Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 270 (2025) 3518–3527

Procedia Computer Science

www.elsevier.com/locate/procedia

29th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2025)

Raman spectroscopy and machine learning can quantitatively asses clindamycin in liquid samples

Eduard C. Milea[a*], Andreia Alecu[a], Alice Stoica[a], Marian Necula[a], Ion Petre[a,c], Simona Litescu[a], Mihaela Paun[a,b]

[a]*National Institute of R&D for Biological Sciences, Splaiul Independentei 296, 060031 Bucharest, Romania*
[b]*University of Bucharest, Bd. Regina Elisabeta no. 4–12, Sector 3, 030108 Bucharest, Romania*
[c]*Department of Mathematics and Statistics, University of Turku, Finland*

Abstract

Raman spectroscopy offers a powerful, non-destructive tool for pharmaceutical quantification, particularly in environments where traditional techniques like HPLC are limited by throughput and sample preparation demands. However, the quantification of low-concentration compounds remains challenging due to weak Raman scattering and high background interference. This study evaluates the use of portable Raman instrumentation coupled with Support Vector Regression (SVR) to quantify clindamycin across various concentrations. Spectral preprocessing steps included Savitzky–Golay smoothing, Standard Normal Variate (SNV) normalisation, and blank subtraction (ΔSNV) to enhance analyte-specific signal fidelity. Three SVR-based models were developed using full spectra, chemically meaningful fingerprint bands, and coefficient-filtered features. Models were evaluated through grouped cross-validation, bootstrapping, and external testing on formulations prepared in different solvent matrix and derived from distinct clindamycin sources (commercial tablet vs. analytical-grade standard). The top performing model achieved R^2 values exceeding 0.98 with root mean squared errors below 2.85 mg/mL. Blind sample predictions, made on fully unseen data, fell within 95% confidence intervals of the true concentrations, demonstrating strong model robustness.

Keywords: Raman spectroscopy; machine learning; clindamycin; portable instrumentation; pharmaceutical quantification; spectral preprocessing

* Corresponding author. *E-mail address:* eduard.milea@incdsb.ro

10.1016/j.procs.2025.09.477

1. Introduction

Quantitative Raman spectroscopy has emerged as a valuable tool in pharmaceutical analysis, particularly in therapeutic drug monitoring [1–3] and validation for drug quality assurance [4,5] and regulatory compliance. Its non-destructive nature, minimal sample preparation, and capability to work in aqueous environments make it especially attractive for real-time or in-process analytical applications. However, the detection and quantification of low-concentration pharmaceutical compounds in aqueous matrices remain challenging due to the inherently weak Raman scattering signals and the dominance of background noise from solvents like water**.** [6]

Clindamycin, a lincosamide antibiotic widely used in dermatological, dental, and systemic infections, requires accurate quantification to ensure therapeutic efficacy and minimize the risk of resistance or toxicity. While high-performance liquid chromatography (HPLC) is the gold standard for drug quantification [7], it is time-consuming, reagent-intensive, and unsuitable for inline or point-of-need applications. Raman spectroscopy, when combined with advanced data modelling, presents a promising alternative for rapid, non-invasive concentration assessment. Nonetheless, the development of robust predictive models for Raman spectra is non-trivial, particularly in low-concentration regimes, where signal-to-noise ratios are critically low.

Traditional chemometric techniques such as Partial Least Squares Regression (PLSR) and Principal Component Regression (PCR) have been widely employed for quantitative Raman analysis [8–10]. These linear multivariate approaches reduce data dimensionality and model the relationship between spectral data and concentration. PLSR, in particular, has shown utility in pharmaceutical quantification tasks. However, these methods often struggle with overfitting, poor generalization to external datasets, and limited adaptability to nonlinear spectral variations or baseline shifts introduced by environmental or instrumental factors [11].

Recent studies have increasingly explored machine learning (ML) techniques to improve Raman quantification accuracy [12,13], particularly under conditions of low analyte concentration and instrumental variability. Support Vector Regression (SVR), especially with linear kernels, has been highlighted for its robustness against multicollinearity and its ability to generalize in high-dimensional feature spaces with limited data [14].

Several practical challenges complicate real-world implementation of ML-based Raman quantification. Instrumental variability, such as wavelength calibration drift, laser instability, and detector sensitivity, is especially pronounced in portable and benchtop Raman systems, which are increasingly used in pharmaceutical QC labs [15] and field applications. Environmental factors such as ambient light [16], temperature fluctuations, and sample positioning can introduce baseline instability and reproducibility issues [17,18]. Moreover, studies often underreport device specifications, preprocessing pipelines, or cross-validation protocols, limiting reproducibility and transferability of the developed models.

