Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Explainable machine learning for infrared spectral analysis of bone

Komal Wavhal
Stevens Institute of Technology
Christopher Blanc
Stevens Institute of Technology
William Querido
wquerido@stevens.edu

Stevens Institute of Technology

Article

Keywords: Bone, Infrared spectroscopy, O-PTIR, Machine learning, Explainable machine learning

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11445581/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Explainable machine learning for infrared spectral analysis of bone

Komal Wavhal[1,2], Christopher Blanc[1], William Querido[1]*

[1] Department of Biomedical Engineering, Stevens Institute of Technology, Hoboken, NJ, USA
[2] Department of Computer Science, Stevens Institute of Technology, Hoboken, NJ, USA
* Corresponding author: wquerido@stevens.edu

Abstract

Infrared spectroscopy provides chemically specific information on bone composition, but conventional analysis often reduces spectra to predefined peak-ratio parameters. Here, we evaluated whether multivariate full-spectrum analysis improves bone phenotype classification. Optical photothermal infrared (O-PTIR) spectroscopy was used as the experimental IR modality, with wild-type (WT) and osteogenesis imperfecta mouse (OIM) bones representing typical and fragile bone phenotypes, respectively. Conventional parameters included mineral content, carbonate content, HPO_4 content, mineral crystallinity, and collagen integrity. Classification using a combination of these parameters achieved 59–63% accuracy across partial least squares discriminant analysis (PLS-DA) and machine-learning classifiers, whereas models using the 1800–800 cm^{-1} fingerprint region reached 87–90%. Models based on the collagen-rich region performed similarly to the full spectrum and better than the mineral-rich region. Similar performance across PLS-DA, logistic regression, and XGBoost showed that classification gains were driven primarily by spectral representation rather than model complexity. Variable importance in projection (VIP) and spectral-zone-based SHapley Additive exPlanations (SHAP) identified prominent contributions from collagen-associated Amide I and II regions, with broader patterns of importance in PLS-DA and logistic regression and more localized attribution in XGBoost. Together, these findings support interpretable multivariate full-spectrum analysis as a complementary approach to conventional peak-ratio analysis of bone IR spectra.

Keywords:

Bone; Infrared spectroscopy; O-PTIR; Machine learning; Explainable machine learning

Introduction

Bone is a hierarchical composite tissue whose mechanical integrity depends not only on the amount of bone present, but also on the composition and organization of its mineralized extracellular matrix[1]. At the tissue level, type I collagen and apatite mineral form the fundamental structural components of bone, and their relative amounts, intrinsic properties, spatial distribution, and interactions contribute to bone quality and strength[2, 3]. Alterations in mineral content and structure, carbonate substitution, collagen organization, and matrix–mineral interactions have been associated with aging, disease, therapeutic intervention, and increased skeletal fragility[4-6].

Infrared (IR) spectroscopy provides a direct means to probe these compositional features by measuring molecular vibrations associated with the organic and mineral phases of bone[7-10]. In particular, Fourier transform infrared (FTIR) spectroscopy and imaging have been widely used to assess bone composition through characteristic bands arising from collagen-associated amide (C=O, NH, CN) and CH vibrations and mineral-associated PO_4, HPO_4, and CO_3 vibrations[7-10]. More recently, optical photothermal infrared (O-PTIR) spectroscopy has emerged as a complementary IR modality that combines mid-infrared excitation with visible-light detection, enabling measurements at submicron spatial resolution[11]. In bone, O-PTIR has been applied to both thick unembedded and resin-embedded samples to assess tissue composition at approximately 500 nm spatial resolution[12-14], extending infrared compositional analysis to length scales not readily accessible with conventional FTIR imaging.

Traditional IR spectroscopic analysis of bone has largely relied on predefined spectral parameters derived from the intensities or areas of selected absorbance bands, often after spectral preprocessing such as second-derivative transformation to better resolve overlapping features[7-10]. These parameters are commonly expressed as peak ratios that provide direct biochemical interpretation; for example, mineral content (or mineral-to-matrix ratio) can be estimated by calculating the ratio of the intensity of the PO_4 band to that of the amide I band. Related metrics have been widely used to assess mineral crystallinity, carbonate content, HPO_4 content (inversely associated with mineral maturity), and collagen integrity. However, an IR spectrum contains hundreds of highly correlated variables and overlapping signals distributed across broad spectral regions, such as within the 1800–800 cm^{-1} spectral fingerprint range. Reducing this information to a small number of predefined parameters therefore provides only a partial representation of the underlying spectral signatures and may overlook coordinated variations across features reflecting collagen and mineral composition.

This analytical limitation has motivated the use of multivariate chemometric approaches that consider spectral variables simultaneously rather than independently. Principal component analysis (PCA) is commonly used to summarize dominant sources of spectral variance, while partial least squares discriminant analysis (PLS-DA) incorporates class information to identify latent combinations of correlated variables that contribute to group separation[15, 16]. In IR spectroscopy, these approaches are particularly useful because neighboring wavenumbers are highly collinear and often reflect overlapping contributions from the same biochemical constituents. More recently, machine-learning methods have provided additional strategies for classification of high-dimensional spectral data, including linear models such as logistic regression, kernel-based methods such as support vector machines (SVM), instance-based methods such as k-nearest neighbors (KNN), and tree-based ensemble approaches such as random forest and XGBoost[17-19]. Together, these approaches provide complementary ways to model complex spectral patterns and assess whether biologically relevant differences are captured consistently across analytical frameworks.

Interpretability is particularly important when multivariate and machine-learning models are applied to spectroscopic data, because classification performance alone does not reveal which chemical features contribute to model predictions. In PLS-DA, variable importance in projection (VIP) scores provide a means of identifying spectral variables that contribute most strongly to discrimination[20], while explainable machine-learning approaches such as SHapley Additive exPlanations (SHAP) can quantify the contribution of individual features to model predictions[21]. However, IR spectra are composed of highly correlated neighboring wavenumbers that collectively form chemically meaningful bands, creating challenges for feature-by-feature attribution. Conventional SHAP analysis may therefore produce fragmented or even opposing contributions among adjacent wavenumbers, complicating chemical interpretation[22]. Recently, spectral-zone-based SHAP has been proposed to address this limitation by evaluating groups of contiguous spectral features rather than isolated wavenumbers, allowing chemically defined spectral regions to be incorporated directly into model interpretation[22].

