separations

MDPI

Article

Rapid Quantification of Low-Level Crystalline Impurities in Dalmelitinib Mesylate Using NIR Spectroscopy and Chemometric Modeling

Runxi Gui [1,2,†], Xiaogang Lian [1,3,†], Maolin Li [1,2], Mingdi Liu [1,4,*], Lina Zhou [1,2,*], Songgu Wu [1,2] and Qiuxiang Yin [1,2]

1 School of Chemical Engineering and Technology, National Engineering Research Center of Industrial Crystallization Technology, Tianjin University, Tianjin 300072, China
2 State Key Laboratory of Chemical Engineering and Low-Carbon Technology, Tianjin University, Tianjin 300072, China
3 Jiangsu Hansoh Pharmaceutical Group Co., Ltd., Lianyungang 222047, China
4 Key Laboratory of Resource Chemistry and Eco-Environmental Protection in Tibetan Plateau, State Ethnic Affairs Commission, Xining 810007, China
* Correspondence: liumingdi@tju.edu.cn (M.L.); linazhou@tju.edu.cn (L.Z.)
† These authors contributed equally to this work.

Abstract

Accurate measurement and control of impurities are critical for ensuring the quality and therapeutic performance of solid-state pharmaceutical formulations. This study introduces a rapid, minimal sample preparation analytical approach for quantifying low-level dalmelitinib impurities in dalmelitinib mesylate, employing near-infrared (NIR) spectroscopy combined with partial least squares regression (PLSR). To mimic actual manufacturing conditions, a mixture system was designed comprising dalmelitinib mesylate, dalmelitinib impurity, and formulation excipients. Various spectral preprocessing strategies were systematically evaluated, including Savitzky–Golay first derivative (SG1st), Savitzky–Golay second derivative (SG2nd), multiplicative scatter correction (MSC), standard normal variate (SNV), wavelet denoising, wavelet compression, and their combinations. The optimal model was obtained using SG1st combined with wavelet denoising. The resulting PLSR model (7 latent variables) showed good predictive performance, with an R^2 of 0.99569 and an RMSECV of 0.00315. The limit of detection (LOD) and limit of quantification (LOQ) were 0.234% and 0.708%, respectively, demonstrating applicability for monitoring low-level impurities in pharmaceutical formulations. Method validation demonstrated satisfactory precision (RSD < 3%), accuracy (100.77–102.01%), and stability over 24 h (RSD $\leq$ 4.75%). Compared with conventional solid-state analytical techniques, the proposed NIR–PLSR framework enables rapid, non-destructive analysis with minimal sample preparation. The combination of derivative preprocessing and wavelet denoising improved extraction of impurity-related spectral information in complex pharmaceutical systems, highlighting the potential of this approach for process analytical technology (PAT) and pharmaceutical quality monitoring.

Keywords: near-infrared spectroscopy; dalmelitinib mesylate; partial least squares regression; low-level impurity quantification

Check for updates

Academic Editor: Gavino Sanna

Received: 22 April 2026
Revised: 2 June 2026
Accepted: 5 June 2026
Published: 9 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Drug impurities are critical quality attributes in pharmaceutical products because they may affect drug safety, efficacy, and stability. Such impurities can originate from residual starting materials, degradation products, manufacturing processes, and solid-state transformations occurring during formulation development or storage. This issue is particularly relevant for organic salt-based drug molecules, where exposure to solvents, humidity, temperature fluctuations, or mechanical stress may induce salt dissociation reactions, leading to regeneration of free acid or free base forms. Therefore, accurate quantification of low-level impurities is essential for robust pharmaceutical quality control and regulatory compliance.

Dalmelitinib is a new innovative anti-tumor drug currently under clinical investigation. However, environmental stress during pharmaceutical processing and storage may induce salt dissociation of DMM (The structure of DMM is shown in Figure 1), resulting in formation of dalmelitinib impurity (DML). Such solid-state transformations may alter dissolution behavior and contribute to variability in formulation performance. Consequently, establishing a rapid, sensitive, and reliable analytical method for DML quantification is of considerable practical importance.