This study addresses these challenges by evaluating Support Vector Regression (SVR) models tailored for high-dimensional spectral data, and by assessing model robustness through grouped cross-validation, bootstrapping, and external validation on samples that differ in both solvent matrix and compound source. The objective is to demonstrate whether portable Raman devices when coupled with modern ML approaches, can serve as a reliable tool for rapid quantitative drug analysis in complex matrices, and to establish a reproducible framework for such applications.

2. Materials and instrumentation

2.1. Hardware

All spectral measurements were acquired using a StellarNet Quasar-TEC-X2 Raman spectrometer, coupled with a RamuLaser-785 excitation source. This compact, modular system is designed for field and benchtop applications, combining flexibility with analytical-grade performance. Its architecture consists of a 785 nm laser, a high-throughput fibre-coupled spectrometer, and a thermoelectrically cooled CCD detector. The setup is particularly suitable for quantitative Raman spectroscopy due to its low baseline drift and stable long-term operation.

2.1.1. Laser excitation and optical characteristics

The RamuLaser-785 is a stabilized diode laser with a nominal output power of 499 mW at 785 nm, enabling strong Stokes scattering while minimizing fluorescence, relevant for organic and aqueous matrices. Laser output power plays a critical role in data acquisition: higher power increases the Raman signal but also introduces risks such as sample heating, photobleaching, or even chemical degradation during prolonged exposure. Thus, a balance between signal intensity and sample integrity, particularly in repetitive or time-series measurements, must be struck.

Another important specification is the laser linewidth, typically <0.2 nm for stabilized sources like the RamuLaser. A narrow linewidth ensures sharp excitation, which translates to better spectral resolution and more accurate Raman peak positions, is essential for quantitative interpretation and chemometric modelling.

2.1.2. Spectrometer and optical resolution

The Quasar-TEC-X2 spectrometer includes a compact, high-throughput optical spectrometer with a fixed diffraction grating, yielding a spectral range from approximately 200 to 2750 cm^{-1} in the Raman shift domain. The optical resolution, specified at ~4 cm^{-1}, refers to the minimum distinguishable difference between two adjacent Raman peaks. It is determined by the spectrometer's slit width, grating density, and the number of pixels on the detector. Higher resolution improves peak separation but may reduce signal intensity, representing another key trade-off in instrument design.

The spectrometer's aperture and fibre optics are optimized for signal throughput while minimizing stray light. Combined with spectral smoothing and baseline correction during preprocessing, this configuration enables high-quality spectra from both high and low concentration samples.

2.1.3. Thermoelectric cooling and noise reduction

An important feature of the Quasar-TEC-X2 is the thermoelectrically cooled CCD detector. CCDs generate a small but not negligible amount of dark current, a source of electronic noise that becomes significant during long integration times or when analysing weak Raman signals. To address this, the detector is cooled via a Peltier thermoelectric cooler (TEC) to approximately -20°C below ambient, substantially reducing the dark current and associated thermal noise.

This cooling is essential for improving the signal-to-noise ratio (SNR) and enabling long exposure durations. Without thermal regulation, extended exposures would elevate the noise floor, blur low-intensity peaks, and introduce quantification errors which are especially problematic when dealing with low-concentration analytes. By suppressing baseline drift and fixed-pattern noise, cooling ensures the analytical stability and reproducibility required for ML applications.

2.2. Spectral acquisition protocol

Raman spectra were acquired using through StellarPro v2.2 software. Samples were placed in Qorpak GLC-00980 4 mL borosilicate glass Clear Type I vials, which were then inserted in the standard opaque black cuvette holder. This holder is specifically designed to minimize ambient light interference and securely align the fibre-coupled Raman probe with the vial wall, ensuring consistent optical geometry across replicates.

Each sample measurement was acquired under the following conditions:

- *Integration Time:* 2000 milliseconds
- *Number of Averaged Scans:* 3
- *Laser Output Power:* Fixed at level 5.

The chosen acquisition parameters were empirically established to ensure the optimal equilibrium between signal intensity, spectral resolution and clarity, all while minimizing photothermal effects. The opaque cuvette holder, in particular, reduces stray light and ensures reproducible probe-to-sample distance, a critical factor for consistent Raman scattering efficiency.