Despite the increasing use of multivariate chemometric and machine-learning approaches in vibrational spectroscopy, relatively few studies have applied these methods to mid-infrared spectral analysis of bone[13, 14, 23-28]. In our previous work[13], O-PTIR spectra from wild-type (WT, typical bone phenotype) and osteogenesis imperfecta mouse (OIM, fragile bone phenotype) bone were analyzed using PCA followed by SVM classification, demonstrating that bone phenotypes could be distinguished from their spectral profiles and that classification was stronger using the collagen-associated spectral range (accuracy of 86–90%) than the mineral-associated range (accuracy of 60–81%). These findings supported the presence of phenotype-associated information across the O-PTIR spectrum, but also raised broader questions regarding how spectral representation, preprocessing, and modeling strategy influence classification and which spectral features ultimately drive model predictions.

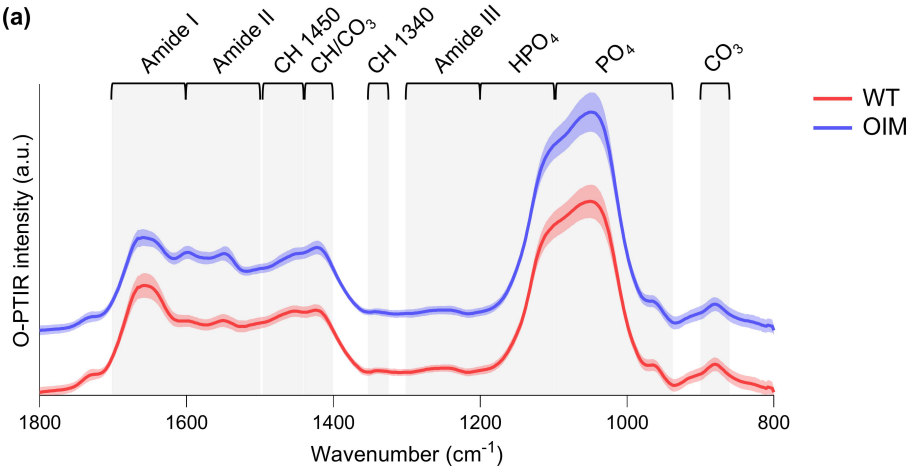
Building on this previous study[13], the present work systematically evaluates how spectral representation, preprocessing, and modeling strategy influence phenotype classification using the same O-PTIR dataset. Here, WT and OIM bones are used as experimentally defined models of typical and fragile bone phenotypes, respectively, rather than to investigate osteogenesis imperfecta as the primary biological focus. We compare conventional peak-ratio parameters with full and region-specific spectral representations, evaluate raw and second-derivative spectra, and benchmark PLS-DA against multiple machine-learning classifiers using animal-level cross-validation. We further examine the spectral basis of classification using VIP and spectral-zone-based SHAP analyses to relate model attribution to chemically defined regions of the bone spectrum. The overarching goal was to determine whether retaining and interrogating the broader IR spectral fingerprint reveals phenotype-associated compositional information that is not captured by conventional spectral parameters, while maintaining interpretation in terms of chemically meaningful features of the bone matrix.

Results

O-PTIR spectra and conventional compositional parameters

O-PTIR spectra of WT and OIM bones showed typical bands associated with collagen (amides, CH) and mineral (PO_4, HPO_4, CO_3) across the 1800–800 cm^{-1} spectral fingerprint range (Fig. 1a). The second-derivative spectra provided greater resolution of overlapping spectral features, with bands appearing as sharper, more distinct features compared with the raw spectra (Supplementary Fig. S1). The spectral zones defined in Fig. 1b were used throughout the subsequent multivariate analyses to relate model-derived spectral information to chemically grounded features of the bone matrix, including spectral regions associated predominantly with collagen (1800–1200 cm^{-1}) and mineral (1200–800 cm^{-1}) signatures.

(a)



(b)

Zone	Range (cm^{-1})	Chemical assignment
Amide I	1700–1600	C=O stretching
Amide II	1600–1500	N–H bending and C–N stretching
CH 1450	1490–1430	CH_2/CH_3 bending
CH/CO_3	1430–1400	Overlapping CH bending and carbonate vibration
CH 1340	1350–1330	CH wagging
Amide III	1300–1200	C–N stretching and N–H bending
HPO_4	1200–1100	Acid phosphate vibration
PO_4	1100–930	Phosphate ν_1/ν_3 stretching
CO_3	890–860	Carbonate ν_2 vibration
Unassigned	Gaps	Regions outside defined zones

Figure 1. O-PTIR spectra and spectral-zone definitions for WT and OIM bones. (a) Mean O-PTIR spectra of WT and OIM bones. (b) Wavenumber ranges and corresponding chemical assignments for each spectral zone used for subsequent multivariate and attribution analyses. Spectra are vertically staggered to facilitate visualization.

At the animal level, conventional quantification of peak ratios reflecting bone tissue composition showed that OIM bones presented higher mineral content and lower HPO_4 content than WT bones, with no significant differences in carbonate content, mineral crystallinity, or collagen integrity (Supplementary Fig. S2). When visualized across the balanced spectrum-level dataset (250 spectra per animal), all five conventional parameters showed substantial overlap between WT and OIM spectra (Fig. 2).

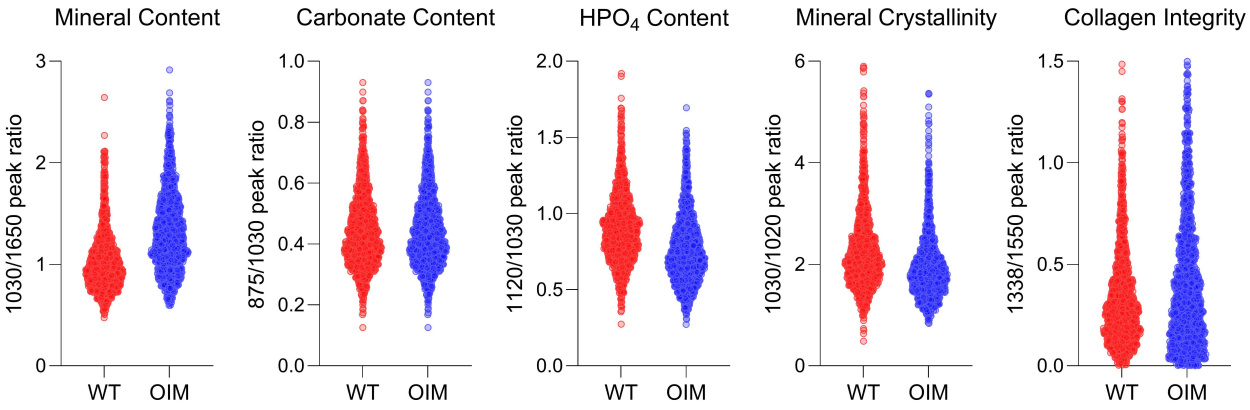


Figure 2. Conventional O-PTIR compositional parameters in WT and OIM bones. Peak-ratio parameters reflecting different bone tissue compositional properties across the balanced spectrum-level dataset (250 spectra per animal, 6 animals per group).

