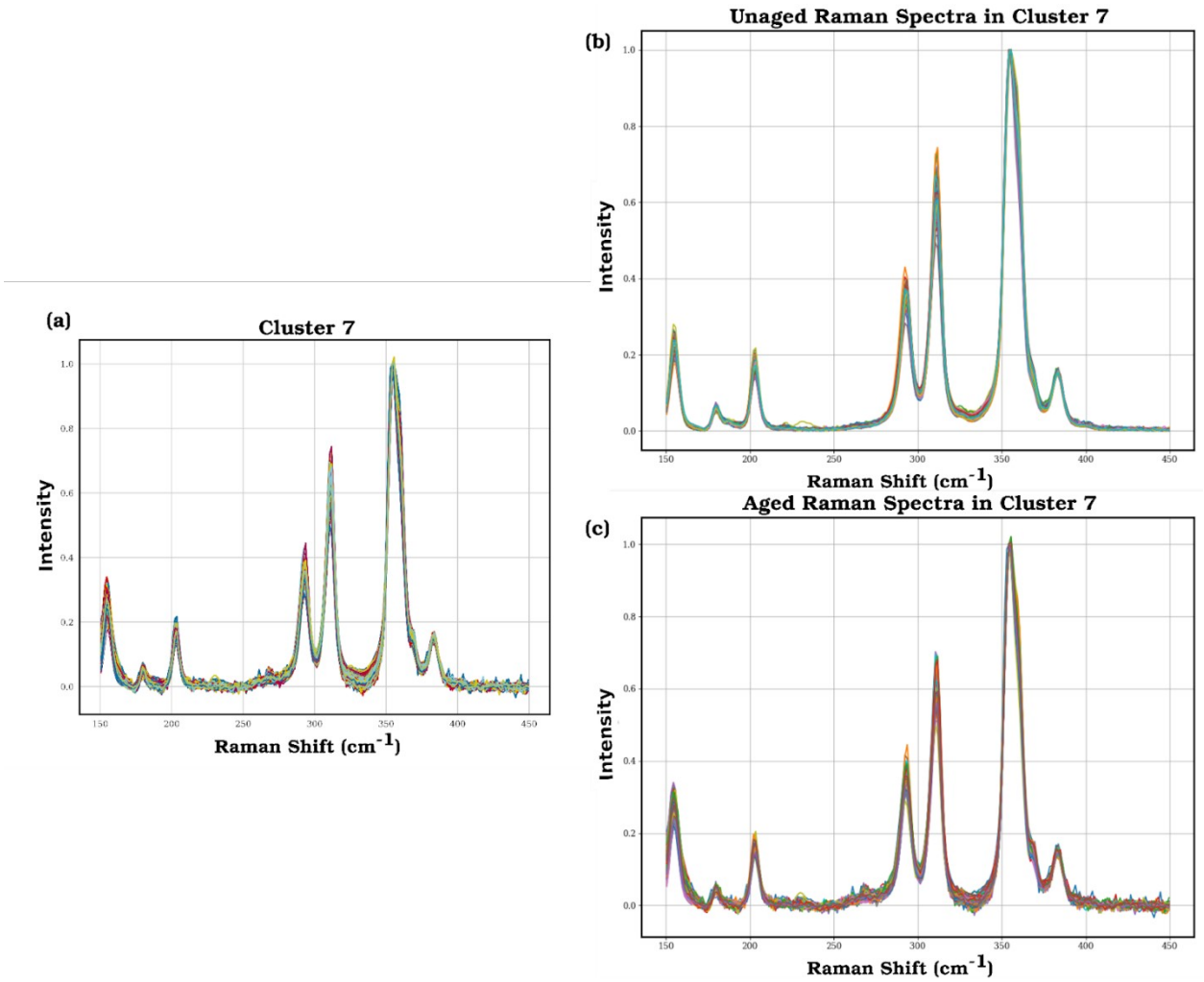


Figure S40. Summary of cluster 7 from the orpiment unaged vs aged unsupervised learning. (a) Spectra from cluster 7, which contains unaged and aged spectra. The spectra contain orpiment and very small arsenolite peaks. (b) The unaged spectra from cluster 7 showing no presence of arsenolite. (c) The aged spectra from cluster 7 showing very small arsenolite peaks.

Calculating Degradation: Unaged versus Aged Orpiment Spectra

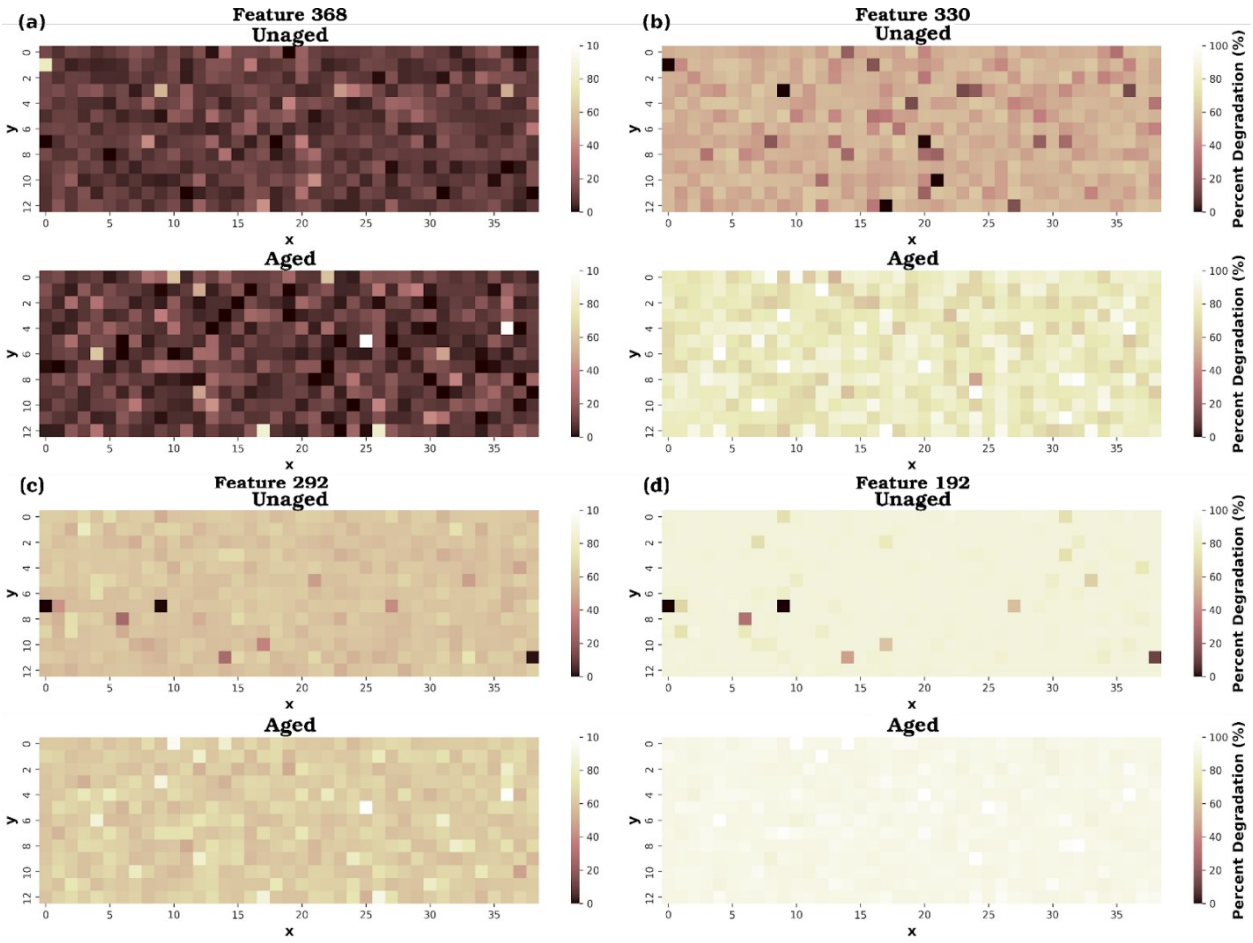


Figure S41. The percent degradation of the unaged and aged orpiment samples using features 368 (a), 330 (b), 292 (c), and 192 (d). A more degraded area of the sample is represented by white, and a more intact area is represented by dark red.

The spectra of the three red cells (Figure S42a-c) show no arsenolite peak, and the spectra of the three white/yellow cells (Figure S42d-f) show very high arsenolite peaks. The cell that appears more yellow has a lower arsenolite peak than the more white cells, which confirms that the method is correctly labelling the cells as intact or degraded. Therefore, the heatmap is able to accurately show areas of high and low

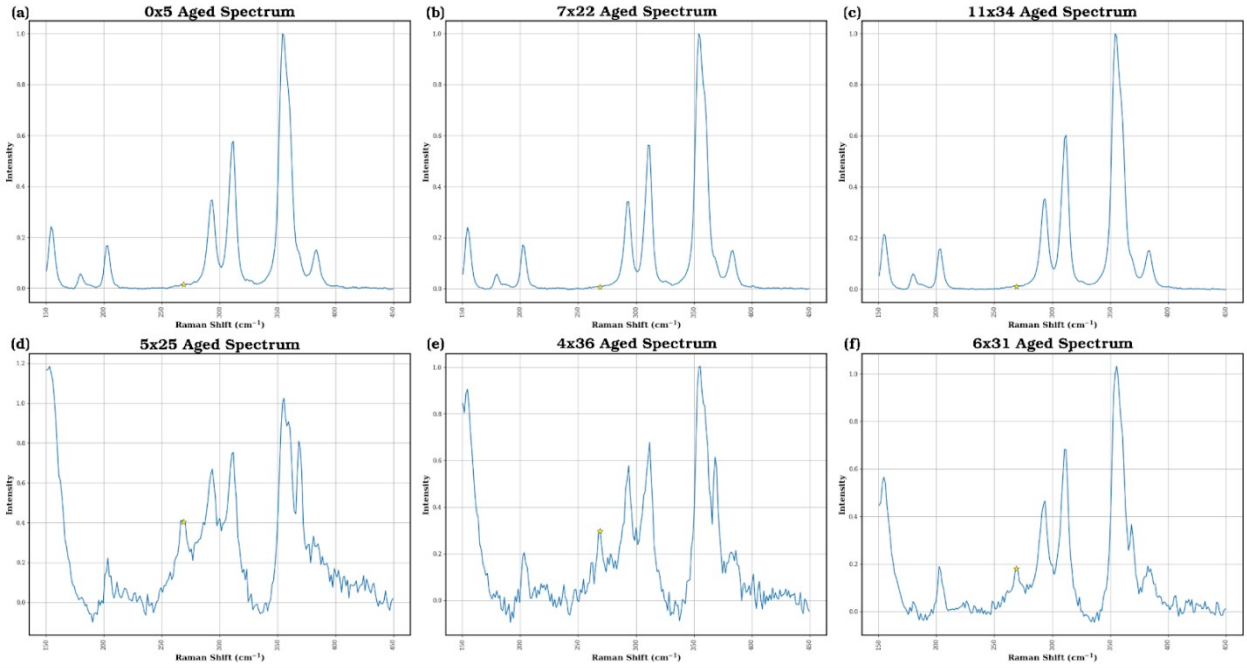


Figure S42. Raman spectra of the selected three red (intact) and three white (degraded) cells from the feature 269 aged orpiment heatmap. (a-c) Raman spectra of the intact cells. (d-f) Raman spectra of the degraded cells.

degradation without prior knowledge of sample labels (unaged or aged).