2.2.1. Sample preparation

Training Dataset: Clindamycin HCl Standard

The training dataset was generated using a clindamycin hydrochloride (HCl) reference standard (>98% purity, Sigma-Aldrich). A series of solutions were prepared via serial dilution in a 0.05 mol·L^{-1} phosphate buffer (pH 7.20). The phosphate buffer provided a stable Raman-transparent medium.

The following clindamycin HCl concentrations were obtained: 50, 40, 30, 20, 17.5, 15, 12.5, 10, 7.5, 5, 2.5, and 0 mg/mL (blank). Each concentration level was independently prepared in triplicate, and each sample was measured with 20 replicate scans, supporting robust statistical modelling, feature selection, and residual analysis.

Test dataset: Commercial Tablet (Clindamycin-MIP 600mg, MIP Pharma)

For the external validation of the model, a separate test set was prepared using a 600 mg clindamycin hydrochloride film-coated tablet (commercial formulation). The tablet was fully dissolved in deionized water, then filtered on PFTE discs, 0.45 micromesh to remove insoluble excipients. From the stock solution (110 mg/mL), the following dilution series was prepared at the following concentration ranges, to be used as the test dataset for the model: 74, 55, 44, 37, 31, 27, 22, and 0 mg/mL

This test dataset differs from the training set in both solvent composition and formulation type. These differences introduce deliberate variability to assess the model's ability to generalize beyond controlled conditions:

- Use of distilled water instead of phosphate buffer as the solvent medium
- Presence of excipients and inactive ingredients inherent to the tablet formulation, in contrast to the analytical-grade standard used in training
- Altered formulation of active molecule due to the tablet's salt form and manufacturing additives

Such variations mirror real-world deployment scenarios and provide a robust benchmark for evaluating model performance under practical, out-of-distribution conditions.

2.3. Representative Raman spectrum and spectral characteristics

The averaged Raman spectrum of both clindamycin formulations (Fig. 1), at 50 mg/mL and 55 mg/mL respectively, overlaid with their set's 0 mg/mL blank, reveal several distinct vibrational features in the fingerprint region (300–1000 cm^{-1}). The three main Raman-active bands are observed at 378 cm^{-1}, 637 cm^{-1}, and 706 cm^{-1}. These peaks are ascribed to active compound, and the confirmation of the correct ascription of the specific wavenumber to clindamycin was done by comparing the spectra to the one available in Wiley's SpectraBase platform. The peak intensity relative to the blank highlights the potential utility for quantitative detection.

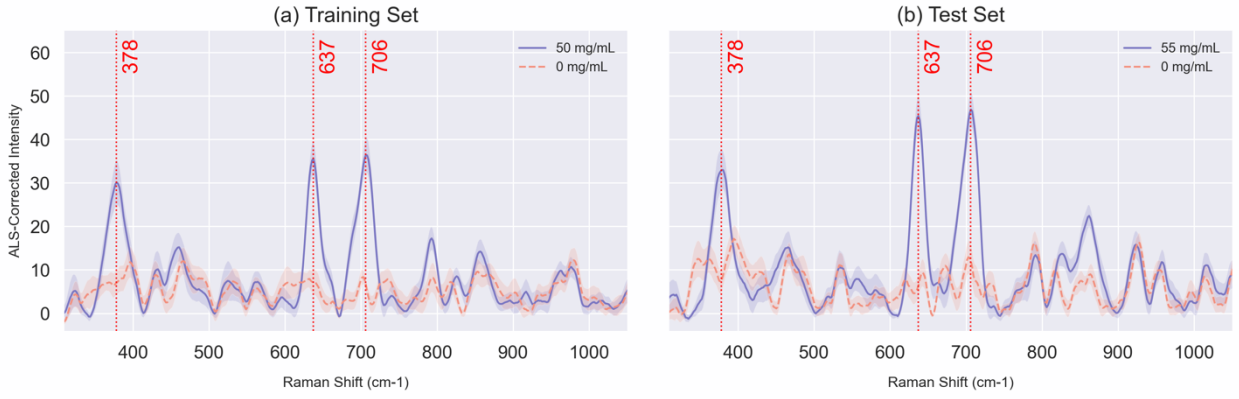

Fig. 1. Raman Spectra of Clindamycin HCl (a) and tablet formulation (b).

Peak at ~378 cm⁻¹

This band is attributed to low-frequency skeletal deformation involving the sugar ring and its heavy substituents. It likely includes bending or torsional motion of the C–Cl bond. The involvement of heavy atoms contributes to its Raman activity [19].