PLS-DA using conventional parameters and full spectra

PLS-DA based on the five conventional peak-ratio parameters showed limited discrimination, with an overall accuracy of 63% (Fig. 3). Mineral content and HPO_4 content were the variables that contributed most strongly to discrimination between WT and OIM spectra (VIP scores > 1). Extending the PLS-DA from the five peak-ratio parameters to the full raw O-PTIR spectral range increased classification accuracy to 87% (Fig. 4a,b). VIP analysis performed across the full spectrum showed that the spectral zones contributing most strongly to discrimination were within the collagen-associated Amide I and Amide II regions and the mineral-associated PO_4 region (Fig. 4c).

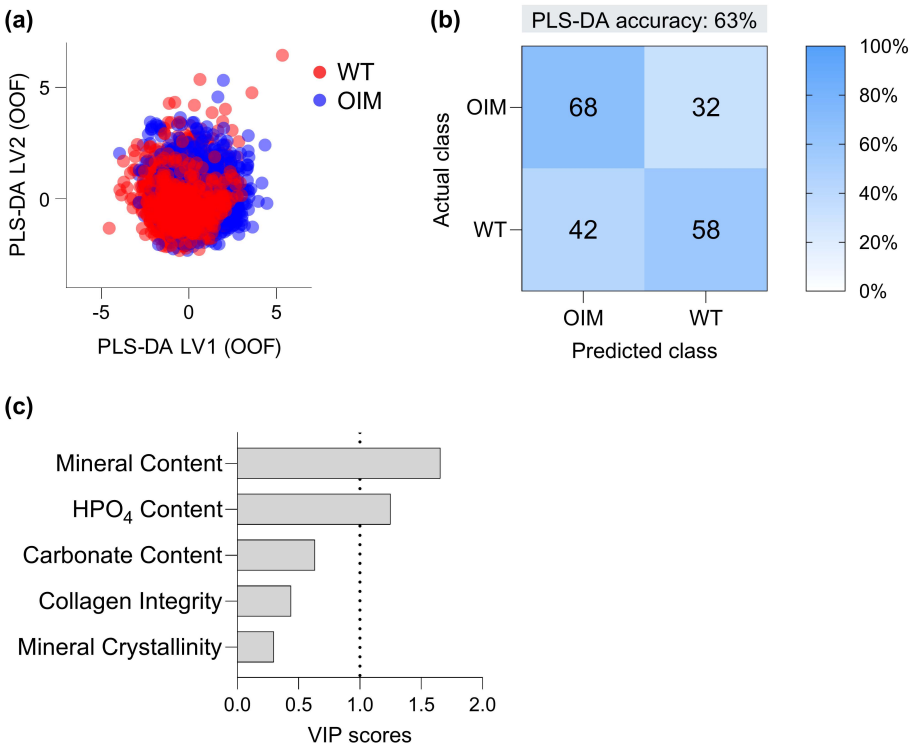


Figure 3. PLS-DA classification using conventional peak-ratio parameters. (a) Out-of-fold (OOF) PLS-DA scores, (b) confusion matrix, and (c) VIP scores for the five compositional parameters. VIP scores > 1 were considered indicative of important contributions to the model.

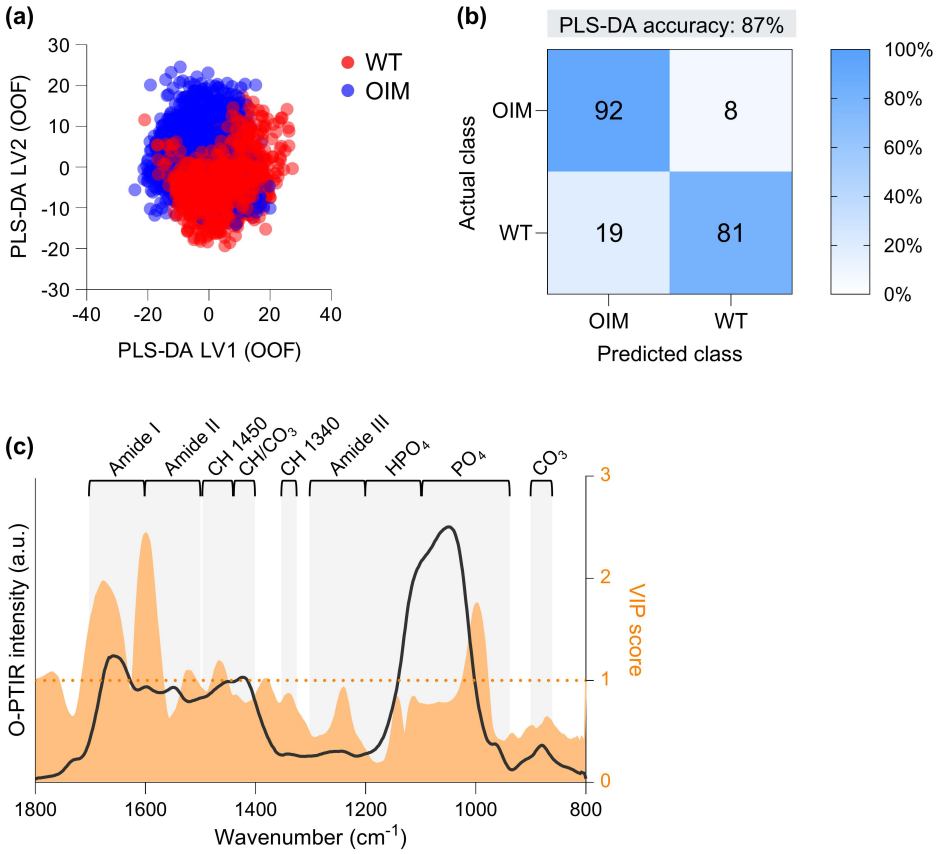


Figure 4. PLS-DA classification using the full O-PTIR spectrum. (a) OOF PLS-DA scores, (b) confusion matrix, and (c) VIP scores across the full spectral fingerprint range. VIP scores > 1 were considered indicative of important contributions to the model. Spectral zones are indicated to facilitate chemical interpretation of the VIP profile.

Classification performance across spectral representations and modeling approaches

Further PLS-DA analyses compared the full fingerprint region with spectral regions associated predominantly with collagen and mineral signatures, using both raw and second-derivative spectra (Fig. 5a). The full raw spectrum provided the highest accuracy, while the collagen-rich region showed similar performance (85%) and the mineral-rich region showed lower accuracy (75%). The same regional performance pattern was observed after second-derivative preprocessing, with no overall improvement in classification performance.