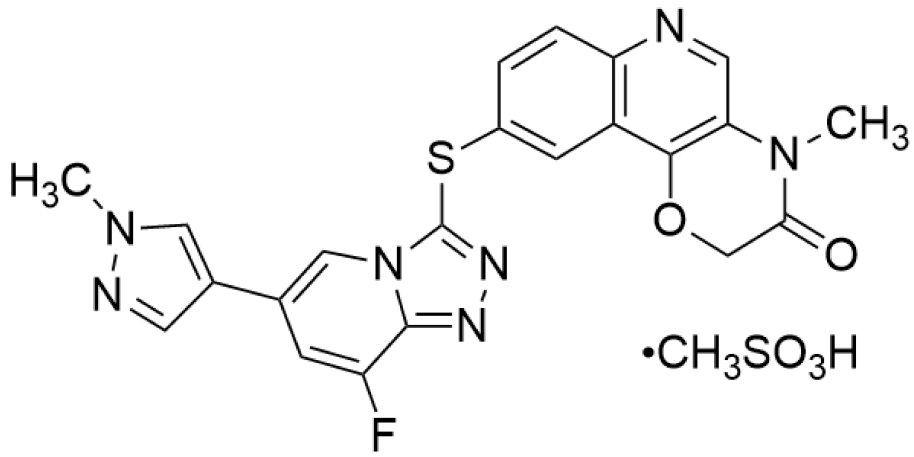


Figure 1. Chemical structure of dalmelitinib mesylate.

Conventional analytical techniques for crystalline pharmaceutical characterization mainly include powder X-ray diffraction (PXRD), Raman spectroscopy, thermal analysis, and chromatographic methods [1–7]. PXRD is widely recognized as the gold standard for polymorph identification because of its excellent structural specificity [4,8–11]; however, quantitative determination of low-level impurities remains challenging owing to peak overlap, weak diffraction signals at low concentrations, and limited sensitivity in multi-component systems. Raman spectroscopy provides molecular fingerprint information but may suffer from fluorescence interference and limited penetration depth [4,12–17]. Thermal analysis reveals phase transitions but generally lacks quantitative capability [4,18–20], whereas chromatographic methods often require destructive sample preparation and labor-intensive procedures [21]. Despite their strengths in structural characterization, these approaches exhibit limited applicability for rapid quantification of low-level impurities in complex pharmaceutical systems.

Compared with conventional techniques, near-infrared (NIR) spectroscopy offers rapid, minimal sample preparation analysis with minimal sample preparation and compatibility with process analytical technology (PAT) [15,22–25]. Owing to deeper penetration into solid matrices, NIR spectroscopy is particularly suitable for intact pharmaceutical dosage forms and bulk powders [2,13,14,26]. More importantly, integration with chemometric

methods such as partial least squares regression (PLSR) enables extraction of subtle spectral variations associated with compositional changes, demonstrating considerable potential for quantification of low-level components in complex pharmaceutical systems [27–29].

Nevertheless, NIR spectroscopy is generally regarded as an indirect analytical technique whose performance strongly depends on appropriate preprocessing and robust chemometric modeling. In complex formulations, severe spectral overlap among APIs, impurities, and excipients may obscure impurity-related information, thereby reducing model selectivity and predictive accuracy [24,30]. Existing NIR–PLSR studies have mainly focused on binary or relatively simple systems, whereas quantitative determination of low-level impurities in complex multicomponent pharmaceutical matrices remains insufficiently explored [31,32]. To address these challenges, preprocessing strategies including multiplicative scatter correction (MSC), standard normal variate (SNV), Savitzky–Golay derivatives, and wavelet transform techniques have been introduced to enhance subtle spectral features and suppress irrelevant variations [33–36].

