

Experimental FTIR and Raman Vibrational Analysis of 4-Amino-2-Chlorobenzonitrile: Cross-Technique Comparison and a Framework for Computational and Machine-Learning Validation

Dr. Surabhi Singhal¹

¹Assistant Professor Physics; Government Girls Degree College Kharkhauda Meerut

Publication Date: 2026/10/03

Abstract: This study organizes the recorded Fourier transform infrared and laser Raman bands of 4-amino-2-chlorobenzonitrile into an experimental benchmark for computational vibrational analysis and interpretable machine learning. The available measurements comprise peak assignments obtained in potassium bromide, Nujol and Raman conditions rather than complete digitized intensity arrays. Accordingly, the work does not report invented density-functional-theory frequencies or classification accuracy. It instead quantifies agreement among the recorded channels, identifies chemically diagnostic regions and defines a reproducible pathway for later simulation and model validation. The nitrile stretching mode is observed at 2235 cm⁻¹ in KBr, 2237 cm⁻¹ in Nujol and 2232 cm⁻¹ in Raman spectra. Amino-related evidence includes associated N–H stretching at 3222–3280 cm⁻¹, C–NH₂ stretching at 1346 cm⁻¹ in KBr and 1336 cm⁻¹ in Raman data, and lower-frequency amino deformations. Chlorine-sensitive modes occur at 738 and 365 cm⁻¹ in the Raman spectrum. Nineteen matched KBr–Nujol modes yield a mean absolute difference of 10.79 cm⁻¹ and a root-mean-square difference of 17.04 cm⁻¹. Nine matched KBr–Raman modes yield corresponding values of 17.67 and 26.67 cm⁻¹. A proposed workflow combines scaled quantum-chemical frequencies, mode-displacement verification, peak-level baselines and dual-branch one-dimensional convolutional learning. The framework emphasizes grouped validation, external testing and chemical inspection of model attention. The resulting paper establishes what is experimentally supported, what can be calculated and what requires new spectral acquisition.

Keywords: 4-Amino-2-Chlorobenzonitrile; FTIR Spectroscopy; Laser Raman Spectroscopy; Vibrational Assignment; Density Functional Theory; Machine Learning; Multimodal Spectral Fusion.

How to Cite: Dr. Surabhi Singhal (2026) Experimental FTIR and Raman Vibrational Analysis of 4-Amino-2-Chlorobenzonitrile: Cross-Technique Comparison and a Framework for Computational and Machine-Learning Validation.

International Journal of Innovative Science and Research Technology, 11(9), 2565-2570.

<https://doi.org/10.38124/ijisrt/26sep1300>

I. INTRODUCTION

Vibrational spectroscopy connects molecular structure with experimentally observable frequencies. Infrared absorption responds to vibrations that change the molecular dipole moment, whereas Raman scattering responds to changes in molecular polarizability [1]–[4]. The two methods therefore provide complementary rather than identical views of molecular motion [1], [2]. Their combination is especially useful for substituted aromatic compounds, where nitrile, amino and halogen substituents modify the electron distribution of a common benzene-ring framework and produce overlapping, substitution-sensitive bands throughout the fingerprint region [5]–[6].

4-Amino-2-chlorobenzonitrile contains four diagnostically relevant structural elements: an aromatic ring, a nitrile group, an amino substituent and a chlorine substituent. The comparatively isolated C≡N stretching band is a useful marker for benzonitrile derivatives, whereas amino vibrations are influenced by association and hydrogen bonding [3]–[5]. Low-frequency C–Cl and skeletal vibrations provide additional evidence of aromatic substitution [4]–[6]. Reliable identification must therefore use a chemically consistent pattern of bands instead of depending on a single frequency.

Quantum-chemical calculations can predict harmonic normal-mode frequencies, infrared intensities, Raman activities and atomic displacement patterns [7]. Machine-

learning methods can integrate distributed spectral information across many wavenumbers and can support classification or structure prediction when they are trained on adequately designed datasets [8]-[10]. Neither approach should be treated as a substitute for measured evidence. A reproducible simulation requires a declared electronic-structure method, basis set, scaling procedure and mode-matching rule, while supervised learning requires multiple independent spectra, leakage-resistant partitioning and an external test set [6], [11], [12]. Peak assignments from one compound can support benchmarking and method design, but they cannot establish predictive accuracy by themselves.