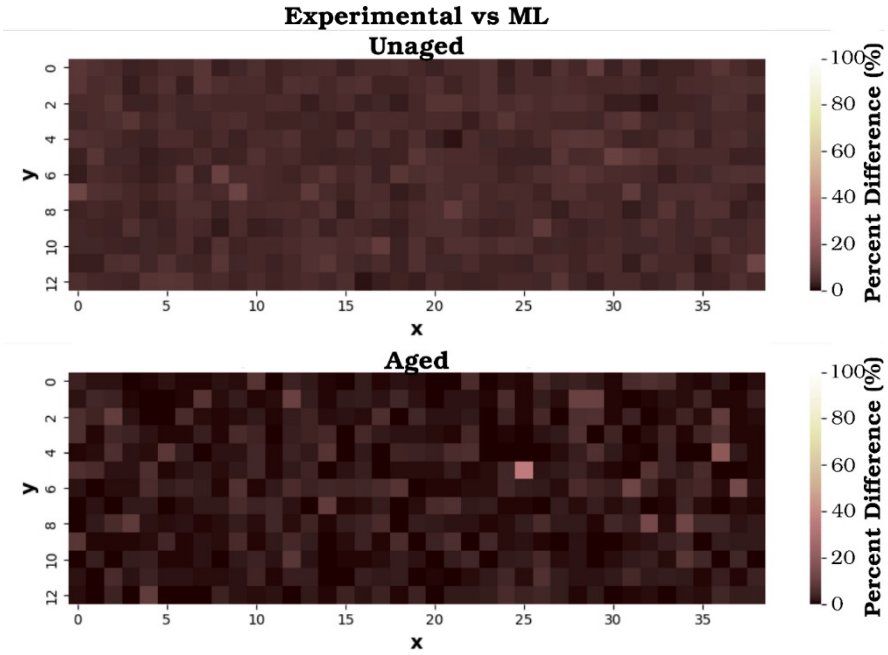


Figure S43. Comparison of the experimental and machine learning unaged and aged orpiment heatmaps. Areas that agree between the two heatmaps are represented by dark red, and areas that differ are white.

Orpiment - Part 3. Aged Orpiment Subsurface Study

Unsupervised Learning: Aged Orpiment Subsurface Study

Each class was placed into more than one cluster, which verifies the heterogeneity among each defocusing step. As shown in Figure S44, class 1 has spectra in clusters that contain spectra from class 0 (clusters 4, 6, and 8) and classes 2-4 (clusters 1-3), which confirms the high variability in classes 1 and 0. Spectra from class 0 appear similar to spectra found on the surface as well as in the deeper defocusing steps. The three deepest defocusing steps appear very similar, with all of the spectra except one being placed in clusters 1-3. As shown in Figure S23, the defocusing steps have less overlap compared to red lead, which is most likely due to the presence of multiple distinct degradation products. The clustering results (Figure S44) confirm the plateau observed when the Light Gradient Boosting Machine regression model predicted the outputs of the ten spectra from each class of the holdout set in Figure 10. It also suggests that all the degradation is captured, meaning that the degradation stops below the third defocusing step and more defocusing steps are not needed to further investigate the degradation. When unsupervised learning is used, some of the pararealgar and realgar spectra are placed in their own cluster, which makes their presence easier to identify, such as for clusters 5 and 9, respectively, in Figure S45b-c. However, these spectra are noisy and hard to distinguish the pigments present. The average spectra for each of the 9 clusters are also shown in Figure S45a. Cluster 8

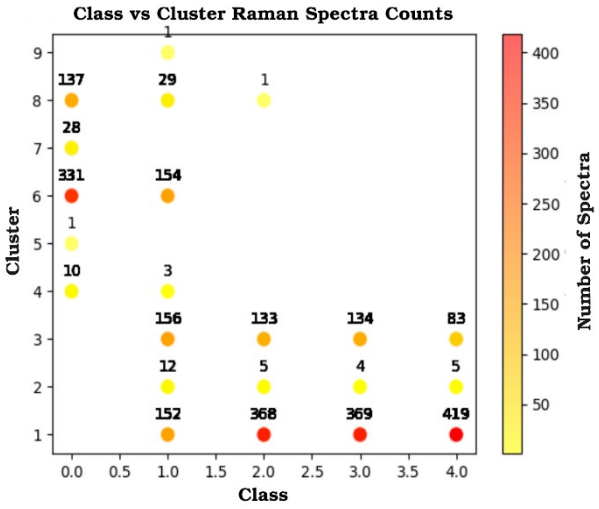


Figure 44. The distribution of the five defocusing steps into nine clusters. Orange represents a high number of spectra in a cluster, and yellow represents a low number of spectra in a cluster.

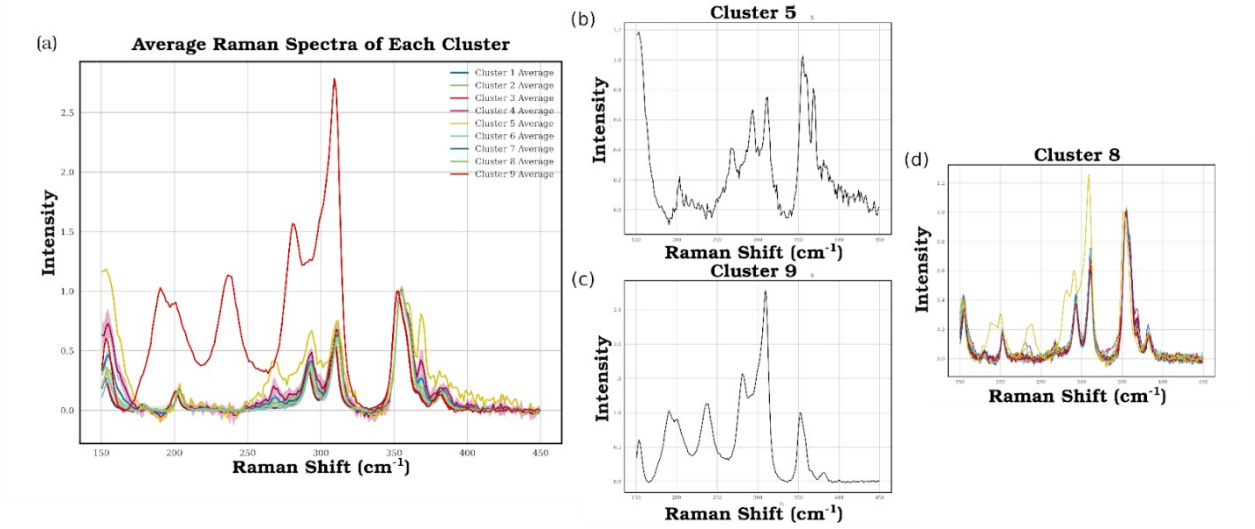


Figure S45. Summary of the unsupervised clusters using the orpiment defocusing steps with the 21 selected features. (a) Average spectra of each of the nine clusters. (b) Cluster 5 showing the possible presence of pararealgar. (c) Cluster 9 showing the presence of realgar. (d) Cluster 8 showing spectra with possible traces of pararealgar and realgar.

shows spectra with realgar and possibly traces of pararealgar present, as indicated by the peaks around 190 cm^{-1} and 230 cm^{-1}.