Peak at ~637 cm⁻¹

This peak corresponds to C–Cl stretching on the chlorinated sugar ring. C–Cl stretches typically appear in this region and are strongly Raman-active due to the polarizability of the chlorine atom. This band serves as a clear spectral marker for the presence of halogen substitution [20].

Peak at ~706 cm⁻¹

This band is assigned to C–S stretching in the methylthio (–S–CH_3) group attached to the sugar ring. Thioether bonds produce occasionally strong Raman signals in this region, owing to the polarizability of sulphur. The intensity of this peak makes it useful for structural identification and quantification [21,22].

3. Methods

3.1. Spectral processing

To prepare the Raman spectra for quantitative modelling, all raw signal data underwent a structured preprocessing workflow consisting of smoothing, normalisation, blank subtraction, and feature extraction. These steps were designed to reduce measurement noise, remove baseline artifacts, and isolate analyte-specific Raman signals critical for quantification.

3.1.1. Smoothing and Standard Normal Variate normalisation

Each spectrum was first smoothed using a Savitzky–Golay filter (window size 9, polynomial order 3). This reduces high-frequency noise while preserving the integrity of peak shapes [23].

Subsequently, Standard Normal Variate (SNV) normalisation was applied to correct for non-chemical variability. SNV transforms each spectrum by subtracting its mean and dividing by its standard deviation, effectively removing additive offsets and multiplicative scaling effects [24] that may arise from sample presentation, optical alignment, or laser intensity fluctuations. These distortions can interfere with quantitative analysis, particularly when comparing peak intensities across samples or building regression models.

While Asymmetric Least Squares (ALS) baseline correction was used elsewhere in the workflow to address broad baseline curvature, it does not correct for multiplicative effects [25] or intensity scaling across samples. SNV was therefore preferred at this stage as it ensures each spectrum is centred and scaled independently, enabling direct comparison of spectral features and supporting robust model generalization.

3.1.2. Blank subtraction: ΔSNV Representation

While SNV normalisation improves spectral comparability by correcting for additive and multiplicative effects, it does not remove systematic background contributions inherent to the experimental setup. These include Raman scattering from the cuvette, solvent matrix, and optical components, all of which can introduce consistent but non-analyte-specific features into the spectra.

To further isolate the chemical signal of clindamycin, we applied a blank subtraction step. Specifically, the average SNV-normalized spectrum of all 0 mg/mL (blank) samples was computed and subtracted from each individual SNV spectrum.

This transformation, referred to as ΔSNV, enhances the analyte-specific contrast in the spectra by removing background features that are consistent across all samples but unrelated to clindamycin concentration and is required in the low-concentration regime, where the target signal is weak and easily masked by background structure. Hence

by suppressing fixed-pattern noise from the sample or instrumentation, ΔSNV improves the SNR for subtle features, thus augmenting the model performance.

This ΔSNV representation was used as the feature input for all learning models in order to predict the compound quantity in the samples.

3.1.3. Replicate-level signal stability assessment

To ensure the reliability of Raman signal acquisition, we assessed potential signal degradation across sequential replicate measurements. Prolonged laser exposure or thermal accumulation can lead to photobleaching or compound degradation, particularly in aqueous systems. Such effects, if present, could introduce artifacts that compromise quantification and detection limit estimates.

For each sample replicate (grouped by sample concentration), we computed the mean Raman intensity over the identified peaks in the fingerprint. A linear regression model was then fitted to capture the trend of signal intensity as a function of replicate index. The resulting slopes, along with their corresponding p-values and correlation coefficients, were recorded for subsequent analysis.

This analysis served as a quality control step to determine whether replicate-level degradation needed to be addressed in preprocessing or modelling.

3.2. Feature selection strategies

To investigate the trade-off between model complexity, interpretability, and accuracy, three distinct strategies were employed to define the input feature space.

Full Feature Set (FFS Model) - In this approach, the model utilized all available Raman shifts, hence persevering the full vibrational landscape of each spectrum. While it offers maximal modelling capacity, this approach increases the risk of overfitting and dilutes model interpretability due to the number of features resulting from the spectral bands.

Top Band Selection (TBS Model) - To reduce dimensionality while preserving chemical relevance, a targeted subset of fingerprint peaks was extracted from the ALS-corrected spectra using a peak-finding algorithm. These peaks represent structurally informative vibrational modes specific to clindamycin. This method emphasizes domain knowledge and spectroscopic specificity.

Coefficient-Based Filtering (CBF Model) - A third strategy involved training an initial model on the top band feature set, then ranking individual features based on the magnitude of their learned SVR coefficients. The top ranking features were retained for the final modelling step. This data-driven filtering approach prioritizes predictive utility over chemical interpretability and is particularly useful for distilling sparse models suited for deployment or explanation.