Building on the PLS-DA analysis, we next evaluated whether alternative machine-learning approaches could improve classification of WT and OIM spectra. Models representing different classification strategies were examined, including linear (logistic regression), kernel-based (SVM), instance-based (KNN), and tree-based ensemble approaches (random forest and XGBoost), allowing us to assess whether increasingly flexible modeling strategies provided additional discriminatory performance beyond PLS-DA (Fig. 5b). Overall classification performance was similar across the different machine-learning approaches, although a few differences were noted. For example, when considering the raw full fingerprint region, logistic regression showed the highest accuracy (90%), whereas KNN showed the lowest (82%). Importantly, the regional and preprocessing trends observed with PLS-DA were maintained across models, with the full and collagen-rich spectral ranges outperforming the mineral-rich

range and second-derivative preprocessing providing no consistent improvement. As observed with PLS-DA, classification based on the conventional peak-ratio parameters produced the lowest accuracy across all machine-learning models (59–62%). The same overall pattern was observed for F1-score and ROC-AUC (Supplementary Fig. S3).

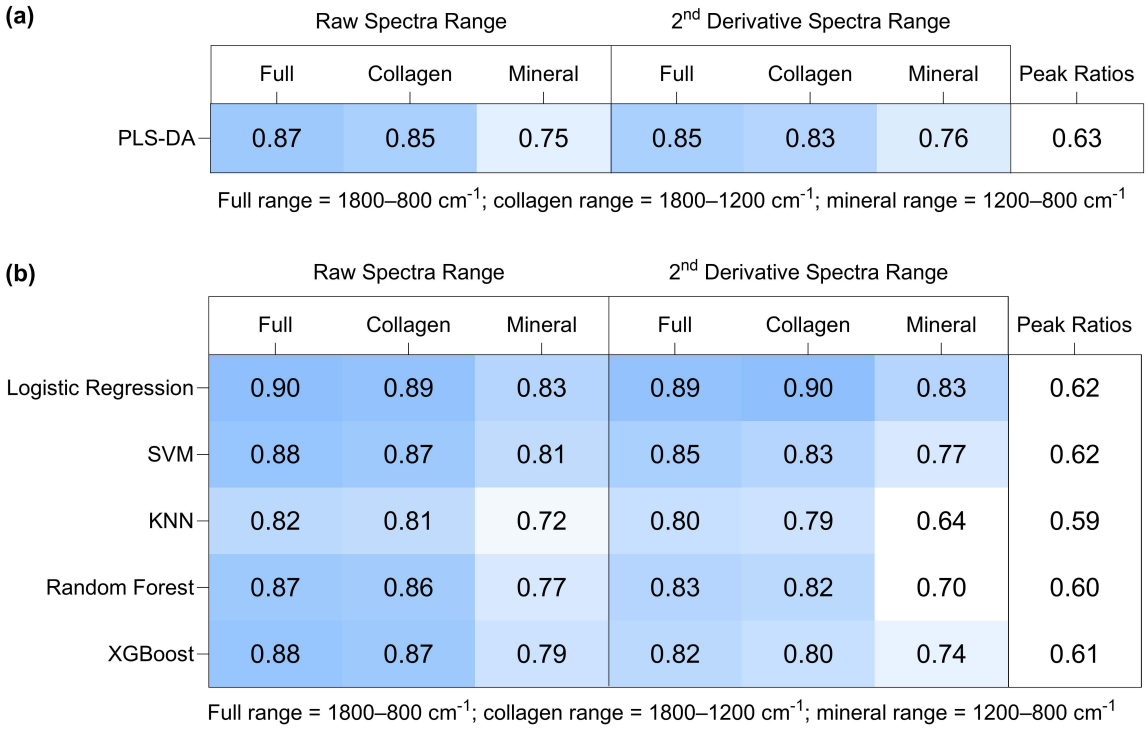


Figure 5. Classification accuracy across spectral representations and modeling approaches. Overall accuracy for (a) PLS-DA and (b) alternative machine-learning models using different spectral ranges (full fingerprint, collagen-rich, mineral-rich) and conventional peak-ratio representations. Raw and second-derivative spectra are compared for each spectral-range analysis.

Explainable spectral attribution across distinct machine-learning models

To further investigate the spectral basis of classification, logistic regression and XGBoost were selected as high-performing models representing distinct linear and tree-based ensemble strategies, respectively, and SHAP analysis was applied to models using the raw full fingerprint spectrum (Fig. 6). The SHAP profiles broadly reinforced the PLS-DA VIP results, with prominent contributions within the collagen-associated Amide I and Amide II regions and additional contributions within the mineral-associated PO_4 region. However, the distribution of feature importance differed between models: logistic regression showed a broader distribution of contributions across the spectrum, whereas XGBoost showed a more localized attribution pattern concentrated at a smaller number of wavenumbers (Fig. 6b,d). The XGBoost SHAP profile highlighted the same major regions identified by PLS-DA and logistic regression, including Amide I, Amide II, and PO_4, while its greater localization also revealed a distinct contribution within the Amide III region that was less apparent with the other approaches.