The purpose of this paper is therefore to connect the available experimental values with a transparent simulation and machine-learning framework. Building on earlier experimental and theoretical studies of substituted benzonitrile vibrations [5], [6], the paper (i) consolidates the measured FTIR and Raman assignments, (ii) quantifies cross-medium and cross-technique frequency agreement, (iii) defines a defensible DFT comparison protocol, (iv) converts chemically meaningful bands into candidate model features, and (v) identifies the additional data required before a trained classifier can be reported.

II. LITERATURE REVIEW AND SCIENTIFIC BACKGROUND

➤ *Vibrational Evidence from Complementary Channels*

A normal mode may be prominent in one spectral channel and weak or absent in another because infrared and Raman activities arise from different molecular properties [1], [2], [4]. This complementarity has long supported the joint interpretation of substituted aromatic molecules [3], [4]. It is also visible in the present recorded data: the associated N–H band is reported in the infrared measurements, whereas the C–Cl stretch at 738 cm^{-1} and C–Cl bend at 365 cm^{-1} are recorded in Raman data. The nitrile stretch appears in all three recorded conditions and consequently provides a useful alignment marker. Comparable use of nitrile, ring and C–Cl modes for substituted benzonitriles has been reported by Singhal [5],[6].

➤ *Role of Computational Simulation*

A calculated harmonic spectrum supplies frequencies, infrared intensities, Raman activities and atomic displacement vectors. Molecular-orbital and density-functional frequency calculations have established the value of theoretical normal modes for supporting experimental assignments [6], [13], [14]. Direct comparison with experiment nevertheless requires care because the harmonic approximation, finite basis sets, temperature, intermolecular association and sample matrix can shift observed bands [13], [14]. A uniform or region-specific scaling factor may reduce systematic frequency error, but the assignment should also agree with the calculated displacement pattern [13]. The associated N–H region is particularly sensitive to intermolecular effects and should not be judged by an isolated gas-phase frequency alone [3], [4].

➤ *Role of Machine Learning*

Machine learning can be introduced at two levels. With limited assignment tables, conventional chemometric models can use peak positions, channel presence, relative intensity categories and cross-channel shifts [15], [16]. With many raw spectra, one-dimensional convolutional networks can learn local peak shapes directly [9], [10]. A two-branch model can preserve distinct FTIR and Raman representations before joining their learned features, while transformer-based approaches may support spectrum-to-structure relationships [8]. Interpretation and validation remain essential: influential regions should correspond to chemically plausible amino, nitrile, chlorine and ring vibrations, and related replicates must not be allowed to leak across training and testing partitions [11], [12].

III. MATERIALS AND METHODS

➤ *Source and Status of the Spectral Data*

The analyzed dataset consists of reported vibrational assignments for 4-amino-2-chlorobenzonitrile under KBr, Nujol and laser Raman conditions, together with reference values for 4-aminobenzonitrile. The available records contain wavenumbers and qualitative intensity labels. They do not contain replicate-level intensity vectors, instrumental metadata or uncertainty estimates. The present statistical analysis is therefore limited to matched peak positions.

➤ *Data Harmonization*

Each assigned normal mode was treated as the matching key across the spectral columns. A dash was interpreted as an unreported or unobserved value, not as a frequency of zero. Intensity labels were retained as categorical descriptors. No missing frequency was imputed. Comparisons were calculated only when both relevant channels contained a value for the same assigned mode. This conservative treatment avoids creating artificial peaks and follows the broader principle that spectral preprocessing must preserve chemically meaningful information [7], [17].

➤ *Agreement Metrics*

For each pair of matched channels, the signed frequency difference was defined as the second-channel value minus the first-channel value. Mean absolute error summarized the typical unsigned difference, while root-mean-square error gave greater weight to larger differences. Pearson correlation was also calculated to describe overall frequency ordering. Because the recorded modes span more than 3000 cm^{-1} , correlation is expected to be high and must not be interpreted alone. The joint reporting of absolute error, correlation and assignment-level inspection is consistent with recommendations to avoid relying on one validation statistic [11]-[13].

➤ *Proposed Quantum-Chemical Protocol*

A later computational study should begin from a geometry optimized without symmetry constraints. A dispersion-aware density-functional method and a polarized, diffuse basis set suitable for C, H, N and Cl should be declared explicitly. The optimized structure should show no imaginary frequency, and harmonic IR intensities and Raman

activities should be calculated at the same level of theory [6], [13], [14]. A scaling factor should be fitted only on a designated calibration subset or adopted from a method-specific benchmark; it should not be tuned separately for every band [13], [14]. Potential-energy distribution or animated displacement inspection should confirm mixed modes.