3.3. Machine learning pipeline

All models were built using a linear Support Vector Regression (SVR) estimator, implemented via a Pipeline in scikit-learn that included z-score standardization of features. SVR was selected due to its suitability for high-dimensional regression tasks, particularly under conditions of collinearity and limited sample size. The use of a linear kernel enabled direct interpretation of feature coefficients, which was essential for the coefficient-based feature selection strategy described earlier.

Model hyperparameters, specifically the regularization parameter C and the margin of tolerance epsilon, were optimized via grid search using a 5-fold GroupKFold cross-validation strategy. In this approach, folds were constructed such that all spectral replicates from the same physical sample were assigned to the same fold and never split between training and validation sets. This grouping was based on a combination of sample identity and

replicate number. This procedure mitigated data leakage and produced a more realistic estimate of generalization performance.

Model evaluation included the following components:

- R^2 (coefficient of determination) and Root Mean Squared Error (RMSE) on held-out test data;
- Grouped cross-validation R^2 to evaluate internal consistency across training folds;
- Bootstrapped RMSE and R^2 distributions (via resampling) to quantify uncertainty;
- Residual diagnostics, including plots of residuals vs. predicted values and predicted vs. true concentration (averaged by sample identity).

4. Results and discussions

4.1. Signal integrity and instrument stability

Before initiating model training, we conducted a signal integrity assessment to ensure that Raman signal degradation, instrumental drift, or photobleaching effects did not compromise the dataset. Given the sequential nature of replicate acquisition and the use of a fibre-coupled probe, the potential for cumulative sample exposure or optical misalignment warranted investigation.

To evaluate the internal consistency of replicates, we computed pairwise Euclidean distances between all replicate spectra within each sample, on the raw spectrum. The mean pairwise distance served as a proxy for spectral variability, and samples were flagged as inconsistent when their mean exceeded a global threshold, defined as two standard deviations above the dataset-wide mean.

Before cleaning, one sample exceeded the global inconsistency threshold (mean pairwise distance = 1.709 vs threshold = 1.646) and was marked as inconsistent. Several individual replicates across multiple samples also showed unusually high distances from their group averages, indicating potential outlier behaviour due to instrumental instability or transient optical misalignment. In total, 11 replicate spectra were removed across 10 samples based on local outlier thresholds.

After removing these outlier replicates, the pairwise analysis was repeated. All remaining samples fell below the updated global threshold of 1.443, indicating improved within-sample consistency.

In parallel, we assessed potential systematic drift across replicates using linear regression applied to the average signal intensity of three Raman peaks (706, 637, and 378 cm^{-1}). For each sample, we computed the slope, R^2, and statistical significance of the trend over replicate number. Across all samples, no significant drift was detected: all slopes were near zero (average absolute slopes < 0.0016), and 0% of peaks showed statistically significant drift ($p < 0.05$). This further supports the conclusion that temporal signal degradation was negligible within the replicate acquisition window.

Together, the results from pairwise replicate analysis and drift regression confirm that the dataset is spectrally stable and suitable for downstream chemometric modelling.

4.2. Model performance across feature strategies

4.2.1. Predictive performance

Table 1. reports the results of the trained models, and we can notice that the *FFS Model*, which incorporated the entire spectral input, exhibited comparatively lower predictive performance, achieving an R^2 of 0.81 and an RMSE of 8.73 on held-out samples. While regularization through GroupKFold cross-validation helped mitigate overfitting, the model's high dimensionality likely introduced noise and redundancy, diminishing its overall accuracy and interpretability.

In contrast, the *TBS Model*, constructed using a curated subset of chemically meaningful spectral bands, demonstrated substantially improved performance, with an R^2 of 0.97 and RMSE of 3.54. This model balances parsimony and interpretability, as the selected bands correspond to known fingerprint regions, thereby reducing sensitivity to irrelevant or low-informative spectral areas while preserving critical signal characteristics.

The *CBF Model*, which selects features based on the magnitude of SVR coefficients from an initial model fit, outperformed both alternatives. It achieved the highest R^2 (0.98) and lowest RMSE (2.85), indicating that a sparse, data-driven feature selection strategy can yield superior predictive accuracy. Moreover, the consistency of this model's performance across cross-validation (CV R^2 = 0.96) and bootstrapping (Boot R^2 = 0.98) suggests strong generalizability and robustness to sample variation.