Therefore, this study aimed to establish and validate an optimized NIR–PLSR analytical framework for sensitive quantification of DML in DMM formulations. Unlike conventional approaches relying on single preprocessing strategies, derivative preprocessing and wavelet denoising were systematically integrated to enhance weak impurity-related spectral information in a complex multicomponent pharmaceutical system. By comparing different preprocessing approaches and evaluating model performance using statistical and validation criteria, an optimized quantitative model was developed. The proposed strategy may contribute to pharmaceutical quality control and provide a general analytical reference for low-level impurity monitoring in complex solid dosage forms.

2. Materials and Methods

2.1. Materials

DMM, DML, excipients, DMM tablets, and stability samples were all provided by Jiangsu Hansoh Pharmaceutical Group Research Institute (Lianyungang, China).

2.2. Sample Preparation

2.2.1. Raw Material Pretreatment

To minimize measurement errors arising from particle size variations, all raw materials were pretreated prior to analysis. DMM, DML, and excipients were gently ground in an agate mortar, followed by sieving through a 100-mesh sieve, and the fraction passing through the sieve was collected.

2.2.2. Tablet and Stability Sample Preparation

Test tablets and stability samples were ground into fine powder using an agate mortar. The powder was subsequently sieved through a 100-mesh sieve, and the fraction passing through the sieve was collected.

2.2.3. Preparation of Calibration Samples

Calibration samples were prepared according to the following standardized procedure. Twenty-six calibration samples with systematically distributed DML concentrations ranging from 0–20% were designed using a concentration-gradient strategy rather than random sampling, ensuring representative coverage across low-, medium-, and high-level impurity regions within the calibration space. Accurately weighed quantities of DMM, DML, and excipients were combined in the mass ratios outlined in Table S1. The composition of each sample was adjusted to achieve varying impurity concentrations while maintaining a constant total sample mass. All components were transferred into a stainless steel ball

milling jar and homogenized using a ball mill (MM400, Retsch, Haan, Germany) at a frequency of 30 Hz for 25 min. The resulting mixtures were collected and transferred into 10 mL round-bottom plastic centrifuge tubes for subsequent analysis. To ensure uniform particle size distribution and compositional homogeneity, all samples were subjected to identical grinding, sieving, and homogenization procedures.

2.3. Solid-State Characterization

2.3.1. Powder X-Ray Diffraction (PXRD)

The PXRD patterns were acquired using a Rigaku Miniflex 600 powder diffraction (Rigaku Co., Tokyo, Japan) equipped with Cu Ka radiation (λ = 1.5405Å). The instrument operated at an optimal ambient condition, with an operating voltage of 40 kV and current of 15 mA. The data were collected over a 2θ range of 2–35° with a step size of 0.02° and a scan rate of $8°\cdot min^{-1}$. All samples underwent three meticulous rounds of PXRD analysis.

2.3.2. Differential Scanning Calorimetry (DSC)

The DSC measurements were performed using a Mettler DSC system (Mettler Toledo DSC1, Greifensee, Switzerland) with a precisely controlled nitrogen flow rate of $50\ mL\cdot min^{-1}$. Each sample of a defined mass was heated from 25 °C to 300 °C at a carefully regulated rate of $10\ °C\cdot min^{-1}$ and tested in triplicate.

2.3.3. Thermogravimetric Analysis (TGA)

TGA measurements were performed using a Mettler TGA system (Mettler Toledo TGA3^{+}, Greifensee, Switzerland) with nitrogen flow precisely controlled at $50\ mL\cdot min^{-1}$. Each sample of a defined mass was heated from 25 °C to 400 °C at a strictly regulated rate of $10\ °C\cdot min^{-1}$, with each test conducted in triplicate.