The simulated spectrum may be generated by broadening each scaled line with a stated Gaussian or Lorentzian width. Comparison with KBr and Nujol values should recognize that condensed-phase association is absent from an isolated-molecule calculation [3], [4], [6]. Cluster or continuum-solvation calculations may be examined where the N–H region shows systematic disagreement. The full input geometry, software version, convergence thresholds and scaling rule should accompany any reported computational results to make the comparison reproducible [13], [14].

➤ Proposed Machine-Learning Workflow

For peak-table analysis, each sample may be represented by peak positions, normalized intensities, binary peak-presence indicators and local band ratios. Suitable transparent baselines include principal-component analysis for visualization and partial least-squares methods for supervised modelling [15], [16], alongside support-vector machines and random forests. These methods require spectra

from multiple compounds and independent preparations; one assignment table cannot support supervised performance estimation [11], [12].

For raw-spectrum analysis, the FTIR and Raman axes should be baseline corrected separately, resampled onto fixed grids and normalized using parameters derived from the training data only [7], [17]. A dual-branch one-dimensional convolutional network can then extract modality-specific features [9], [10]. The branch embeddings may be concatenated before classification. Comparisons must include FTIR-only, Raman-only and fused models. Grouped splitting should keep all technical replicates from one physical preparation in the same partition, and final performance should be estimated on different acquisition days, batches or instruments [11], [12].

IV. RESULTS

➤ Principal Recorded Bands

The measured assignments identify the amino, nitrile, aromatic and chlorine contributions expected from the molecular structure, in agreement with established group-frequency interpretation and prior studies of substituted benzonitriles [3]–[6]. Table 1 summarizes the bands with the greatest value for experimental confirmation, simulation benchmarking and later model interpretation for 4-amino-2-chlorobenzonitrile.

Table 1 Bands with the Greatest Value for Experimental Confirmation, Simulation Benchmarking for 4-Amino-2-Chlorobenzonitrile

KBr cm^{-1}	Nujol cm^{-1}	Raman cm^{-1}	Assignment	Diagnostic interpretation
3222	3280	—	Associated N–H stretch	Amino association and matrix sensitivity
3127, 3092, 3022	3098	—	Aromatic C–H stretches	Aromatic framework
2235	2237	2232	C≡N stretch	Stable nitrile identity marker
1582, 1496, 1453	1550, 1464	1606	Ring C–C stretches	Distributed aromatic fingerprint
1346	—	1336	C–NH ₂ stretch	Amino-specific skeletal feature
1314	1325	—	Kekulé C–C stretch	Ring bond alternation
1271, 1207	1271, 1217	—	In-plane C–H bends	Substitution-sensitive fingerprint
1164	1164	1169	C–CN stretch	Nitrile-to-ring skeletal link
1008	—	—	Ring breathing	Aromatic confirmation
950, 865, 847	961, 875	935	Out-of-plane C–H bends	Substitution-pattern information
—	—	738	C–Cl stretch	Chlorine-sensitive Raman feature
728	730	720	Ring deformation	Strong cross-channel agreement
558	569	534	C≡N in-plane bend	Secondary nitrile feature
383	—	395	C–NH ₂ in-plane bend	Low-frequency amino feature
—	—	365	C–Cl bend	Low-frequency chlorine feature
237	238	—	C–NH ₂ out-of-plane bend	Highly consistent amino feature
216	216	—	C–CN out-of-plane bend	Highly consistent skeletal feature

➤ Cross-Medium and Cross-Technique Agreement

Nineteen assigned modes were common to the KBr and Nujol columns. The two infrared preparations preserved almost identical frequency ordering, with $r = 0.9999$. Their mean absolute difference was 10.79 cm^{-1} , root-mean-square difference was 17.04 cm^{-1} and median absolute difference was 10 cm^{-1} . The mean signed Nujol-minus-KBr shift was $+5.42 \text{ cm}^{-1}$. The largest difference, 58 cm^{-1} , occurred in the associated N–H region, consistent with the known environmental sensitivity of amino association [3], [4].