Table 1. Trained model results.

Model	Test R^2	RMSE	CV R^2	Bootstrapped R^2	Bootstrapped RMSE
Full Feature Set	0.81	8.73	0.95	0.81	8.72
Top Band Selection	0.97	3.54	0.97	0.97	3.60
Coefficient-Based Filtering	0.98	2.85	0.96	0.98	2.84

4.2.2. Model diagnostic

To support visual inspection, we plotted the predicted vs. true concentration values (averaged per sample) and corresponding residuals vs. predicted for each model. We can see in Fig.2 (a) the predicted vs. true plot, all three SVR models align closely with the identity line, particularly the *TBS* and *CBF* models, which exhibit tighter clustering around the true values. The full feature model displays a larger RMSE, especially at higher concentrations, consistent with its larger feature space and tendency to overfit.

The residuals vs. predicted plot in Fig.2. (b) confirms these observations: residuals from the *TBS* and *CBF* models are symmetrically distributed around zero, with no strong drift or trend across the prediction range. In contrast, the full model exhibits a subtle downward bias at higher concentrations, reflected in the more negative residuals. These visual diagnostics validate the models' calibration and reinforce the advantages of targeted feature selection in maintaining predictive accuracy while minimizing overfitting.

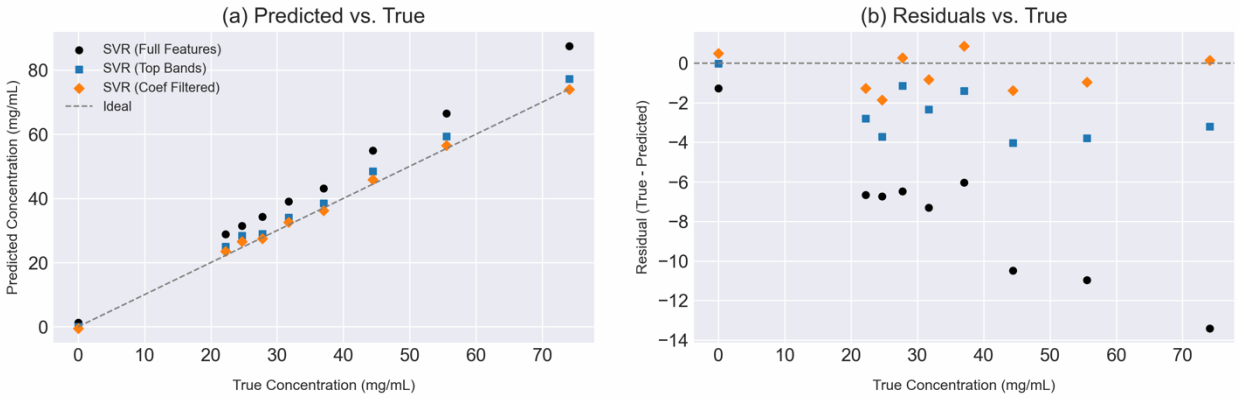

Fig. 2 (a) Predicted vs. True Concentrations; (b) Residuals vs. True

Finally, to further evaluate the robustness of the trained models, they were challenged with two blind clindamycin samples that were entirely excluded from both the training and testing datasets. The true concentrations (28.61 mg/mL and 21.46 mg/mL) were disclosed only after predictions were completed, ensuring an unbiased

evaluation. Each model produced concentration estimates by averaging predictions across spectral replicates, accompanied by standard errors and 95% confidence intervals (Table 2).

Table 2. Blind sample predictions.

Model	Sample mg/mL	Predicted mg/mL	Std. Err.	95% C.I.
Full Feature Set	28.61	29.80	0.53	28.69 - 30.74
	21.46	23.46	0.53	22.42 - 24.50
Top Band Selection	28.61	29.99	0.56	28.88 - 31.06
	21.46	24.09	0.69	22.74 - 25.44
Coefficient-Based Filtering	28.61	29.30	0.66	28.00 - 30.60
	21.46	23.75	0.73	22.33 - 25.17

For both blind samples, all models produced consistent internal predictions with relatively narrow 95% confidence intervals. The *CBF Model* successfully captured the true concentration of the 28.61 mg/mL sample within its 95% confidence interval, and all models provided close approximations of the concentration levels. This blind validation demonstrates the models' ability to generalise reliably to entirely new, unseen samples.