Calculating Degradation: Aged Orpiment Defocusing Steps

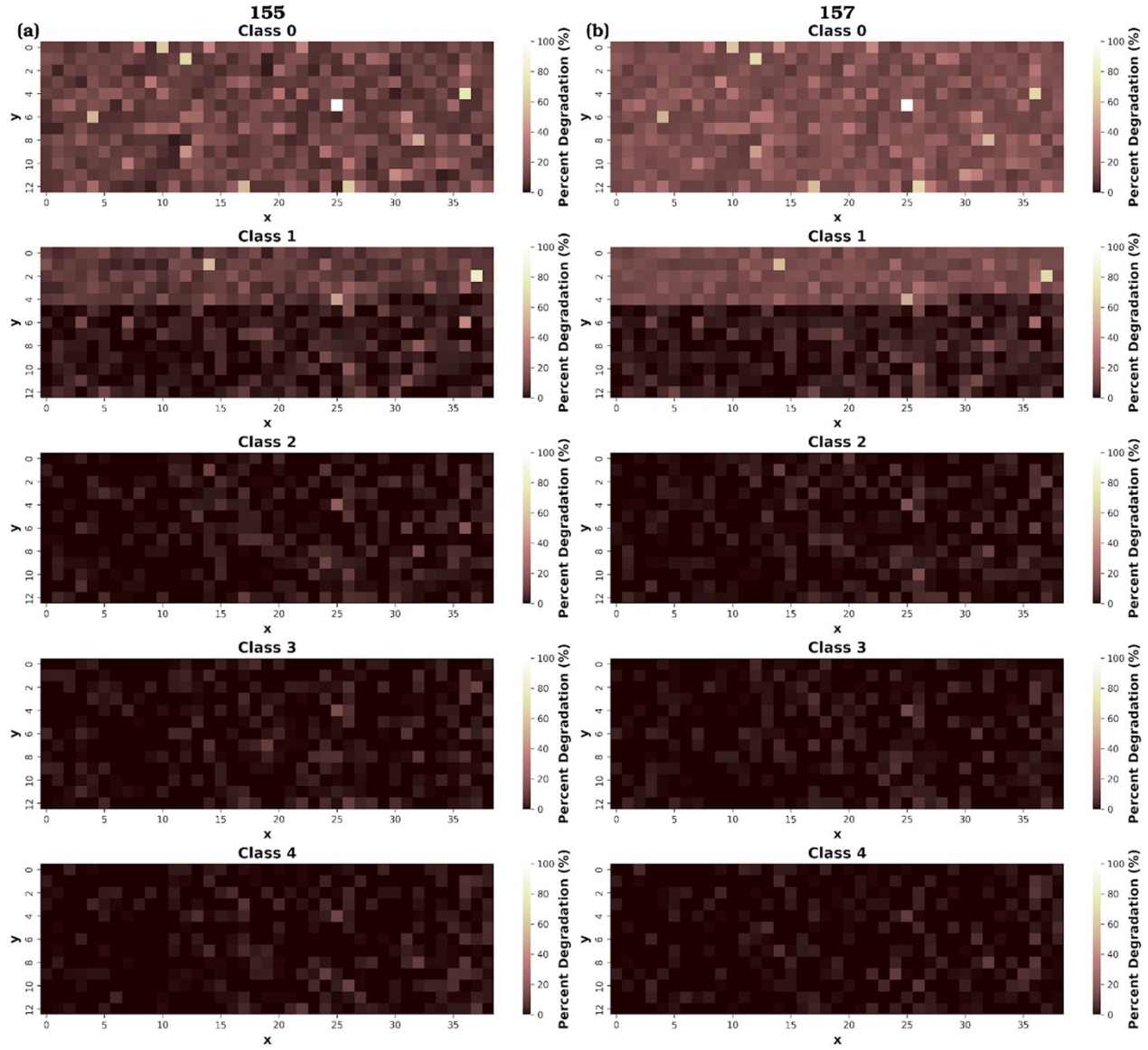


Figure S46. The percent degradation of the orpiment defocusing steps using features 155 (a) and 157 (b). A more degraded area of the sample is represented by white, and a more intact area is represented by dark red.

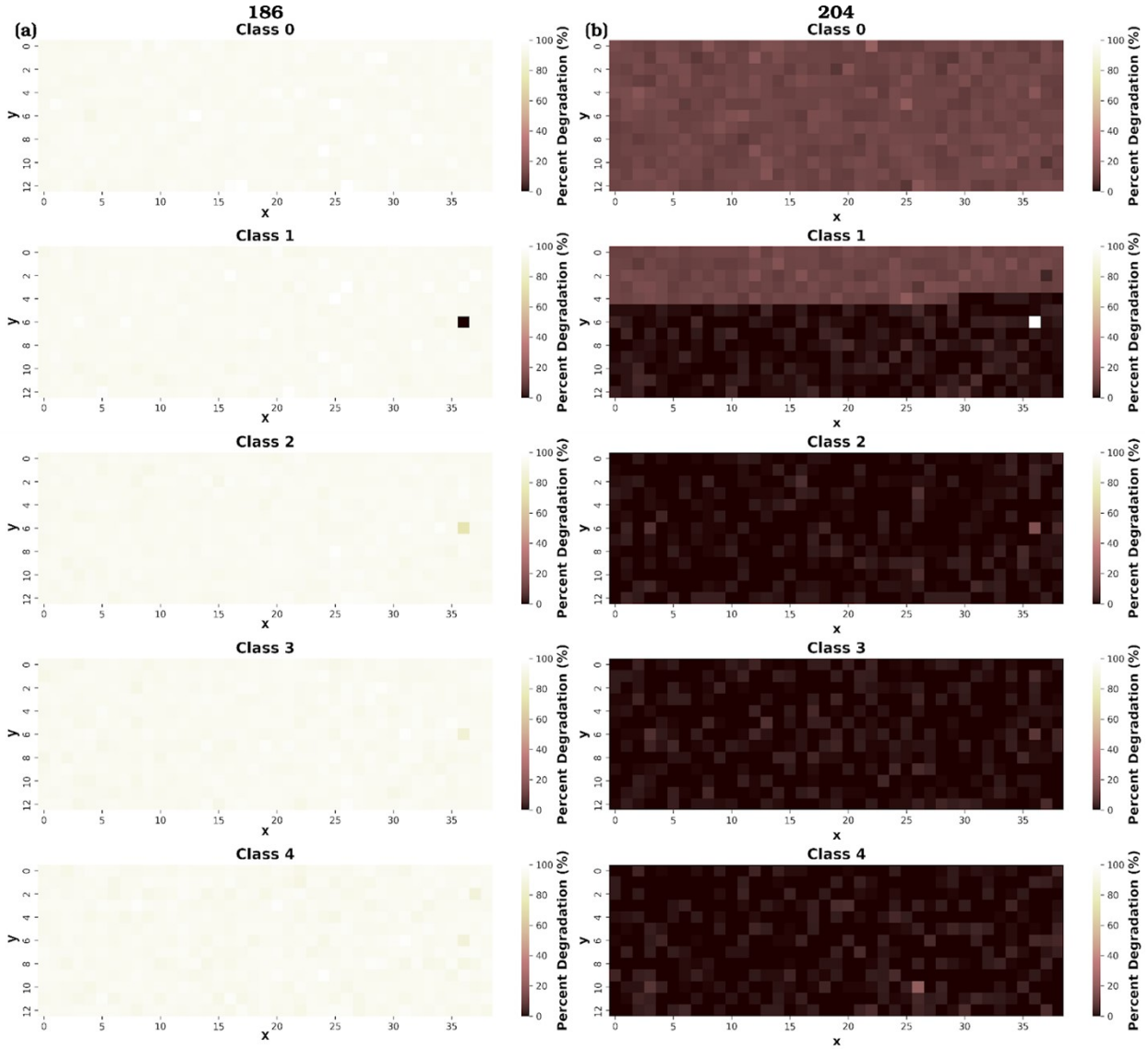


Figure S47. The percent degradation of the orpiment defocusing steps using features 186 (a) and 204 (b). A more degraded area of the sample is represented by white, and a more intact area is represented by dark red.

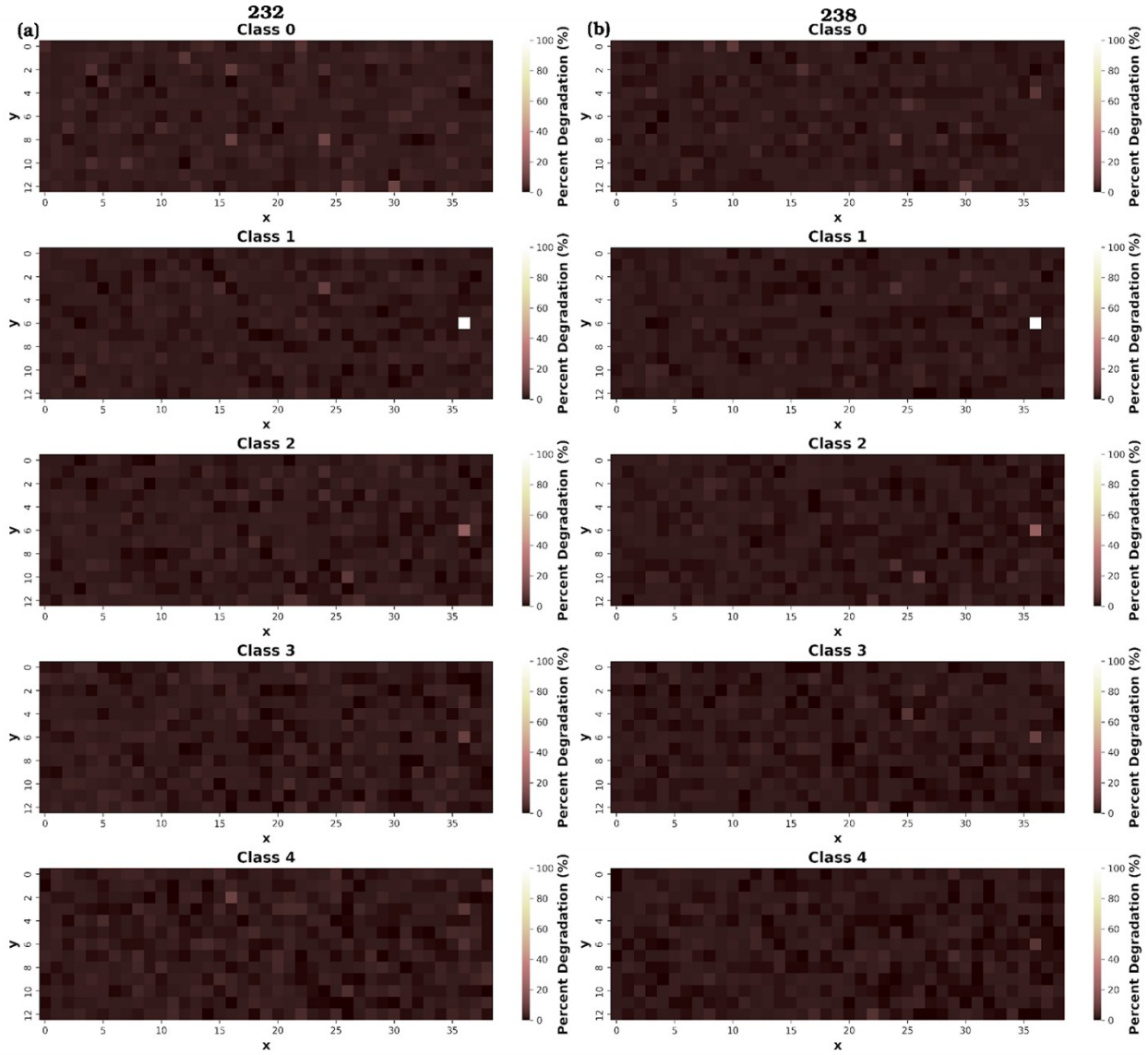


Figure S48. The percent degradation of the orpiment defocusing steps using features 232 (a) and 238 (b). A more degraded area of the sample is represented by white, and a more intact area is represented by dark red.