Nine assigned modes were common to the KBr and Raman columns. They produced $r = 0.9990$, a mean absolute difference of 17.67 cm^{-1} and a root-mean-square difference of 26.67 cm^{-1} . The mean signed Raman-minus-KBr shift was $+7.00 \text{ cm}^{-1}$. The 865 cm^{-1} KBr and 935 cm^{-1} Raman C–H out-of-plane assignments produced the largest matched difference and warrant inspection of the original spectra and calculated displacement vectors before being treated as the same normal mode [5], [6], [13], [14].

Table 2 Statistical Agreement of Matched Vibrational Modes Across KBr, Nujol and Raman Measurements

Comparison	Matched modes	Pearson r	MAE cm^{-1}	RMSE cm^{-1}	Mean signed shift cm^{-1}
KBr vs Nujol	19	0.9999	10.79	17.04	+5.42
KBr vs Raman	9	0.9990	17.67	26.67	+7.00

➤ Experimental Anchors for Simulation

The 2232–2237 cm^{-1} nitrile stretch is the strongest candidate for testing global frequency scaling because it is isolated and recorded in every channel [3]–[6]. The 1164–1169 cm^{-1} C–CN stretch and 720–730 cm^{-1} ring deformation provide additional cross-channel anchors. Amino-related bands should be interpreted with more caution because hydrogen bonding and association can perturb N–H frequencies [3], [4]. The 3222–3280 cm^{-1} associated N–H signal reflects condensed-phase interactions, while the C–NH₂ stretch and bending modes may couple with ring motion. C–Cl modes at 738 and 365 cm^{-1} should be verified through

calculated displacement vectors because the low-frequency region contains several skeletal deformations; displacement-based checking is also central to reliable scaled-frequency comparison [13], [14].

➤ Experimental Anchors for Interpretable Machine Learning

These chemically assigned regions convert established spectral interpretation into explicit, testable model features. Their use should remain subordinate to independent validation and inspection of whether model importance follows chemically meaningful bands [3], [4], [8]–[10].

Table 3 Chemically Informed Spectral Features and Interpretation Criteria for Multimodal Machine Learning

Feature family	Recorded region	Suggested representation	Interpretation test
Nitrile stretch	2232–2237 cm^{-1}	Peak position, height, area and width	Model importance should identify the isolated C≡N region
Associated amino stretch	3222–3280 cm^{-1}	Broad-band area, centroid and width	Check sensitivity to preparation and hydrogen bonding
Amino skeletal modes	1336–1346, 383–395 and 237–238 cm^{-1}	Peak presence and local intensity ratios	Should separate amino-bearing compounds
Chlorine modes	738 and 365 cm^{-1} Raman	Peak presence and Raman activity	Should support discrimination from non-chlorinated analogues
Aromatic fingerprint	1450–1610 and 795–1008 cm^{-1}	Full local windows or learned convolution features	Should capture distributed substitution effects
Cross-modal consistency	Matched FTIR and Raman modes	Frequency difference and joint embedding	Should improve confidence without duplicating information

V. DISCUSSION

➤ Chemical Interpretation of the Recorded Pattern

The observed spectrum is internally consistent with an amino- and chloro-substituted aromatic nitrile [3]–[6]. The nitrile stretch is narrowly grouped across all measurement conditions, whereas the associated N–H band shows the greatest KBr–Nujol displacement. This contrast is chemically useful. The nitrile band can act as a robust identity anchor, while the N–H region can report local association and sample environment. The amino group is also represented through C–NH₂ stretching and bending, which reduces reliance on a single broad high-frequency signal [3], [4].

The aromatic modes form a dense multiband fingerprint rather than an isolated diagnostic line. Ring stretches between approximately 1450 and 1610 cm^{-1} , in-plane bends near 1200–1271 cm^{-1} , ring breathing near 1008 cm^{-1} and out-of-plane C–H modes below 1000 cm^{-1} collectively encode the substitution pattern [3]–[6]. The Raman C–Cl stretch and bend add low-frequency evidence that is absent or weak in the reported infrared columns. This distribution supports modality-specific processing followed by feature fusion.

➤ Meaning of the Agreement Statistics

The cross-channel correlations are close to unity, but their magnitude is partly a consequence of the wide frequency

range. Two lists that preserve the ordering of high-, middle- and low-frequency modes will naturally correlate strongly. Absolute differences therefore provide the more useful measure. Most KBr–Nujol pairs differ by approximately 0–11 cm^{-1} , while the associated N–H band is an exception. The greater KBr–Raman error reflects both channel-specific responses and uncertainty in pairing mixed modes. These statistics should be treated as descriptive agreement measures rather than predictive-validation results [11]–[13].