2.4. NIR Spectral Acquisition

The NIR spectra were analyzed using a near-infrared spectrometer (Thermo Fisher Scientific Antaris II, Waltham, MA, USA) with a resolution of 4 cm^{-1} and a wavenumber interval of 1 cm^{-1}, in the range of 10,000 cm^{-1} to 4000 cm^{-1}. Each sample was scanned at different positions under rotation to collect a total of nine spectra (in three sets, with three scans per set). The average spectrum was then calculated and used as the final spectrum of the sample to minimize sampling errors and instrumental noise.

2.5. Spectral Preprocessing and PLSR Modeling

To achieve precise quantitative determination of DML in DMM tablet, a total of 26 calibration samples were prepared, with impurity contents ranging from 0% to 20%. This range was designed based on the possible distribution of impurities in real pharmaceutical formulations, encompassing low-level as well as higher concentration ranges to promote the model's broad applicability and generalizability. Near-infrared (NIR) spectra are susceptible to non-chemical variations such as particle size distribution, sample packing density, light scattering effects, and instrumental noise, which may introduce redundant information and reduce model predictive performance. To mitigate these effects and accentuate chemically relevant spectral features, preprocessing is essential. In this study, multiple spectral preprocessing methods were systematically evaluated, including multiplicative scatter correction (MSC), standard normal variate (SNV), Savitzky–Golay first derivative (SG1st), Savitzky–Golay second derivative (SG2nd), wavelet denoising, wavelet compression and their combinations. These methods improve spectral quality by correcting scattering effects, eliminating baseline drift, enhancing spectral resolution, and suppressing high-frequency noise.

Based on the preprocessed spectral data, partial least squares regression (PLSR) models were established to correlate NIR spectra with DML concentrations. All spectral data processing and model development were performed using MATLAB (R2014a, MathWorks, Natick, MA, USA). The optimal number of latent variables was determined by cross-validation to prevent overfitting and improve model robustness. The calibration model performance was evaluated using the coefficient of determination (R^2), root mean square error of calibration (RMSEC), root mean square error of cross-validation (RMSECV), and root mean square error of prediction (RMSEP). Higher R^2 values and lower RMSE values indicate better predictive performance of the model.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (1)$$

where y_i, $\hat{y}_i$, and n in Equation (1) represented theoretical value, calculated value, and the number of samples, respectively.

2.6. Model Validation

To assess the reliability and practical applicability of the developed PLSR model, validation experiments were conducted to evaluate its precision, accuracy, and stability using reference samples with predetermined levels of DML. Precision and accuracy were quantified by analyzing standard samples spiked with 2.0% and 3.0% DML, each in six replicate measurements. NIR spectra were collected under the same experimental conditions, and impurity contents were predicted using the established PLSR model. To evaluate sample stability and model robustness over time, standard samples containing 1.0% and 1.5% DML were analyzed at different time intervals (0, 4, 6, 8, 12, and 24 h).The relative standard deviation (RSD%) was calculated for each PLSR model according to Equation (2), while the limit of detection (LOD) and limit of quantification (LOQ) were determined for each PLSR model utilizing Equations (3) and (4),

$$\mathrm{RSD} = \left(\sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{n-1}} / \bar{y}\right) \times 100\% \quad (2)$$

$$\mathrm{LOD} = 3.3\sigma \quad (3)$$

$$\mathrm{LOQ} = 10\sigma \quad (4)$$

where $\hat{y}_i$, $\bar{y}$, n, σ, s in Equations (2)–(4) were predicted content values, average of the predicted content values of the samples, number of samples, standard deviation of predicted content values, and slope of the calibration curve, respectively.

3. Results and Discussion

3.1. Characterization of DMM, DML and Excipients

The PXRD patterns of DMM, DML and excipients are presented in Figure 2. Significant differences in diffraction features can be observed. The crystalline components exhibit distinct diffraction peaks at different 2θ regions, whereas most excipients display broad amorphous halos with only limited sharp diffraction signals. These pronounced differences indicate clear structural distinctions among the three components and provide a solid basis for subsequent analysis.