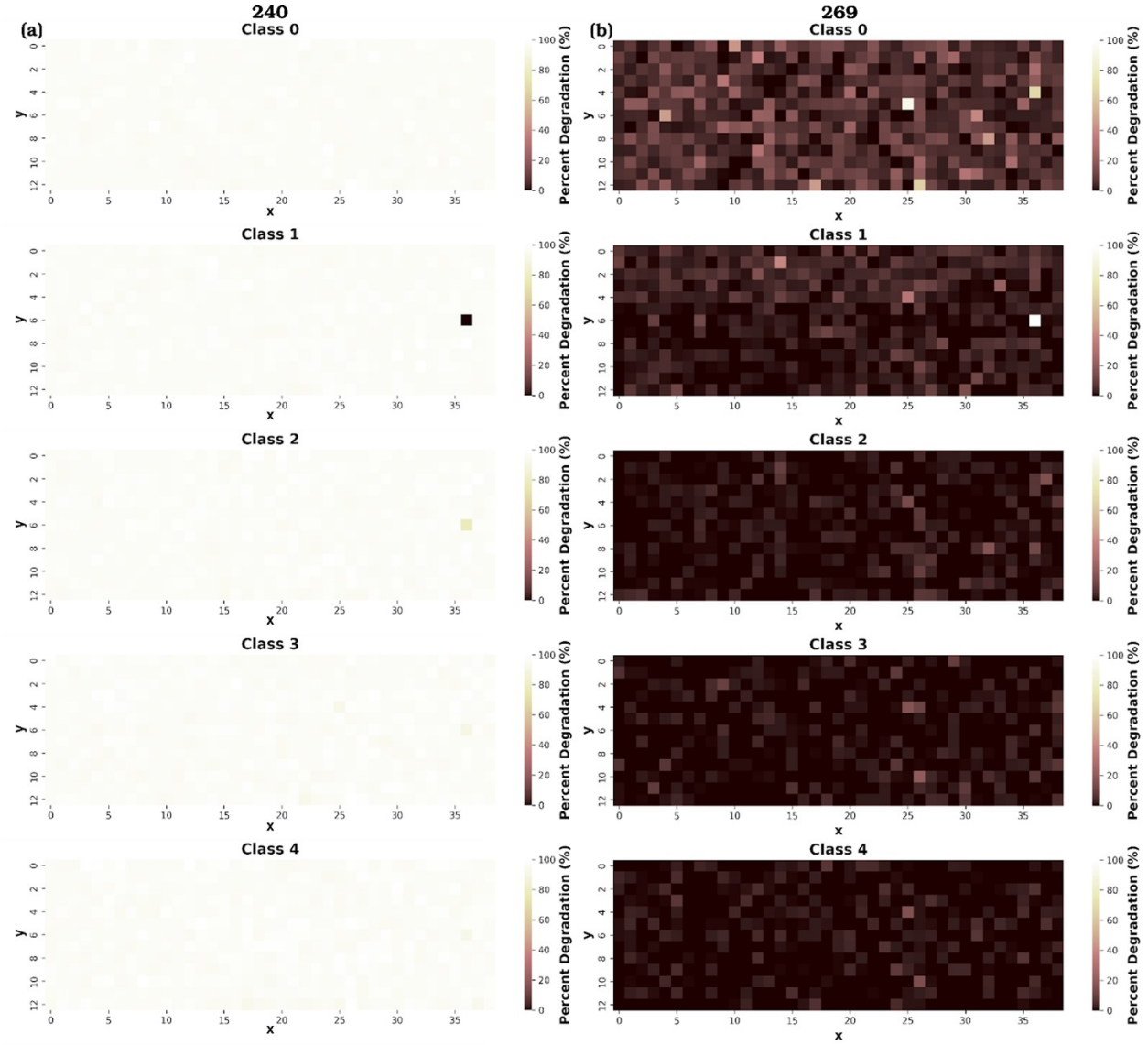


Figure S49. The percent degradation of the orpiment defocusing steps using features 240 (a) and 269 (b). A more degraded area of the sample is represented by white, and a more intact area is represented by dark red.

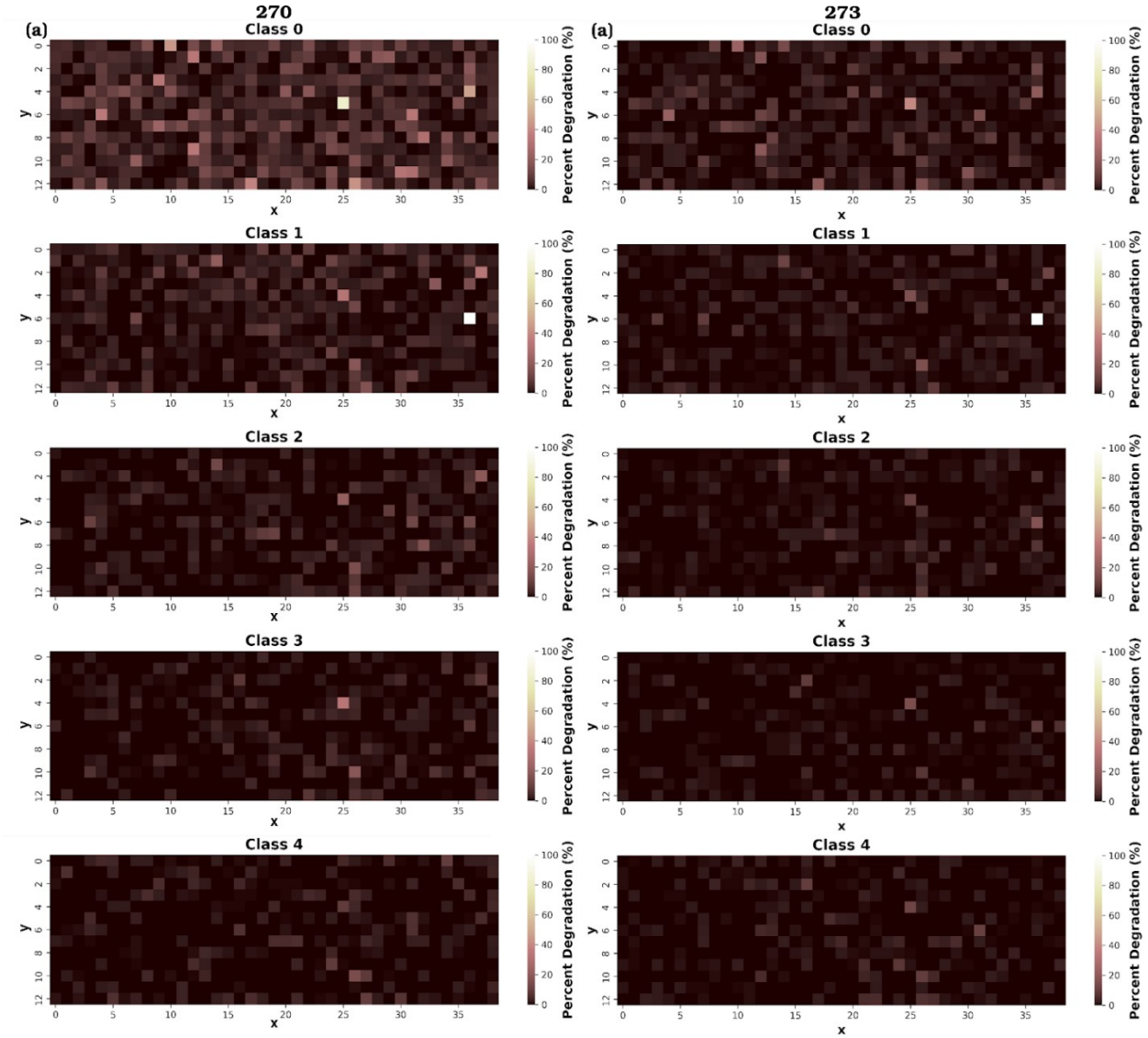


Figure S50. The percent degradation of the orpiment defocusing steps using features 270 (a) and 273 (b). A more degraded area of the sample is represented by white, and a more intact area is represented by dark red.