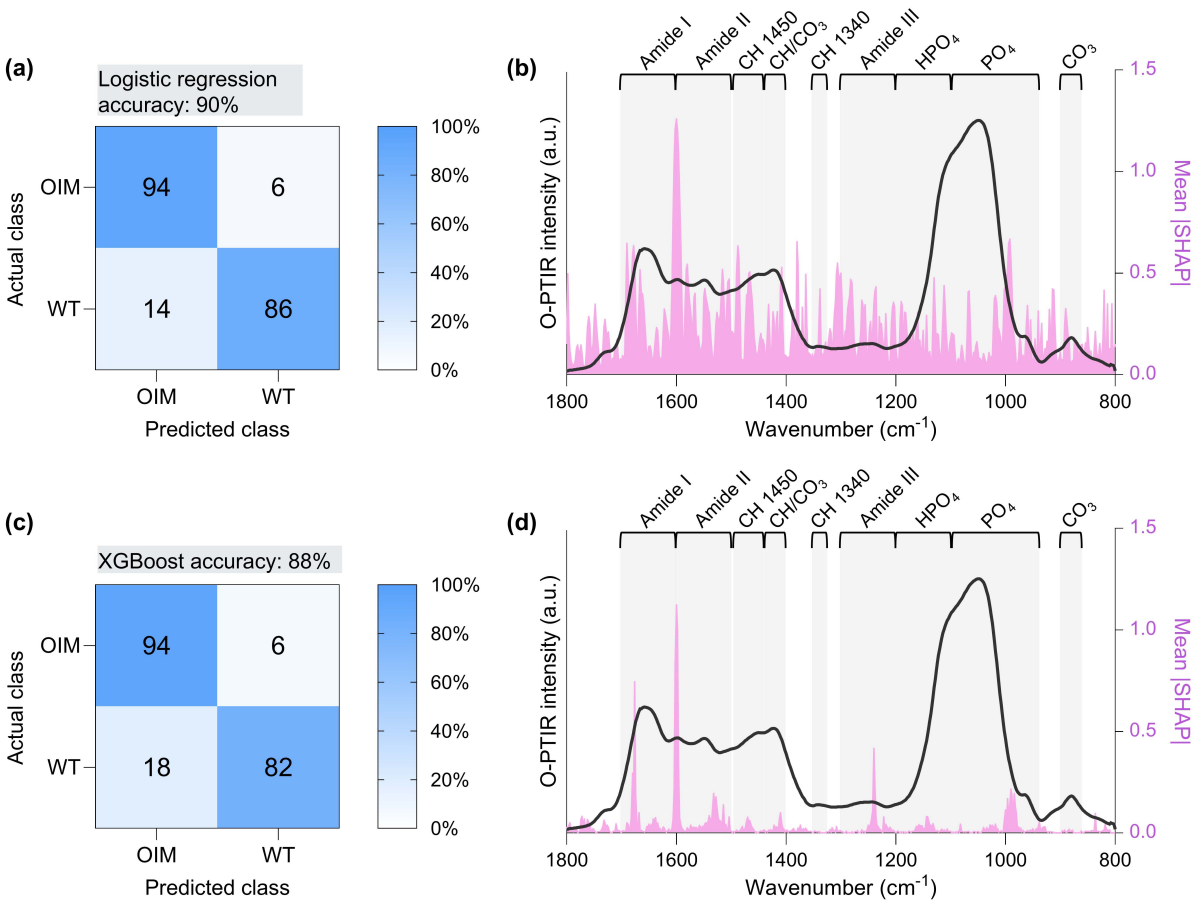


Figure 6. Explainable spectral attribution for logistic regression and XGBoost models. Confusion matrices for (a) logistic regression and (c) XGBoost, with corresponding mean absolute SHAP values across the full spectral fingerprint range for (b) logistic regression and (d) XGBoost. Spectral zones are indicated to facilitate chemical interpretation of the attribution profiles.

To extend the SHAP interpretation from individual wavenumbers to chemically defined regions of the bone spectrum, feature attribution was evaluated at the level of the spectral zones defined above (Fig. 7). Spectral-zone-based SHAP analysis again identified Amide I and Amide II as the strongest contributors in both logistic regression and XGBoost, followed by the PO_4 region. Beyond these leading regions, the lower-ranked contributions differed between models; for example, unassigned spectral gaps contributed more prominently in logistic regression, whereas XGBoost showed a clearer contribution from the Amide III region. Importantly, the zone-level analysis also showed how each spectral region influenced classification direction, with Amide I contributing in opposite directions for WT and OIM spectra and showing particularly clear class separation.

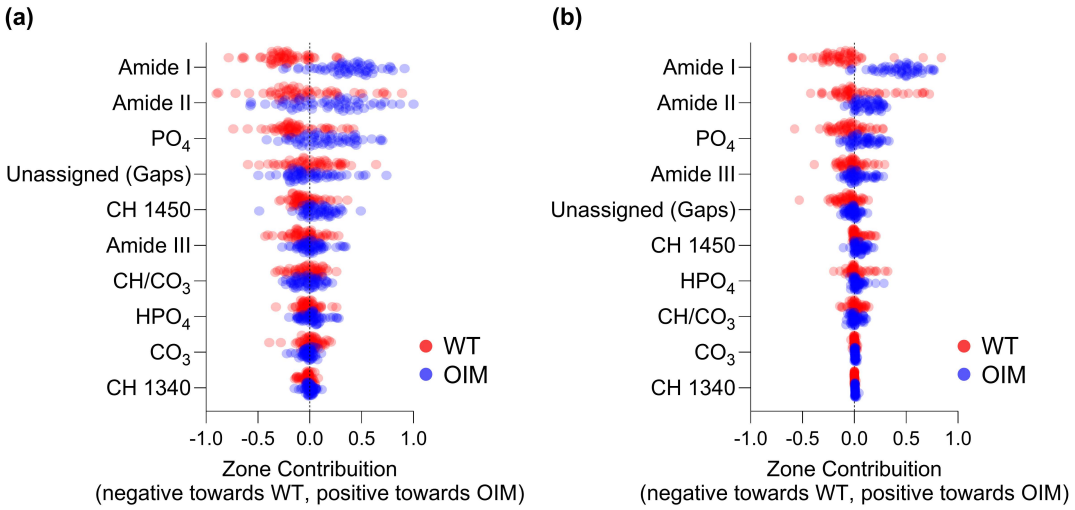


Figure 7. Spectral-zone-based SHAP beeswarm plot for logistic regression and XGBoost models. SHAP contributions for chemically defined spectral zones in (a) logistic regression and (b) XGBoost models. Positive and negative SHAP values indicate contributions toward OIM and WT classification, respectively.

Discussion

The present study extends our previous work demonstrating that O-PTIR spectral data can distinguish WT and OIM bone phenotypes[13] by making three main contributions. First, we show that classification depends on how the spectrum is represented: the full and collagen-rich spectral ranges consistently outperform the mineral-rich range and conventional peak-ratio parameters across PLS-DA and multiple machine-learning classifiers. Second, we show that classification performance is broadly similar across model families despite their different underlying assumptions, indicating that the primary gain arises from retaining richer multivariate spectral information rather than from increasing model complexity. Third, we establish an interpretable framework for these predictions by combining PLS-DA VIP and SHAP analyses, including spectral-zone-based attribution, to relate model outputs to chemically meaningful regions of the bone spectrum. Together, these findings expand the original PCA-SVM analysis into a broader, animal-level validated framework for determining both how O-PTIR spectra distinguish bone phenotypes and which spectral features drive that discrimination.

In this study, WT and OIM bones were used as experimentally defined models of typical and fragile bone phenotypes rather than as the primary biological focus of the study. The OIM model provides a well-characterized contrast in bone matrix composition and mechanical integrity[29-31], making it a useful test system for evaluating whether IR spectral analysis can detect and interpret phenotype-associated compositional differences. Accordingly, the objective was not to identify disease-specific biomarkers of osteogenesis imperfecta, but to use this biological contrast to examine how different representations and analytical approaches capture compositional information encoded within bone IR spectra.