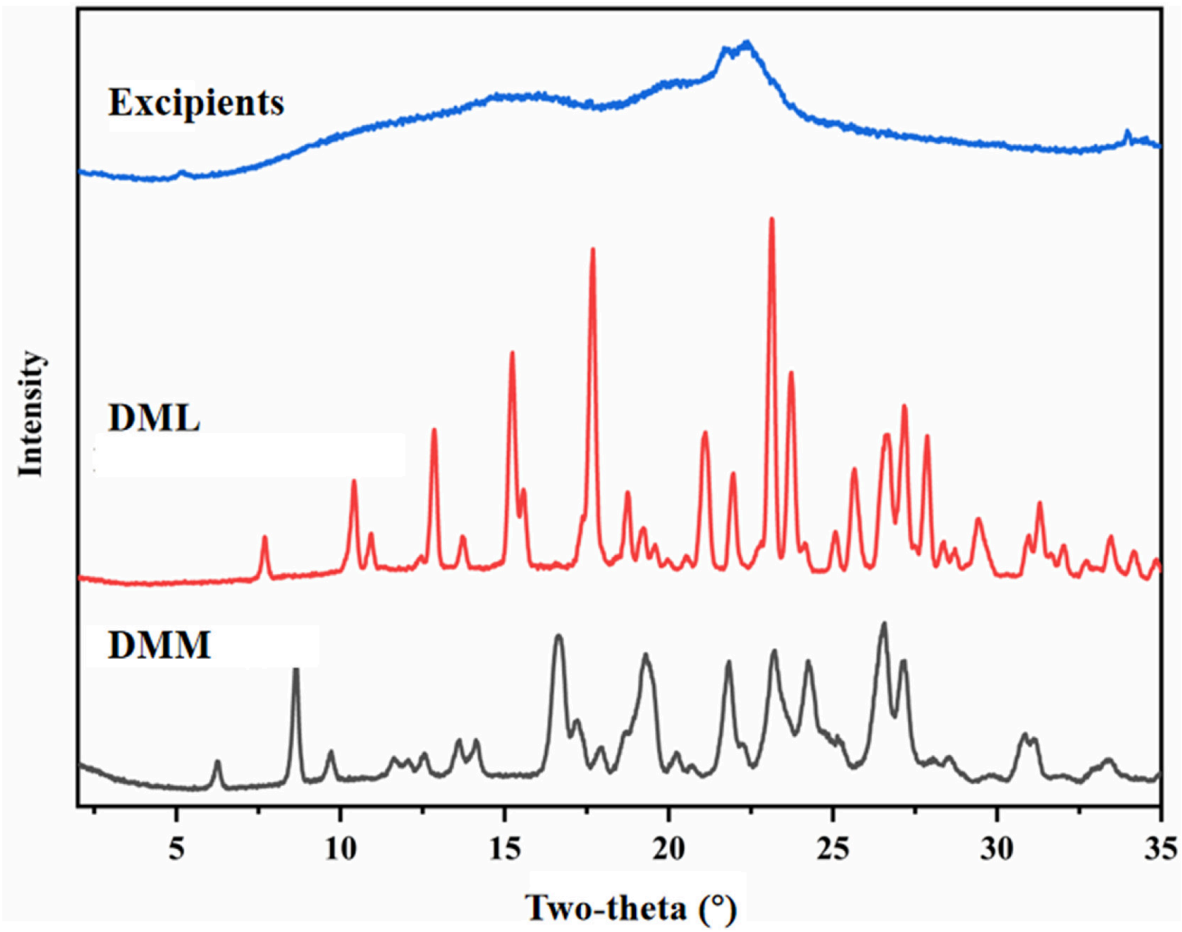


Figure 2. PXRD pattern of different DMM solid forms.

The TGA curves of DMM, DML and excipients are presented in Figure 3. Among these, DMM begins weight loss at 250 °C. The DML starts weight loss around 325 °C, while the excipients exhibit weight loss from the initial 25 °C, with another weight loss peak occurring at 275 °C, resulting in a two-segment weight loss curve.