5. Conclusions

This study demonstrates the feasibility and reliability of Support Vector Regression (SVR) for the quantitative analysis of clindamycin using portable Raman spectroscopy under realistic, aqueous conditions. By combining robust spectral preprocessing, targeted feature selection, and grouped cross-validation protocols, we developed and evaluated three SVR-based models capable of accurately quantifying clindamycin across a wide concentration range.

Among the evaluated models, the *Top Band Selection* and *Coefficient-Based Filtering* ones outperformed the *Full Feature Set* model, offering improved accuracy, interpretability, and resistance to overfitting. These models showed strong agreement with true concentrations on both held-out test data and fully blind samples, with predictions falling within tight 95% confidence intervals. The blind sample evaluation - conducted under conditions of sample anonymity - confirmed that the models generalize effectively to entirely unseen formulations, reinforcing their potential for deployment in quality control and in-process analytical settings.

The use of ΔSNV as the spectral representation, combined with fingerprint driven feature engineering, has allowed enhanced sensitivity to low-intensity signals while minimizing the influence of non-analyte background variation. Additionally, replicate-level quality control and confidence estimation allowed for uncertainty-aware predictions, which can be critical in pharmaceutical applications where decision-making often depends on precise quantification.

Future work will investigate whether the current model remains valid when using Surface-Enhanced Raman Scattering (SERS), or whether retraining is necessary to accommodate signal amplification effects, as well as how variations in matrix composition and differences in chemical structure across compound classes influence model performance and stability.

Acknowledgements

This work was supported in part by the MRID, project PNRR-I8 no 842027778, contract no 760096 and by the Core Program within the National Research, Development and Innovation Plan 2022-2027, carried out with the support of MRID, project no. 23020101(SIA-PRO), contract no 7N/2022.

References

[1] U. Neugebauer, P. Rösch, J. Popp, Raman spectroscopy towards clinical application: drug monitoring and pathogen identification, International Journal of Antimicrobial Agents 46 (2015) S35–S39. https://doi.org/10.1016/j.ijantimicag.2015.10.014.

[2] Y. Wang, C. Yu, H. Ji, Z. Liu, X. Wang, Y. Ji, X. Sun, Y. Zhao, X. Qiu, T. Zhang, J. Li, X. Liu, X. Lv, B. Cai, Y. Zhao, J.-A. Huang, Y. Li, Label-free therapeutic drug monitoring in human serum by the 3-step surface enhanced Raman spectroscopy and multivariate analysis, Chemical Engineering Journal 452 (2023) 139588. https://doi.org/10.1016/j.cej.2022.139588.

[3] A. Jaworska, S. Fornasaro, V. Sergo, A. Bonifacio, Potential of Surface Enhanced Raman Spectroscopy (SERS) in Therapeutic Drug Monitoring (TDM). A Critical Review, Biosensors 6 (2016) 47. https://doi.org/10.3390/bios6030047.

[4] K.A. Esmonde-White, M. Cuellar, I.R. Lewis, The role of Raman spectroscopy in biopharmaceuticals from development to manufacturing, Anal Bioanal Chem 414 (2022) 969–991. https://doi.org/10.1007/s00216-021-03727-4.

[5] F. Adar, Geiger ,Ruth, J. and Noonan, Raman Spectroscopy for Process/Quality Control, Applied Spectroscopy Reviews 32 (1997) 45–101. https://doi.org/10.1080/05704929508001133.

[6] Applications of Raman Spectroscopy in Biopharmaceutical Manufacturing: A Short Review - Kevin Buckley, Alan G. Ryder, 2017, (n.d.). https://journals.sagepub.com/doi/full/10.1177/0003702817703270 (accessed April 11, 2025).

[7] J. Peris-Vicente, E. Peris-García, J. Albiol-Chiva, A. Durgbanshi, E. Ochoa-Aranda, S. Carda-Broch, D. Bose, J. Esteve-Romero, Liquid chromatography, a valuable tool in the determination of antibiotics in biological, food and environmental samples, Microchemical Journal 177 (2022) 107309. https://doi.org/10.1016/j.microc.2022.107309.

[8] J.I.S. da S. de Jesus, R. Löbenberg, N.A. Bou-Chacra, Raman Spectroscopy for Quantitative Analysis in the Pharmaceutical Industry, J Pharm Pharm Sci 23 (2020) 24–46. https://doi.org/10.18433/jpps30649.

[9] T.S. Alstrøm, M.N. Schmidt, T. Rindzevicius, A. Boisen, J. Larsen, A pseudo-Voigt component model for high-resolution recovery of constituent spectra in Raman spectroscopy, in: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017: pp. 2317–2321. https://doi.org/10.1109/ICASSP.2017.7952570.