Conventional peak-ratio parameters remain valuable because they provide direct biochemical readouts of selected aspects of bone composition[7-10], but their interpretability is achieved through substantial dimensional reduction. Each ratio collapses a broader spectral region into a single summary metric, which is useful for testing specific compositional hypotheses but necessarily excludes information contained in band shape, shoulders, neighboring features, and covariance among spectral regions. The lower discriminatory performance of the peak-ratio representation therefore supports the idea that phenotype-associated differences are distributed across the spectrum rather than confined to a small set of predefined markers. This distinction is important because multivariate spectral analysis should not be viewed as replacing conventional compositional metrics, but rather as complementing them by retaining additional phenotype-associated information that those metrics are not designed to capture.

The stronger performance of the full and collagen-rich spectral representations further indicates that the discriminatory information is distributed across broader regions of the spectrum rather than concentrated within isolated compositional markers. Notably, the collagen-rich region consistently approached the performance of the full fingerprint range and outperformed the mineral-rich region across PLS-DA and machine-learning models. This pattern is consistent with our previous O-PTIR study using PCA and SVM[13], in which collagen-associated spectra provided stronger WT/OIM discrimination than mineral-associated spectra. This finding is consistent with the pathogenesis of osteogenesis imperfecta, in which abnormalities of type I collagen are central to the skeletal phenotype[29-33]. At the same time, the slightly higher performance of the full spectral range suggests that mineral-associated information also contributes to classification, supporting a distributed matrix-level signature rather than a purely collagen-specific effect.

Second-derivative preprocessing provided an additional test of whether enhancing resolution of overlapping spectral features would alter phenotype discrimination. As expected, the transformation produced sharper and more clearly resolved bands[34, 35], but classification

performance remained broadly similar to that obtained from the raw spectra, with no consistent advantage across spectral regions or modeling approaches. This distinction is useful because it shows that improved visual resolution of spectral features does not necessarily translate into a meaningful gain in classification performance. In the present dataset, both raw and second-derivative representations retained sufficient multivariate information to support discrimination, while preserving the same overall hierarchy among full, collagen-rich, and mineral-rich spectral ranges.

The relatively similar performance across PLS-DA and the different machine-learning classifiers suggests that increasingly flexible model architectures did not provide a major advantage for this dataset. Logistic regression, SVM, random forest, and XGBoost all achieved broadly comparable discrimination when supplied with the same spectral representations. These classifiers rely on different assumptions about the structure of the feature space[18, 19], yet produced similar performance in the present dataset. This finding suggests that the dominant phenotype-associated information is sufficiently strong and consistently represented across the spectra to be captured by both relatively simple and more flexible classifiers. From a methodological perspective, this is important because it argues against assuming that greater model complexity will necessarily yield better performance in vibrational spectroscopic classification. Instead, the choice of spectral representation appears to have a larger influence on performance than the specific classifier used, reinforcing the value of evaluating simpler and more complex models within the same analytical framework.

The interpretability analyses are important because they show that the classification can be related back to chemically meaningful regions of the bone spectrum rather than treated as a purely predictive exercise. Agreement between PLS-DA VIP and SHAP-based attribution supports the robustness of the identified spectral information across distinct analytical frameworks. Of particular interest, the spectral-zone-based SHAP approach provides a more spectroscopy-centered interpretation by grouping neighboring wavenumbers into chemically defined regions, reducing the fragmentation that can arise when highly correlated spectral variables are considered individually[22]. This makes the model output easier to relate to established bone matrix features and provides a direct way to compare the relative influence and direction of contributions from major spectral regions associated with collagen and mineral composition. To our knowledge, applying this type of spectral-zone-based explainability to bone IR spectroscopy represents a new extension of current approaches and may offer a useful framework for improving interpretability in future spectroscopic machine-learning studies.

Several limitations should be considered when interpreting the present findings. The study included a relatively small biological sample size, with six animals per group contributing 250 spectra per specimen. This creates a potential risk of overestimating performance if spectra from the same animal are allowed to contribute to both training and testing; accordingly, we used animal-level cross-validation so that all spectra from the held-out animal were excluded from model training. Even with this safeguard, the present results should be viewed as demonstrating analytical feasibility rather than establishing generalizable biomarkers of bone fragility. The WT and OIM groups represent one specific contrast between typical and fragile bone phenotypes, and additional studies will be needed to evaluate classification performance and identify chemically informative spectral features across other skeletal diseases, phenotypes, and sample conditions. Generalizability across infrared modalities also remains to be established. O-PTIR spectra of bone have been shown to exhibit broadly comparable spectral features and compositional parameters to conventional FTIR measurements[14], but modality-specific spectral differences remain. Further studies applying the same analytical framework to FTIR datasets will therefore be needed before extending these conclusions to bone IR spectroscopy more generally.

Overall, this study shows that retaining multivariate full-spectrum information provides a richer basis for bone phenotype classification than reducing spectra to a small number of conventional peak-ratio parameters. The consistency of this finding across PLS-DA and multiple machine-learning models indicates that spectral representation is more influential than increasing model complexity, while the strong performance of collagen-rich regions highlights the importance of matrix-associated information within the broader spectral fingerprint. At the same time, VIP and SHAP analyses demonstrate that these classifications can be interpreted in chemically meaningful terms rather than treated as black-box predictions. In particular, spectral-zone-based SHAP offers a useful bridge between multivariate modeling and conventional spectroscopic interpretation by expressing feature importance at the level of chemically defined spectral regions. More broadly, this framework may facilitate the characterization of compositional differences across other skeletal phenotypes by enabling both classification and chemically interpretable identification of the spectral features underlying those differences. Together, these findings support an interpretable, full-spectrum analytical framework for extracting compositional information from bone IR spectra that complements established peak-based approaches.

Methods

Bone samples and O-PTIR spectral dataset

The present study reanalyzed the O-PTIR spectral dataset originally reported by Dev et al.[13] Bones with a typical phenotype were obtained from wild-type (WT) mice, whereas bones with a fragile phenotype were obtained from homozygous osteogenesis imperfecta mice (OIM; *oim/oim*), a murine model characterized by defective type I collagen formation and increased skeletal fragility[32, 33]. Six animals per group were analyzed. Animals were female and 26 weeks old. O-PTIR spectra were acquired from cross-sectional segments at the tibial mid-diaphysis using a mIRage sub-micron infrared microscope (Photothermal Spectroscopy Corp.) equipped with a quantum cascade laser source and a 532 nm visible probe. Spectra were collected as linescans across the cortical bone from the endosteal to periosteal surface at 500 nm spacing, with three linescans acquired per sample. Full details of sample preparation and spectral acquisition have been described previously[13].