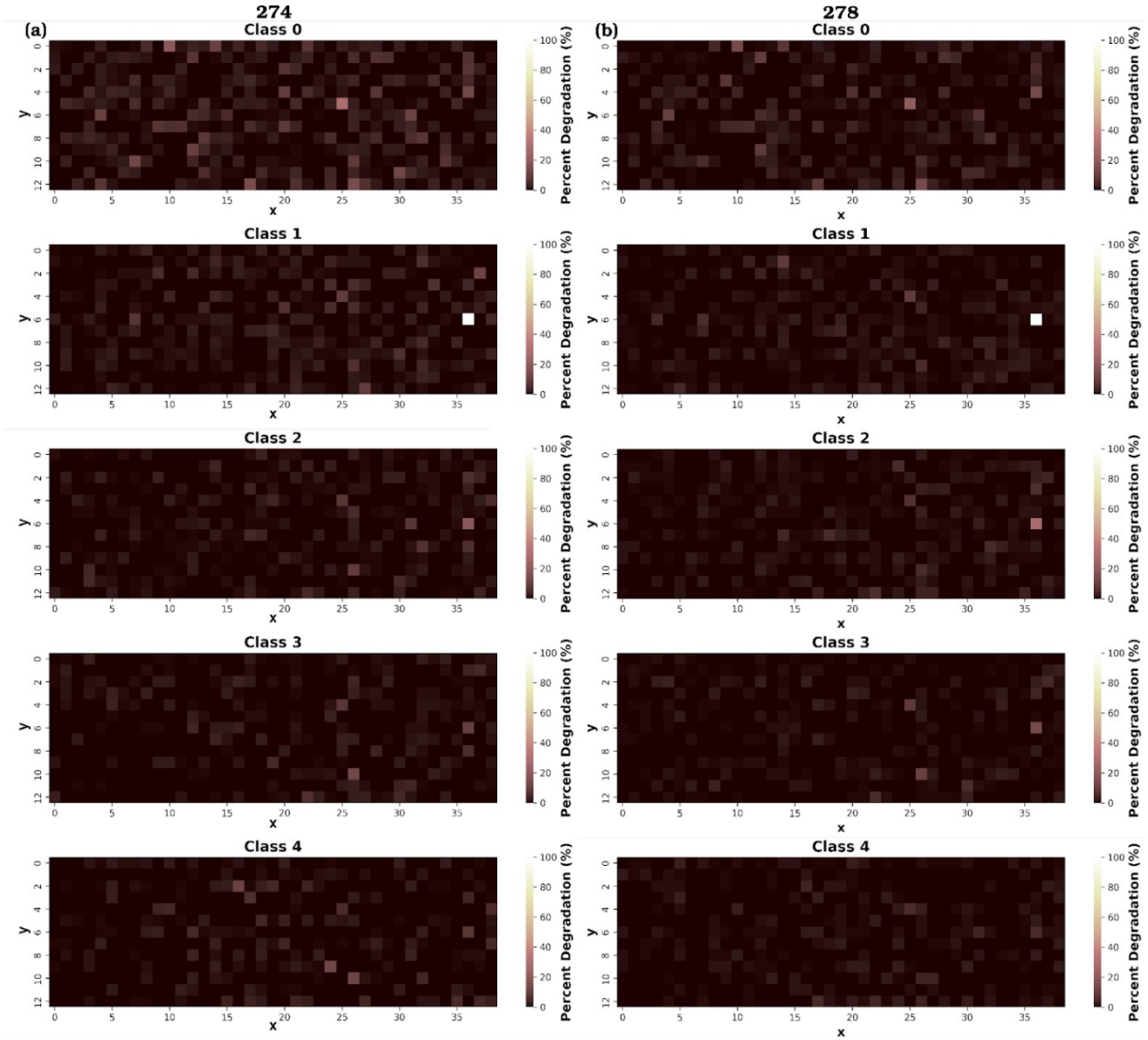


Figure S51. The percent degradation of the orpiment defocusing steps using features 274 (a) and 278 (b). A more degraded area of the sample is represented by white, and a more intact area is represented by dark red.

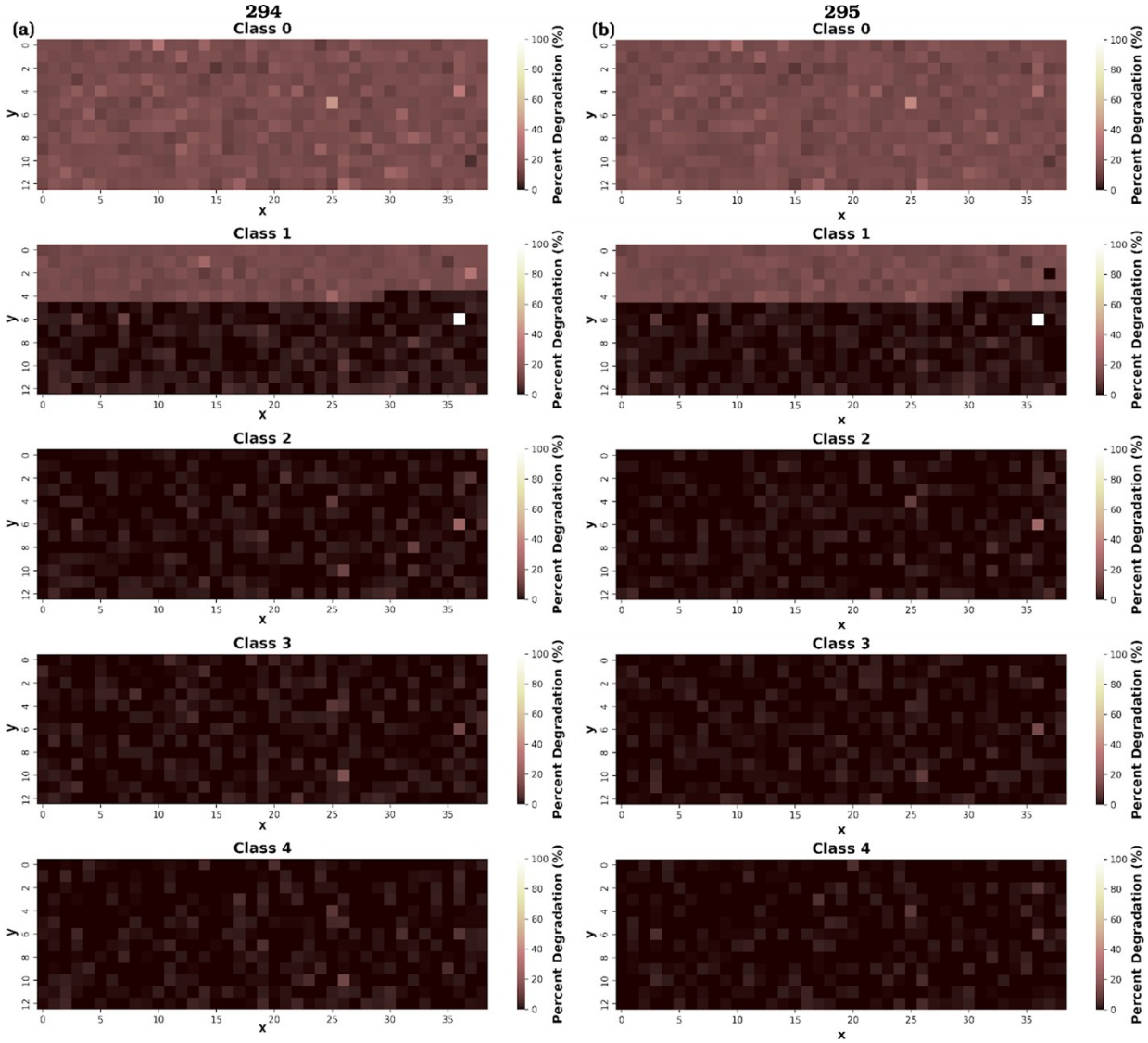


Figure S52. The percent degradation of the orpiment defocusing steps using features 294 (a) and 295 (b). A more degraded area of the sample is represented by white, and a more intact area is represented by dark red.

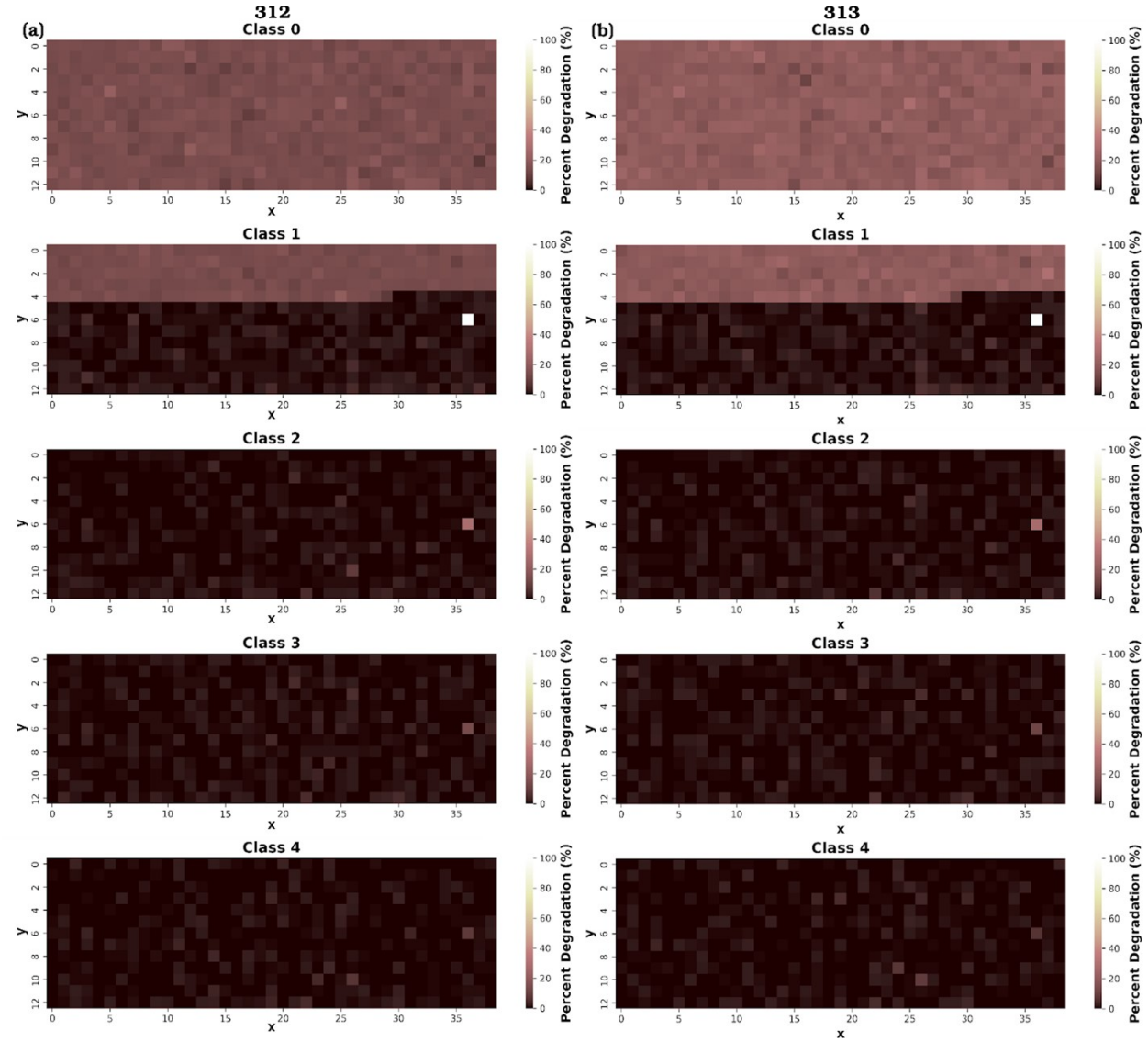


Figure S53. The percent degradation of the orpiment defocusing steps using features 312 (a) and 313 (b). A more degraded area of the sample is represented by white, and a more intact area is represented by dark red.