[10] M. Moores, K. Gracie, J. Carson, K. Faulds, D. Graham, M. Girolami, Bayesian modelling and quantification of Raman spectroscopy, (2018). https://doi.org/10.48550/arXiv.1604.07299.

[11] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, J. Pagès, Calibration in non-linear near infrared reflectance spectroscopy: a comparison of several methods, Analytica Chimica Acta 384 (1999) 207–214. https://doi.org/10.1016/S0003-2670(98)00814-9.

[12] S. Guo, J. Popp, T. Bocklitz, Chemometric analysis in Raman spectroscopy from experimental design to machine learning–based modeling, Nat Protoc 16 (2021) 5426–5459. https://doi.org/10.1038/s41596-021-00620-3.

[13] Y. Qi, D. Hu, Y. Jiang, Z. Wu, M. Zheng, E.X. Chen, Y. Liang, M.A. Sadi, K. Zhang, Y.P. Chen, Recent Progresses in Machine Learning Assisted Raman Spectroscopy, Advanced Optical Materials 11 (2023) 2203104. https://doi.org/10.1002/adom.202203104.

[14] U. Thissen, M. Pepers, B. Üstün, W.J. Melssen, L.M.C. Buydens, Comparing support vector machines to PLS for spectral regression applications, Chemometrics and Intelligent Laboratory Systems 73 (2004) 169–179. https://doi.org/10.1016/j.chemolab.2004.01.002.

[15] A. Silge, K. Weber, D. Cialla-May, L. Müller-Bötticher, D. Fischer, J. Popp, Trends in pharmaceutical analysis and quality control by modern Raman spectroscopic techniques, TrAC Trends in Analytical Chemistry 153 (2022) 116623. https://doi.org/10.1016/j.trac.2022.116623.

[16] J. Zhao, M.A. Short, T.A. Braun, H.L. M.d, D.I.M. M.d, H. Zeng, Clinical Raman measurements under special ambient lighting illumination, JBO 19 (2014) 111609. https://doi.org/10.1117/1.JBO.19.11.111609.

[17] M. Xiong, J. Ye, Reproducibility in surface-enhanced Raman spectroscopy, J. Shanghai Jiaotong Univ. (Sci.) 19 (2014) 681–690. https://doi.org/10.1007/s12204-014-1566-7.

[18] A. Averkiev, R.D. Rodriguez, M. Fatkullin, A. Lipovka, B. Yang, X. Jia, O. Kanoun, E. Sheremet, Towards solving the reproducibility crisis in surface-enhanced Raman spectroscopy-based pesticide detection, Science of The Total Environment 935 (2024) 173262. https://doi.org/10.1016/j.scitotenv.2024.173262.

[19] G. Socrates, Infrared and Raman Characteristic Group Frequencies: Tables and Charts, John Wiley & Sons, 2004.

[20] J.S. Böke, J. Popp, C. Krafft, Optical photothermal infrared spectroscopy with simultaneously acquired Raman spectroscopy for two-dimensional microplastic identification, Sci Rep 12 (2022) 18785. https://doi.org/10.1038/s41598-022-23318-2.

[21] S.F. Barakh Ali, Dharani ,Sathish, Afrooz ,Hamideh, Khan ,Mansoor A., Mohamed ,Eman M., Kohli ,Kanchan, Z. and Rahman, Application of salt engineering to reduce/mask bitter taste of clindamycin, Drug Development and Industrial Pharmacy 45 (2019) 1871–1878. https://doi.org/10.1080/03639045.2019.1672715.

[22] A. Bankapur, R.S. Krishnamurthy, E. Zachariah, C. Santhosh, B. Chougule, B. Praveen, M. Valiathan, D. Mathur, Micro-Raman Spectroscopy of Silver Nanoparticle Induced Stress on Optically-Trapped Stem Cells, PLoS One 7 (2012) e35075. https://doi.org/10.1371/journal.pone.0035075.

[23] Å. Rinnan, F. van den Berg, S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, TrAC Trends in Analytical Chemistry 28 (2009) 1201–1222. https://doi.org/10.1016/j.trac.2009.07.007.

[24] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra, Appl Spectrosc 43 (1989) 772–777. https://doi.org/10.1366/0003702894202201.

[25] R. teja Vulchi, V. Morgunov, R. Junjuri, T. Bocklitz, Artifacts and Anomalies in Raman Spectroscopy: A Review on Origins and Correction Procedures, Molecules 29 (2024) 4748. https://doi.org/10.3390/molecules29194748.