Ethics Declarations

The present study reanalyzed spectral data previously generated and reported by Dev et al.[13] and did not involve any new animal experiments. The bone specimens used to generate the original dataset were female mouse tibiae obtained as residual specimens from a separate study approved by the Institutional Animal Care and Use Committee (IACUC) at the Hospital for Special Surgery (HSS), New York, NY. All animal procedures associated with the original study were performed in accordance with relevant institutional guidelines and regulations.

Spectral preprocessing and conventional compositional parameters

Spectra were processed using Solo (Eigenvector Research Inc.) as described previously[13]. Briefly, spectra with substantial PMMA contamination were excluded, and outliers were identified and removed using the interquartile range (IQR) method. Spectra were then processed using a second-derivative transformation, and the inverted second-derivative intensities of selected bands were used to calculate established compositional parameters. These included mineral content (1030/1650 cm^{-1}), carbonate content (875/1030 cm^{-1}), HPO_4 content (1120/1030 cm^{-1}; inverse to mineral maturity), mineral crystallinity (1030/1020 cm^{-1}), and collagen integrity (1338/1550 cm^{-1}). For animal-level comparisons, spectral values were averaged within each animal to obtain one value per animal (n = 6 animals per group). WT and OIM groups were compared using the Mann–Whitney test in Prism (GraphPad), with statistical significance defined as $p < 0.05$.

Spectral representations and spectral-zone definitions

For multivariate analysis, spectra were evaluated using three spectral representations: the full fingerprint region (1800–800 cm-1), a collagen-rich region (1800–1200 cm-1), and a mineral-rich region (1200–800 cm-1). Analyses were performed using both raw and second-derivative spectra to assess the influence of spectral preprocessing on classification performance. To facilitate interpretation of the multivariate analyses, the fingerprint region was further divided into chemically defined spectral zones based on the established vibrational assignments (Fig. 1).

Data structure and analytical platform

For the present reanalysis, spectral data from six WT and six OIM animals were analyzed using a custom interactive analytical platform developed in Python using Streamlit. Spectral data were organized by wavenumber in raw and second-derivative form, together with the calculated peak-ratio parameters. Each spectrum or set of peak ratios was associated with its corresponding animal ID and phenotype to enable animal-level grouping during model validation. The platform enabled class-balanced sampling by selecting 250 spectra per animal, yielding 3,000 spectra in

total, while retaining the animal identity and phenotype associated with each spectrum. The platform integrated a range of functions, including data import and visualization, spectral-range selection, multivariate analysis, supervised machine-learning classification, model performance evaluation, and spectral-zone-based SHAP interpretation within a common workflow. The implementation used standard Python scientific-computing and machine-learning libraries, including pandas, NumPy, scikit-learn, XGBoost, and SHAP. The analytical platform is publicly available for interactive use at https://bonespec-ai.streamlit.app/.

Chemometrics and machine-learning classification

Partial least squares discriminant analysis (PLS-DA) and supervised machine-learning classification were performed within the same analytical workflow and animal-level validation framework. Machine-learning classifiers included logistic regression (LR), support vector machine (SVM), k-nearest neighbors (KNN), random forest (RF), and XGBoost (XGB). Each model was evaluated across the spectral representations and preprocessing conditions described above. Because 250 spectra were used from each animal, spectra were not treated as independent biological replicates during model validation. Animal-level grouping was therefore maintained throughout cross-validation to avoid pseudoreplication and information leakage between training and validation sets. PLS-DA models were constructed using three latent variables (LVs). For logistic regression, per-fold feature scaling was performed using StandardScaler fitted to the training animals and subsequently applied to the held-out animal. Model performance was evaluated using leave-one-group-out (LOGO) cross-validation, with animal ID used as the grouping variable. For each fold, spectra from the held-out animal were completely excluded from model fitting, ensuring that spectra from the same animal were not represented in both training and validation data. Predictions generated for the held-out spectra in each fold were subsequently combined to obtain out-of-fold (OOF) predictions across the complete dataset. Classification accuracy was calculated as the proportion of correctly classified spectra among all OOF predictions. Confusion matrices were generated from the pooled OOF predictions and row-normalized to show the proportion of spectra assigned to each predicted class within each true phenotype. For the machine-learning classifiers, F1 score and receiver operating characteristic area under the curve (ROC-AUC) were additionally calculated as complementary measures of classification performance.

Model interpretation using VIP and spectral-zone-based SHAP

From PLS-DA models, variable importance in projection (VIP) scores were used to identify spectral variables contributing most strongly to class discrimination. For logistic regression and XGBoost, SHapley Additive exPlanations (SHAP) were used to quantify model-based feature attribution. To account for the strong correlation among neighboring spectral variables and to provide chemically interpretable explanations, SHAP analysis was performed using a spectral-zone perturbation approach based on Contreras et al.[22] In this approach, each predefined spectral zone was treated as a single feature, and all wavenumbers within a zone were perturbed together using the corresponding values from the training-set mean spectrum. SHAP analysis was performed within the animal-level validation framework, such that explanations were generated from models that had not been trained on the corresponding held-out animal. Ten spectra per animal were analyzed, yielding 120 explained spectra across the 12 animals. SHAP beeswarm plots were used to visualize the distribution and direction of spectral-zone contributions across the explained spectra.

Code availability

Source code is available through GitHub at https://github.com/wavhalkomal/bonespec-ai.

Acknowledgements

This research was supported by NIH/NIAMS award R21AR082129.

Author Contributions

K.W. performed the computational analyses, developed the analytical platform, prepared data visualizations, and wrote and revised the manuscript. C.B. contributed to computational analyses and revised the manuscript. W.Q. conceived and supervised the study, contributed to data analysis and interpretation, and wrote and revised the manuscript. All authors reviewed and approved the final manuscript.

Competing Interests

The authors declare no competing interests.

Data Availability Statement

The O-PTIR spectral dataset analyzed in the present study was previously generated and reported by Dev et al.[13]. The data supporting the findings of the current reanalysis are available from the corresponding author upon reasonable request.