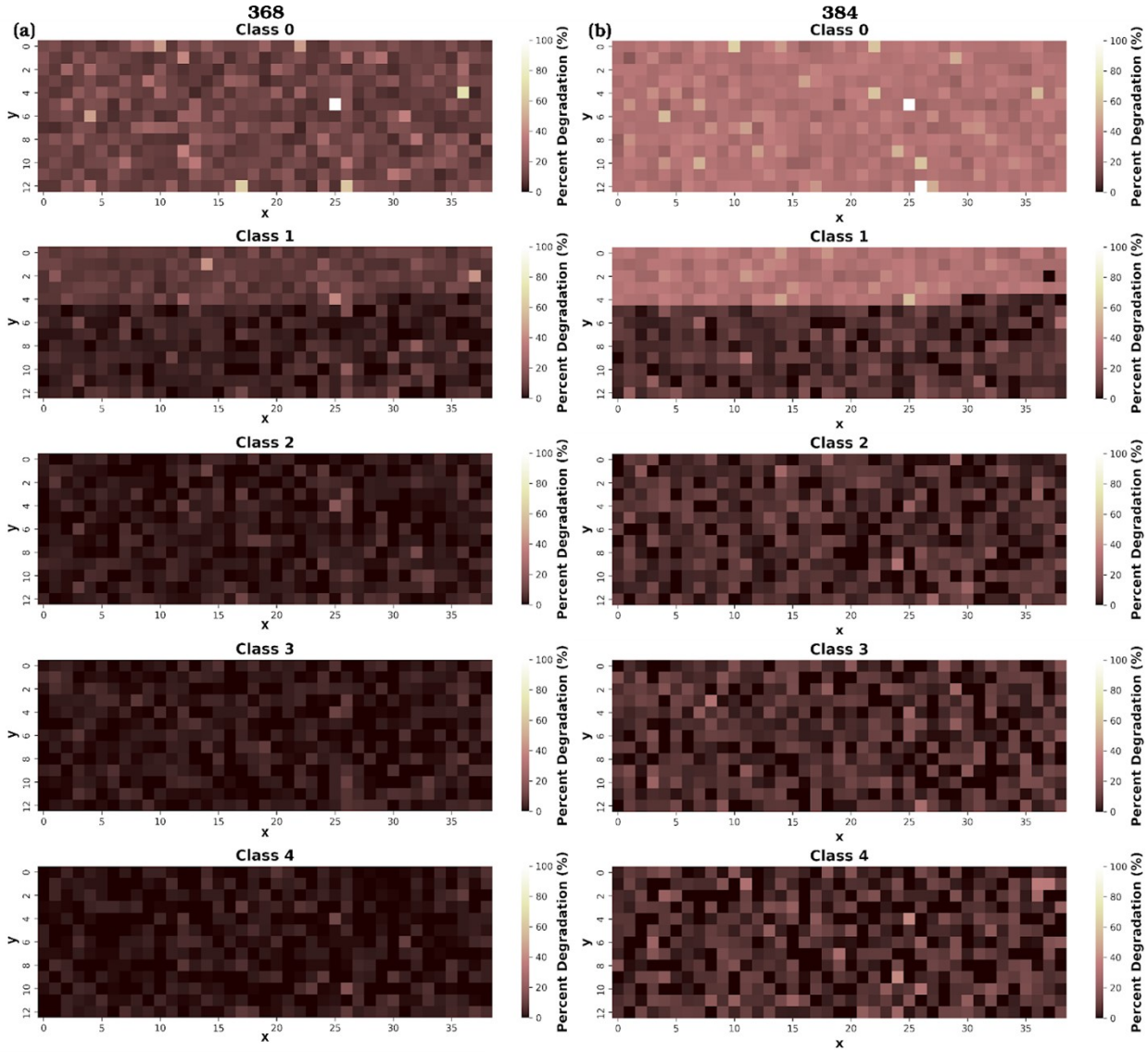


Figure S54. The percent degradation of the orpiment defocusing steps using features 368 (a) and 384 (b). A more degraded area of the sample is represented by white, and a more intact area is represented by dark red.

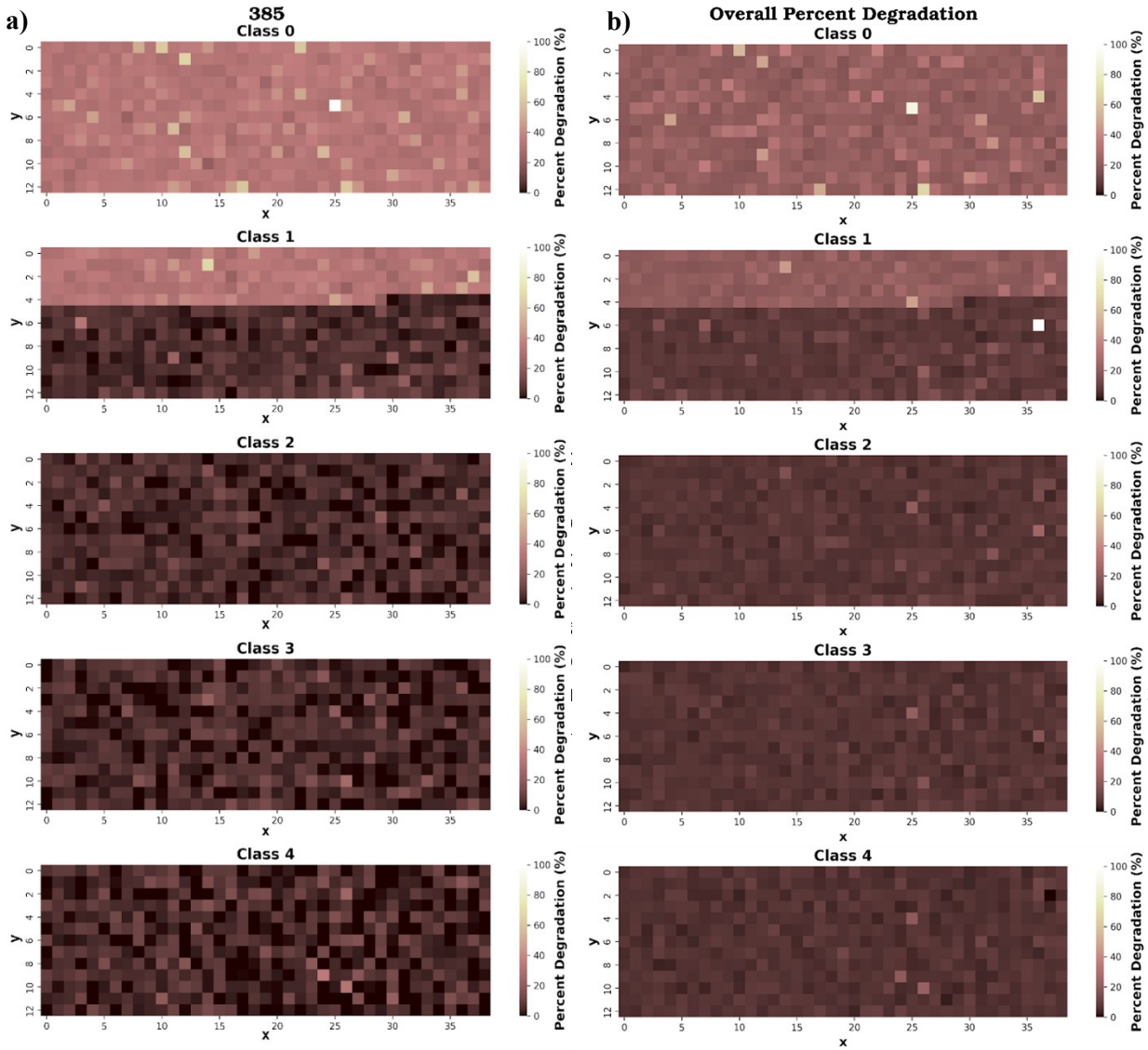


Figure S55. The percent degradation of the orpiment defocusing steps using (a) feature 385 or (b) all features. A more degraded area of the sample is represented by white, and a more intact area is represented by dark red.

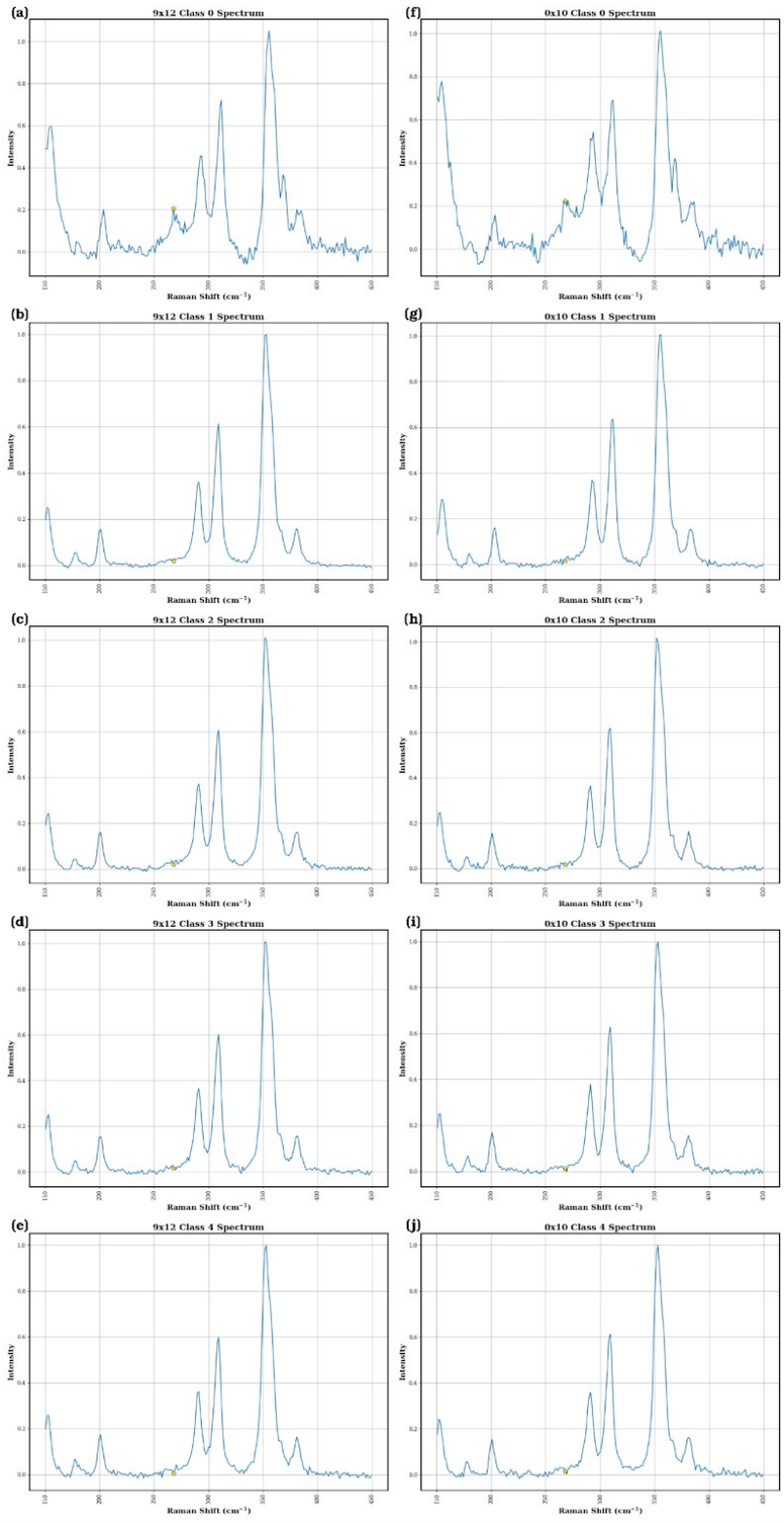


Figure S56. Raman spectra of two degraded (white) cells in the class 0 orpiment defocusing step heatmap showing how the spectra change as the defocusing steps (classes) increase. (a-e) The Raman spectra in each defocusing step corresponding to the 9x12 cell in the heatmap. (f-j) The Raman spectra in each defocusing step corresponding to the 0x10 cell in the heatmap. The yellow star denotes the position of feature 268.