➤ How Simulation should be Correlated with Experiment

A defensible calculation should report both numerical error and assignment consistency. After scaling, simulated and experimental frequencies can be matched through a constrained procedure that penalizes large frequency separation but also requires compatible atomic motion [13], [14]. Mean absolute and root-mean-square errors should be given for the complete set and for separate high-, middle- and low-frequency regions. Reporting only the best-matching bands would bias the comparison. The N–H region should be discussed separately because an isolated-molecule model may not reproduce association in KBr or Nujol [3], [4].

➤ How Machine Learning should use the Recorded Values

The present table can define expected regions and interpretation checks, but it is not a training dataset. A peak-based pilot model would require multiple spectra per

material, including independent preparations and realistic sources of variation. Full-spectrum deep learning would require still more observations [9], [10]. Synthetic broadening of the listed peaks may support software testing or pretraining, but synthetic copies must never be distributed between training and testing and reported as independent experimental evidence [11],[12].

After suitable data collection, a staged evaluation is recommended. PCA may first reveal batch, baseline and instrument effects [15]. PLS-DA and support-vector machines can provide conventional baselines [11],[16]. FTIR-only and Raman-only convolutional networks should then be compared with the proposed fused network [9], [10]. Saliency or integrated-gradient maps should be aggregated across test samples and checked against the nitrile, amino, chlorine and aromatic regions identified here [8]. If a model relies mainly on background edges, substrate signals or acquisition artifacts, its apparent accuracy would not represent chemical learning.

➤ Comparison with 4 Aminobenzonitrile

The supplied reference values for 4-aminobenzonitrile allow cautious qualitative comparison [14,15]. Its nitrile stretch appears at 2197 cm^{-1} in IR and 2210 cm^{-1} in Raman, while the chlorinated compound is reported at $2235\text{--}2237\text{ cm}^{-1}$ in IR and 2232 cm^{-1} in Raman. Several ring and deformation modes also shift after chlorine substitution, and new C–Cl modes appear. These differences are appropriate candidate features for discrimination. Nevertheless, the measurements should be repeated under matched instrumental and sample conditions before attributing every difference solely to substitution.

VI. RECOMMENDED REPRODUCIBLE VALIDATION DESIGN

The validation design below follows established chemometric guidance on preprocessing, model assessment, partitioning and leakage prevention [7], [11], [12], [17].

Table 4 Recommended Reproducible Validation Framework for Multimodal FTIR–Raman Machine-Learning Models

Stage	Minimum requirement	Reason
Sample design	Multiple independent preparations of each compound	Separates chemical identity from preparation-specific artifacts
Acquisition	Paired FTIR and Raman spectra from the same samples	Enables valid multimodal fusion
Metadata	Instrument, day, operator, laser power, scans, resolution and sample condition	Supports confounder analysis and reproducibility
Partitioning	Group all technical replicates by physical preparation	Prevents leakage between training and testing
Baselines	PCA, PLS-DA, SVM and single-channel models	Tests whether complex fusion adds real value
Metrics	Balanced accuracy, macro-F1, sensitivity, specificity, calibration and confidence intervals	Avoids reliance on accuracy alone
Interpretation	Compare model attention with assigned chemical bands	Tests whether predictions use meaningful spectral evidence
External test	New batch, day and preferably a second instrument	Measures generalization under realistic variation
Unknown handling	Confidence threshold or open-set rejection	Prevents forced assignment of unfamiliar materials

VII. LIMITATIONS

The current evidence is based on assigned peak positions rather than raw numerical spectra. Peak heights, areas, widths, baselines and replicate variation are unavailable. Instrument settings and uncertainty values are not reported in the assignment table. Consequently, the study cannot train or externally validate a classifier, estimate detection limits or determine whether spectral fusion improves prediction [11], [12]. The simulation section specifies a protocol but does not claim calculated frequencies, energies or optimized geometry [13], [14]. These boundaries are necessary to keep the conclusions proportional to the available evidence.

Some reported normal modes may contain mixed atomic motions, and the frequency matching across IR and Raman columns may not always represent the same dominant displacement. The relatively large difference between the 865 cm^{-1} KBr and 935 cm^{-1} Raman assignments illustrate this concern. Original spectra, computational animations and potential-energy distributions are needed for final confirmation.

cm^{-1} KBr and 935 cm^{-1} Raman assignments illustrate this concern. Original spectra, computational animations and potential-energy distributions are needed for final confirmation.