The DSC curves of DMM, DML and excipients are presented in Figure 4. The DMM has a melting point of 262.2 °C and a phase transition enthalpy of −155.17 J/g. DML has a melting point of 276.18 °C and a phase change enthalpy of −130.68 J/g. The excipients have a phase transform at 52.96 °C and the enthalpy of −156.35 J/g. The significant difference in melting points further confirms that DMM and DML correspond to distinct solid forms.

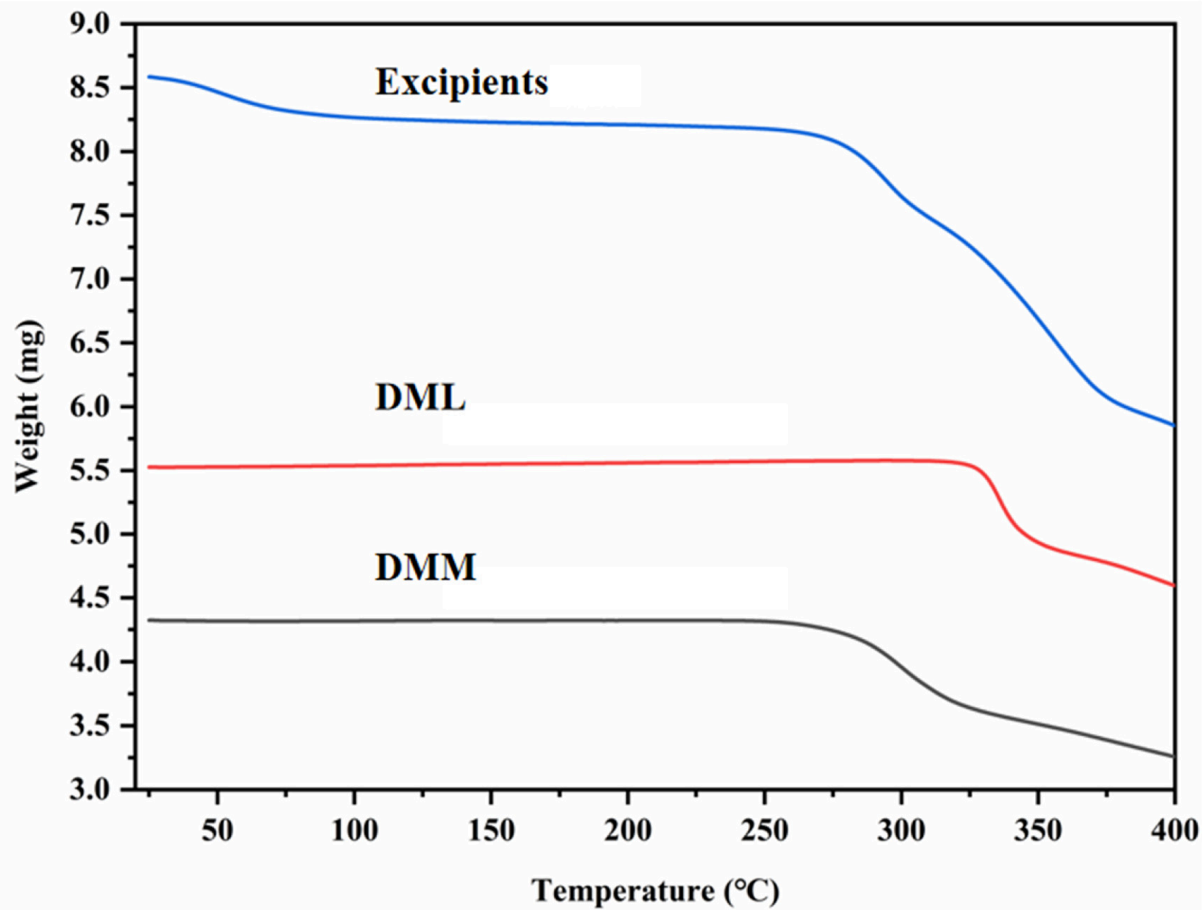


Figure 3. TGA curves of different DMM solid forms.