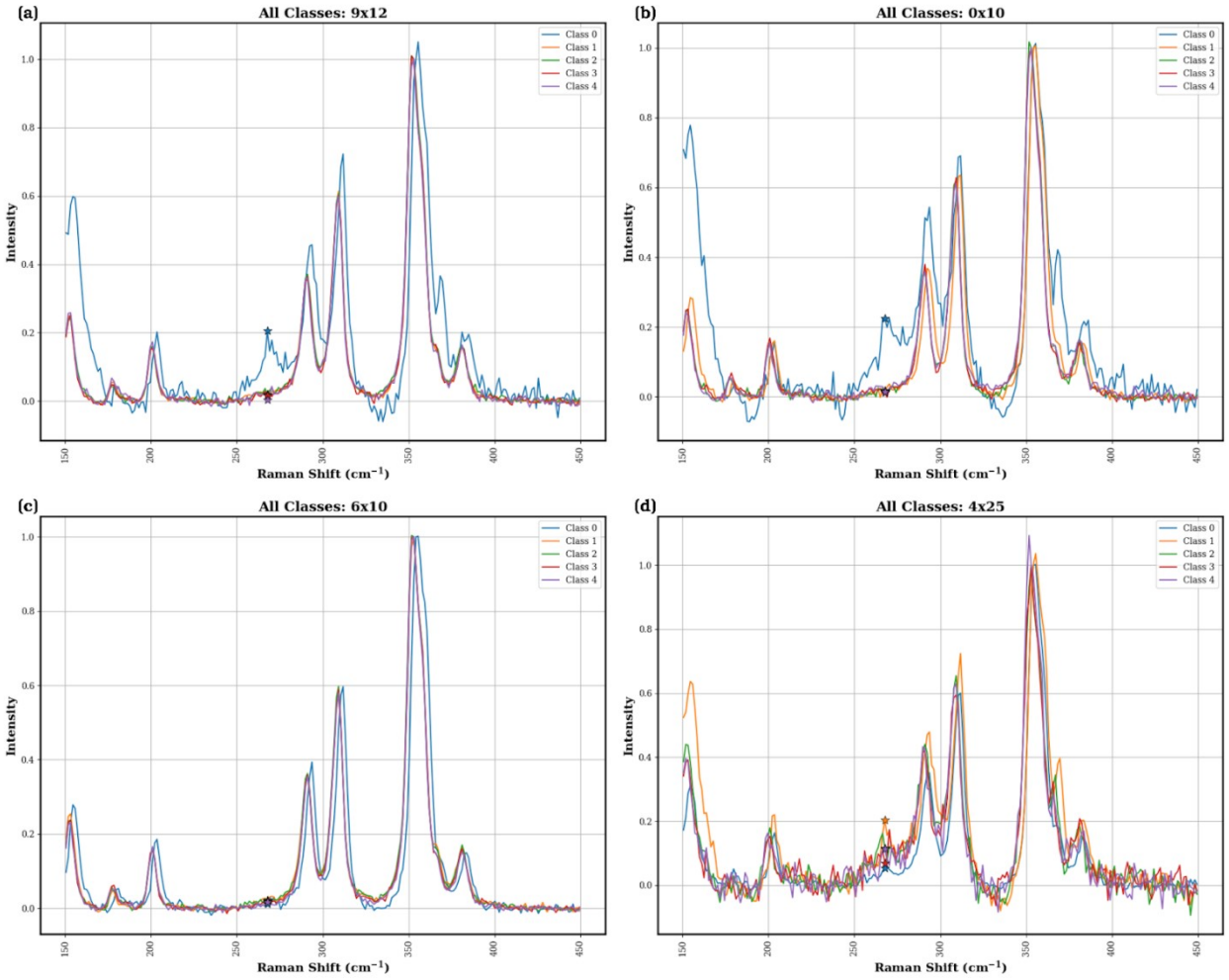


Figure S57. Summary of the selected cells in the orpiment defocusing step heatmaps to show the intensity (and degradation) patterns. (a-b) The selected degraded (white) cells in the heatmaps. (c-d) The selected intact (red) cells in the heatmaps. The star denotes the position of feature 268.

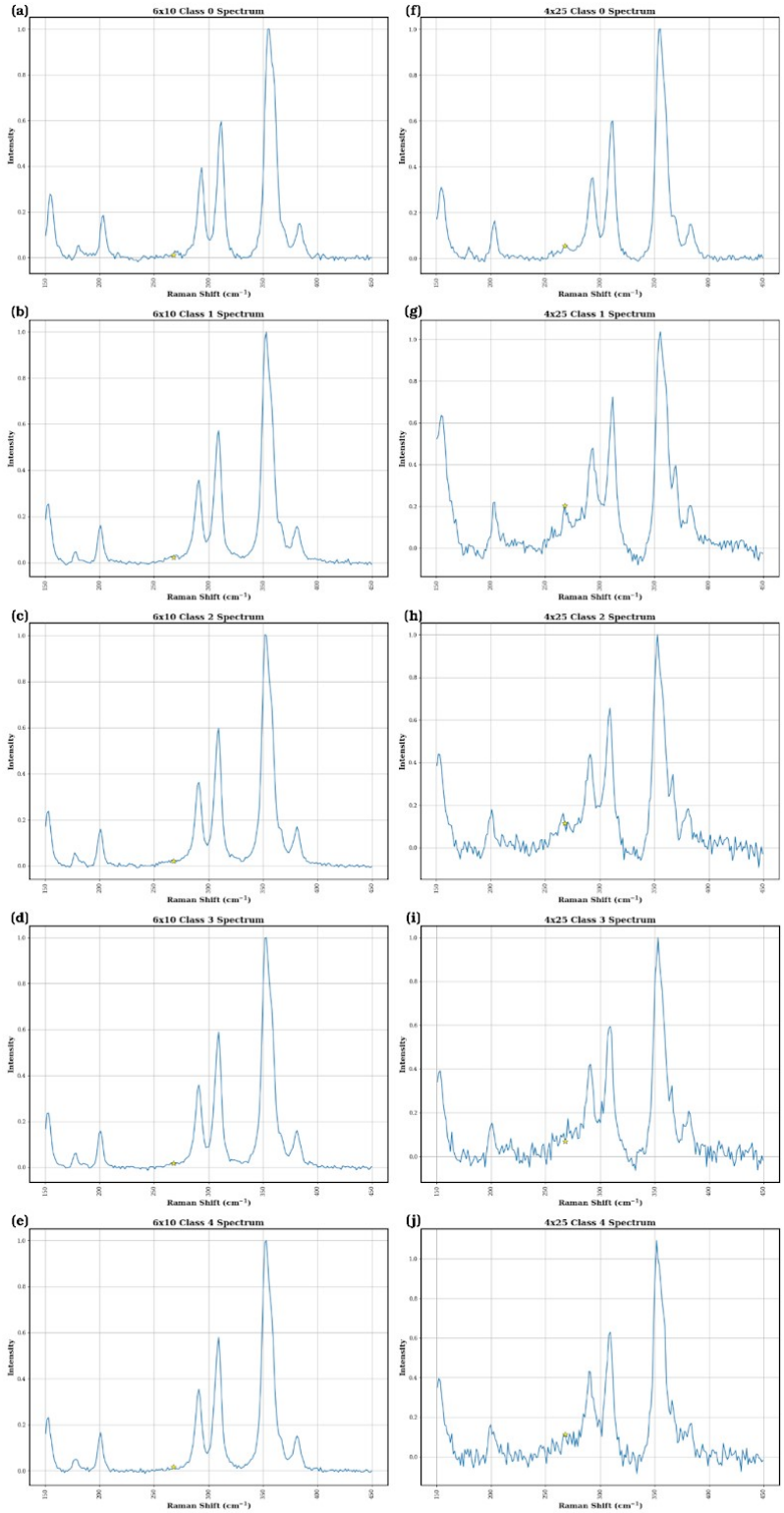
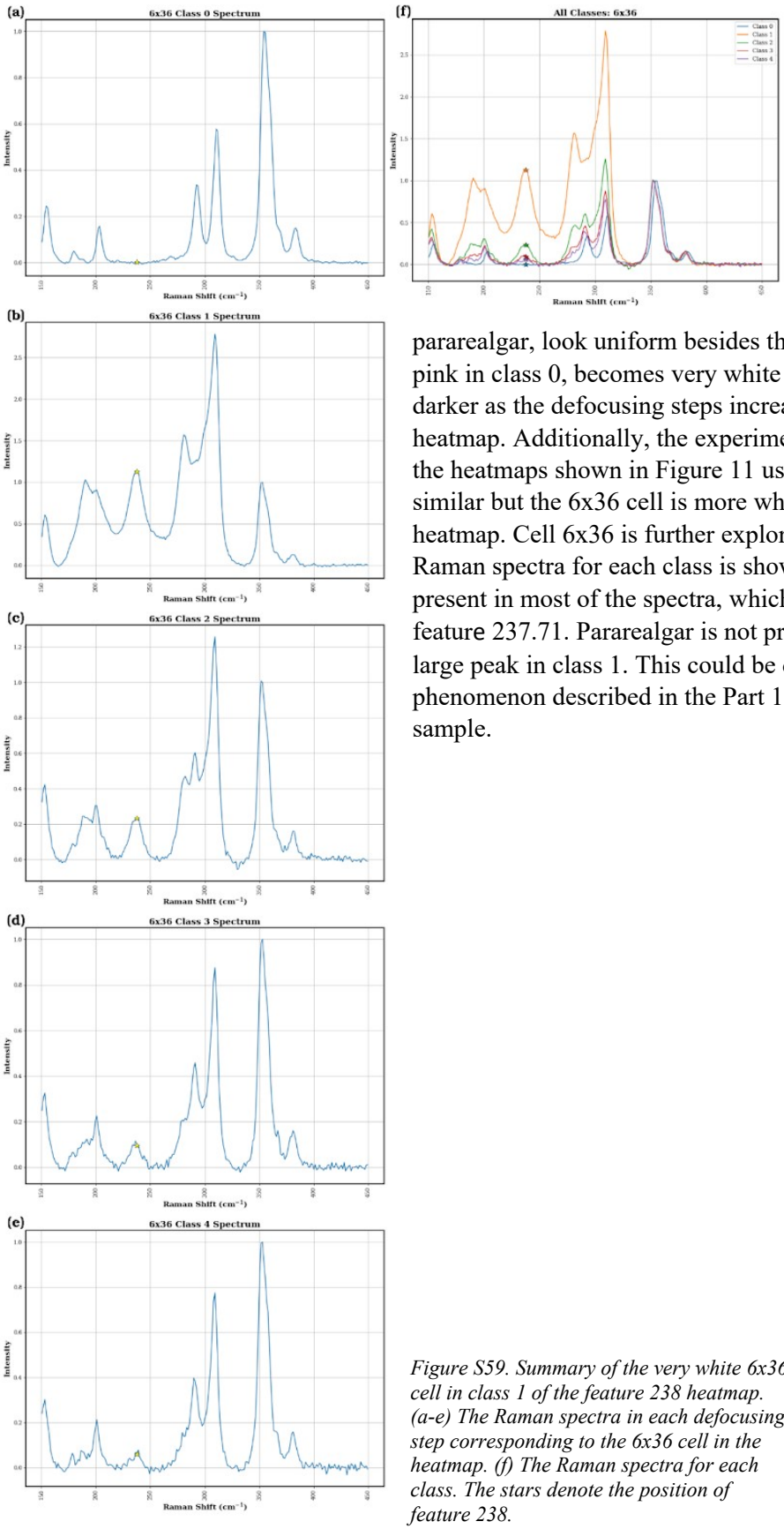