VIII. CONCLUSIONS

The recorded bands of 4-amino-2-chlorobenzonitrile provide a coherent experimental benchmark for later quantum-chemical and machine-learning studies. The nitrile stretch at $2232\text{--}2237\text{ cm}^{-1}$ is the most stable cross-modal marker. Associated N–H stretching at $3222\text{--}3280\text{ cm}^{-1}$ and the C–NH₂ stretching and bending modes establish the amino contribution. Raman bands at 738 and 365 cm^{-1} supply chlorine-sensitive evidence, while the aromatic fingerprint distributes structural information across several regions.

The KBr and Nujol assignments show close agreement for most matched modes, whereas larger differences appear

for environment-sensitive and cross-technique comparisons. These results justify a workflow in which DFT calculations are evaluated with scaled frequency errors and displacement-vector checks, and ML models are evaluated with grouped external validation and chemical interpretation. The available assignments can guide feature design and simulation validation, but raw replicate spectra are essential before predictive performance is claimed.

➤ *Novelty and Contribution*

The contribution of this work is the integration of three levels of evidence without conflating them. First, it organizes the recorded FTIR and Raman values of 4-amino-2-chlorobenzonitrile into a chemically interpreted benchmark. Second, it quantifies agreement between the available measurement conditions. Third, it translates the assignments into explicit validation targets for quantum-chemical simulations and interpretable multimodal machine learning. The framework also defines the boundary between results supported by existing measurements and outcomes that require new calculations or experiments.

REFERENCES

- [1]. Hollas, J. M. (2004). *Modern Spectroscopy* (4th ed.). Wiley.
- [2]. Nakamoto, K. (2009). *Infrared and Raman Spectra of Inorganic and Coordination Compounds* (6th ed.). Wiley.
- [3]. Silverstein, R. M., Webster, F. X., Kiemle, D. J., and Bryce, D. L. (2014). *Spectrometric Identification of Organic Compounds* (8th ed.). Wiley.
- [4]. Socrates, G. (2004). *Infrared and Raman Characteristic Group Frequencies: Tables and Charts* (3rd ed.). Wiley.
- [5]. Singhal, S. (2025). Comparative study of spectroscopic analysis of 2-chloro-5-nitrobenzonitrile and 2,5-dichlorobenzonitrile. *International Journal of Scientific Development and Research*, 10(7), b147–b160. <https://doi.org/10.56975/ijisdr.v10i7.304166>
- [6]. Singhal, S. (2026). Vibrational frequency, force constant, and thermodynamic correlation study of substituted benzonitrile compounds. *Journal of Emerging Technologies and Innovative Research*, 13(5), j29–j46.
- [7]. Rinnan, Å., van den Berg, F., and Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, 28(10), 1201–1222. <https://doi.org/10.1016/j.trac.2009.07.007>
- [8]. Alberts, M., Patel, H., and Aspuru-Guzik, A. (2024). Leveraging infrared spectroscopy for automated structure prediction. *Communications Chemistry*, 7. <https://doi.org/10.1038/s42004-024-01341-w>
- [9]. Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.
- [10]. LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444. <https://doi.org/10.1038/nature14539>
- [11]. Szymańska, E., Saccenti, E., Smilde, A. K., and Westerhuis, J. A. (2012). Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics*, 8, 3–16. <https://doi.org/10.1007/s11306-011-0330-3>
- [12]. Xu, Y., and Goodacre, R. (2018). On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *Journal of Analysis and Testing*, 2, 249–262.
- [13]. Merrick, J. P., Moran, D., and Radom, L. (2007). An evaluation of harmonic vibrational frequency scale factors. *The Journal of Physical Chemistry A*, 111(45), 11683–11700. <https://doi.org/10.1021/jp073974n>
- [14]. Scott, A. P., and Radom, L. (1996). Harmonic vibrational frequencies: An evaluation of Hartree–Fock, Møller–Plesset, quadratic configuration interaction, density functional theory, and semiempirical scale factors. *The Journal of Physical Chemistry*, 100(41), 16502–16513. <https://doi.org/10.1021/jp960976r>
- [15]. Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1–3), 37–52. [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9)
- [16]. Wold, S., Sjöström, M., and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1)
- [17]. Savitzky, A., and Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8), 1627–1639. <https://doi.org/10.1021/ac60214a0>