The near-infrared spectra (NIR) of DMM, DML, and excipients are presented in Figure 5. Characteristic absorption bands observed in the spectra are mainly associated with overtone and combination vibrations of C–H, O–H, and N–H functional groups. For DMM, characteristic absorption peaks located at 8549.0 cm^{-1} and 7258.4 cm^{-1} are mainly attributed to the second overtone and combination vibrations of C–H stretching modes, while the absorption band at 6047.8 cm^{-1} may be associated with O–H/N–H related combination vibrations. In contrast, DML exhibits distinct absorption features at 8782.1 cm^{-1} and 5914.7 cm^{-1}, indicating differences in molecular environment and intermolecular interactions. The excipients show characteristic absorption bands at 6773.3 cm^{-1} and 5172.4 cm^{-1}, which are distinguishable from those of DMM and DML.

Although partial spectral overlap exists due to the complexity of the formulation matrix, these characteristic spectral variations reflect chemically meaningful differences among DMM, DML, and excipients. Combined with spectral preprocessing and multivariate PLSR analysis, the relevant spectral information could be effectively extracted to enable selective quantification of low-level DML impurities in pharmaceutical formulations. However, the current model was developed and validated using a single impurity system (DML only), and its applicability to simultaneous quantification of multiple impurities remains to be explored in future studies.