References

1. Reznikov, N., R. Shahar, and S. Weiner, *Bone hierarchical structure in three dimensions.* Acta Biomaterialia, 2014. **10**(9): p. 3815-3826.
2. Unal, M., A. Creecy, and J.S. Nyman, *The Role of Matrix Composition in the Mechanical Behavior of Bone.* Curr Osteoporos Rep, 2018. **16**(3): p. 205-215.
3. Hunt, H.B. and E. Donnelly, *Bone quality assessment techniques: geometric, compositional, and mechanical characterization from macroscale to nanoscale.* Clin Rev Bone Miner Metab, 2016. **14**(3): p. 133-149.
4. Donnelly, E., et al., *Contribution of Mineral to Bone Structural Behavior and Tissue Mechanical Properties.* Calcified Tissue International, 2010. **87**(5): p. 450-460.
5. Boskey, A.L. and L. Imbert, *Bone quality changes associated with aging and disease: a review.* Ann N Y Acad Sci, 2017. **1410**(1): p. 93-106.
6. Wang, Z.X., et al., *Altered distributions of bone tissue mineral and collagen properties in women with fragility fractures.* Bone, 2016. **84**: p. 237-244.
7. Taylor, E.A. and E. Donnelly, *Raman and Fourier transform infrared imaging for characterization of bone material properties.* Bone, 2020. **139**: p. 115490.
8. Paschalis, E.P., S. Gamsjaeger, and K. Klaushofer, *Vibrational spectroscopic techniques to assess bone quality.* Osteoporos. Int., 2017. **28**(8): p. 2275-2291.
9. Querido, W., S. Kandel, and N. Pleshko, *Applications of Vibrational Spectroscopy for Analysis of Connective Tissues.* Molecules, 2021. **26**(4).
10. Boskey, A. and N.P. Camacho, *FT-IR imaging of native and tissue-engineered bone and cartilage.* Biomaterials, 2007. **28**(15): p. 2465-2478.
11. Prater, C.B., M. Kansiz, and J.-X. Cheng, *A tutorial on optical photothermal infrared (O-PTIR) microscopy.* APL Photonics, 2024. **9**(9): p. 091101.
12. Ahn, T., et al., *Matrix/mineral ratio and domain size variation with bone tissue age: A photothermal infrared study.* J Struct Biol, 2022. **214**(3): p. 107878.
13. Dev, I., et al., *Assessment of submicron bone tissue composition in plastic-embedded samples using optical photothermal infrared (O-PTIR) spectral imaging and machine learning.* J Struct Biol X, 2024. **10**: p. 100111.
14. Reiner, E., et al., *Application of Optical Photothermal Infrared (O-PTIR) Spectroscopy for Assessment of Bone Composition at the Submicron Scale.* Appl Spectrosc, 2023. **77**(11): p. 1311-1324.
15. Rosipal, R. and N. Krämer. *Overview and Recent Advances in Partial Least Squares*. in *Subspace, Latent Structure and Feature Selection*. 2006. Berlin, Heidelberg: Springer Berlin Heidelberg.
16. Wold, S., M. Sjöström, and L. Eriksson, *PLS-regression: a basic tool of chemometrics.* Chemometrics and Intelligent Laboratory Systems, 2001. **58**(2): p. 109-130.
17. Westermayr, J. and P. Marquetand, *Machine learning spectroscopy to advance computation and analysis.* Chemical Science, 2025. **16**(46): p. 21660-21676.
18. Egbewole, B.I., et al., *Machine learning–enabled vibrational spectroscopy for translational bioclinical applications.* Microchemical Journal, 2026. **230**: p. 119784.
19. Yang, J., et al., *Deep learning for vibrational spectral analysis: Recent progress and a practical guide.* Analytica Chimica Acta, 2019. **1081**: p. 6-17.
20. Mehmood, T., S. Sæbø, and K.H. Liland, *Comparison of variable selection methods in partial least squares regression.* Journal of Chemometrics, 2020. **34**(6): p. e3226.
21. Lundberg, S.M. and S.-I. Lee, *A unified approach to interpreting model predictions*, in *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017, Curran Associates Inc.: Long Beach, California, USA. p. 4768–4777.

22. Contreras, J., et al., *Spectral Zones-Based SHAP/LIME: Enhancing Interpretability in Spectral Deep Learning Models Through Grouped Feature Analysis.* Anal Chem, 2024. **96**(39): p. 15588-15597.
23. Chaber, R., et al., *Distinguishing Ewing sarcoma and osteomyelitis using FTIR spectroscopy.* Sci Rep, 2018. **8**(1): p. 15081.
24. Wang, Q., et al., *Human and non-human bone identification using FTIR spectroscopy.* Int J Legal Med, 2019. **133**(1): p. 269-276.
25. Yu, K., et al., *Identification of antemortem and postmortem fractures in a complex environment by FTIR spectroscopy based on a rabbit tibial fracture self-control model.* Int J Legal Med, 2021. **135**(6): p. 2385-2394.
26. Yu, K., et al., *Identification of antemortem, perimortem and postmortem fractures by FTIR spectroscopy based on a rabbit tibial fracture model.* Spectrochim Acta A Mol Biomol Spectrosc, 2020. **239**: p. 118535.
27. Chowdhury, M.P., et al., *Machine learning ATR-FTIR spectroscopy data for the screening of collagen for ZooMS analysis and mtDNA in archaeological bone.* Journal of Archaeological Science, 2021. **126**: p. 105311.
28. Bayarı, S.H., et al., *Application of ATR-FTIR spectroscopy and chemometrics for the discrimination of human bone remains from different archaeological sites in Turkey.* Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2020. **237**: p. 118311.
29. Bi, X., et al., *Correlations Between Bone Mechanical Properties and Bone Composition Parameters in Mouse Models of Dominant and Recessive Osteogenesis Imperfecta and the Response to Anti-TGF-β Treatment.* J Bone Miner Res, 2017. **32**(2): p. 347-359.
30. Bart, Z.R., M.A. Hammond, and J.M. Wallace, *Multi-scale analysis of bone chemistry, morphology and mechanics in the oim model of osteogenesis imperfecta.* Connect Tissue Res, 2014. **55 Suppl 1**: p. 4-8.
31. Camacho, N.P., et al., *The material basis for reduced mechanical properties in oim mice bones.* J Bone Miner Res, 1999. **14**(2): p. 264-72.
32. Alcorta-Sevillano, N., et al. *Murine Animal Models in Osteogenesis Imperfecta: The Quest for Improving the Quality of Life*. International Journal of Molecular Sciences, 2023. **24**, DOI: 10.3390/ijms24010184.
33. Chipman, S.D., et al., *Defective pro alpha 2(I) collagen synthesis in a recessive mutation in mice: a model of human osteogenesis imperfecta.* Proceedings of the National Academy of Sciences, 1993. **90**(5): p. 1701-1705.
34. Rieppo, L., et al., *Application of second derivative spectroscopy for increasing molecular specificity of fourier transform infrared spectroscopic imaging of articular cartilage.* Osteoarthritis and Cartilage, 2012. **20**(5): p. 451-459.
35. Rinnan, Å., F.v.d. Berg, and S.B. Engelsen, *Review of the most common pre-processing techniques for near-infrared spectra.* TrAC Trends in Analytical Chemistry, 2009. **28**(10): p. 1201-1222.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- WavhaletalSupplementalMaterial102026.pdf