Figure S58. Raman spectra of two intact (dark red) cells in the class 0 orpiment defocusing step heatmap showing how the spectra change as the defocusing steps (classes) increase. (a-e) The Raman spectra in each defocusing step corresponding to the 6x10 cell in the heatmap. (f-j) The Raman spectra in each defocusing step corresponding to the 4x25 cell in the heatmap. The yellow star denotes the position of feature 268.



Another area of interest on the heatmaps in Figure 11 and S48b is cell 6x36, which appears very degraded in the heatmaps in class 1 for features 267.73 and 237.71. Interestingly, the heatmaps for classes 0-4 created using feature 237.71, which represents pararealgar, look uniform besides this very white cell. It is a light pink in class 0, becomes very white in class 1, and then becomes darker as the defocusing steps increase in the feature 237.71 heatmap. Additionally, the experimental heatmaps (Figure 5) and the heatmaps shown in Figure 11 using feature 267.73 are very similar but the 6x36 cell is more white in the machine learning heatmap. Cell 6x36 is further explored in Figure S59, where the Raman spectra for each class is shown. Pararealgar appears to be present in most of the spectra, which causes the high intensity for feature 237.71. Pararealgar is not present in class 0 but has a very large peak in class 1. This could be due to the same instrumental phenomenon described in the Part 1 analysis of the orpiment sample.

Figure S59. Summary of the very white 6x36 cell in class 1 of the feature 238 heatmap. (a-e) The Raman spectra in each defocusing step corresponding to the 6x36 cell in the heatmap. (f) The Raman spectra for each class. The stars denote the position of feature 238.

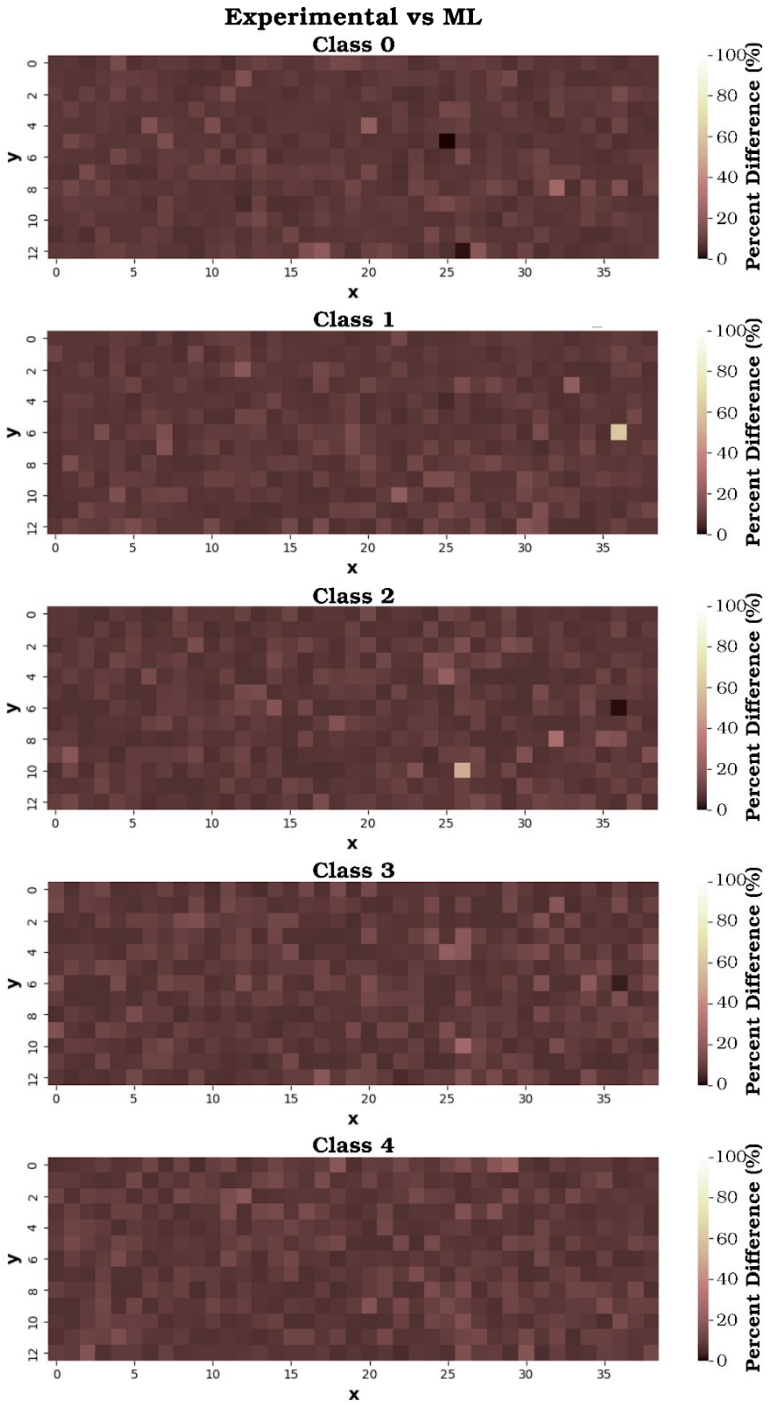


Figure S60. Comparison of the experimental and machine learning orpiment defocusing steps heatmaps. Areas that agree between the two heatmaps are represented by dark red, and areas that differ are white.

Bibliography

1 P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. Pietro Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G.-L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko and Y. Vázquez-Baeza, *Nat. Methods*, 2020, 17, 261–272.

2 J. Li, *PLoS One*, 2024, 19, e0316019.

3 G. R. Kumar, G. Thippanna and K. P. Kumar, *J. Comput. Sci. Syst. Softw.*, 2024, 1, 13–17.

4 L. Burgio, R. J. H. Clark and S. Firth, *Analyst*, 2001, 126, 222–227.

5 Y. Li, J. Ma, K. He and F. Wang, *Materials (Basel).*, 2024, 17, 770.

6 A. Amat, F. Rosi, C. Miliani, P. Sassi, M. Paolantoni and S. Fantacci, *J. Cult. Herit.*, 2020, 46, 374–381.

7 J. P. Vigouroux, E. Husson, G. Calvarin and N. Q. Dao, *Spectrochim. Acta Part A Mol. Spectrosc.*, 1982, 38, 393–398.

8 S. Bruni, F. Cariati, F. Casadio and L. Toniolo, *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, 1999, 55, 1371–1377.

9 H. G. M. Edwards, D. W. Farwell, E. M. Newton and F. R. Perez, *Analyst*, 1999, 124, 1323–1326.

10 M. Vagnini, R. Vivani, A. Sgamellotti and C. Miliani, *J. Raman Spectrosc.*, 2020, 51, 1118–1126.

11 M. Vagnini, R. Vivani, E. Viscuso, M. Favazza, B. G. Brunetti, A. Sgamellotti and C. Miliani, *Vib. Spectrosc.*, 2018, 98, 41–49.

12 R. L. Frost, W. N. Martens and J. T. Kloprogge, *Neues Jahrb. für Mineral. - Monatshefte*, 2002, 2002, 469–480.

13 R. L. Frost, J. T. Kloprogge and P. A. Williams, *Neues Jahrb. für Mineral. - Monatshefte*, 2003, 2003, 529–542.

14 C. G. Ryan, E. Clayton, W. L. Griffin, S. H. Sie and D. R. Cousens, *Nucl. Instruments Methods Phys. Res. Sect. B Beam Interact. with Mater. Atoms*, 1988, 34, 396–402.