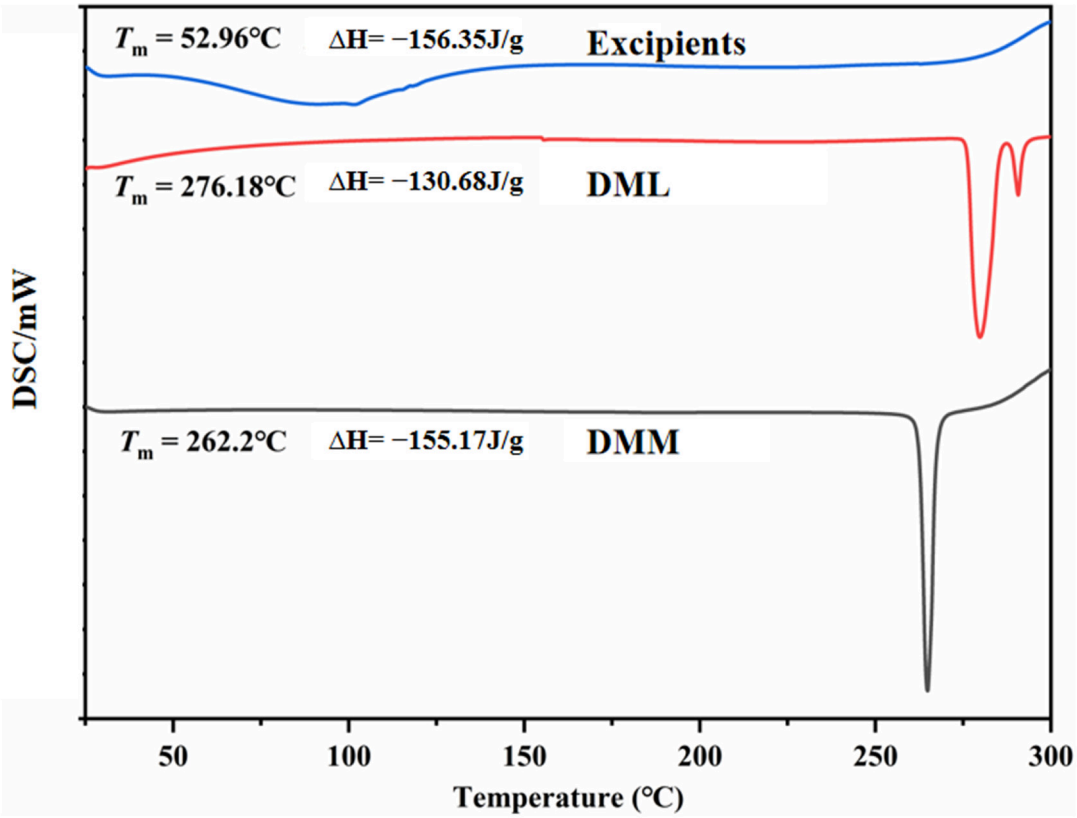


Figure 4. DSC curves of different DMM solid forms.

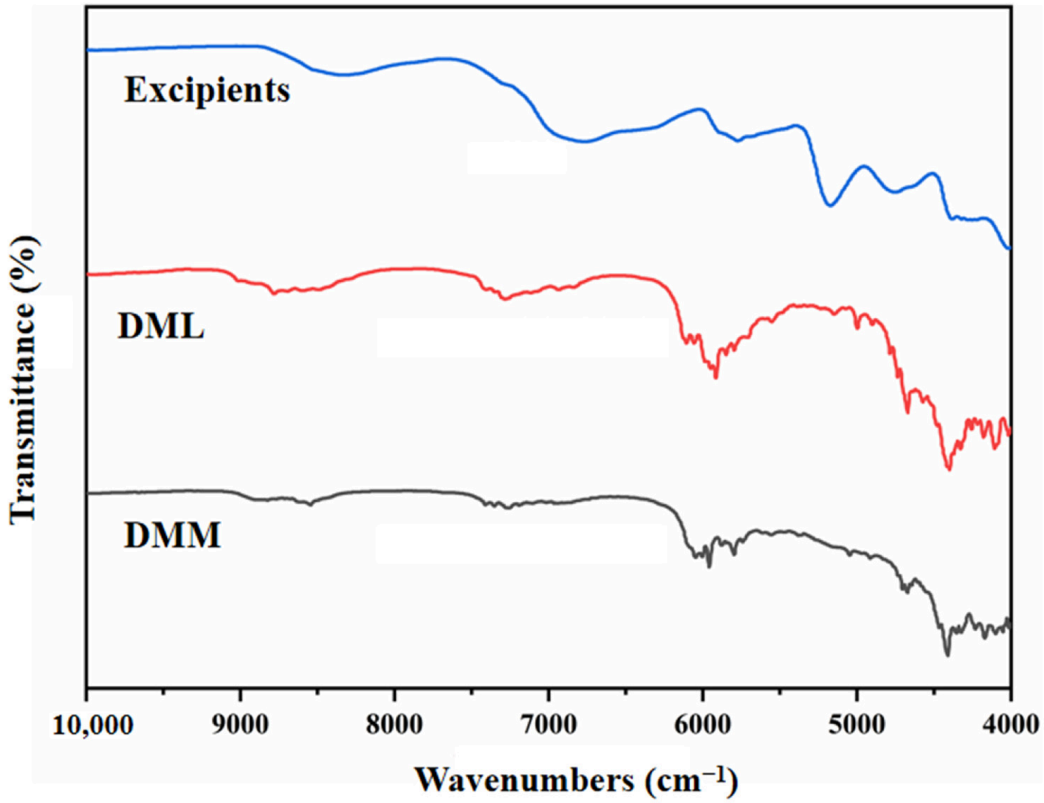


Figure 5. NIR spectra of different DMM solid forms.

3.2. Analysis of NIR Spectral Characteristics and Pretreatment Effect

The original NIR spectra of the 26 mixed samples are shown in Figure 6. It can be observed that, with increasing impurity concentration, the spectra exhibit systematic variations in specific regions (e.g., around 8782 cm^{-1} and 5914 cm^{-1}). However, these concentration-dependent changes are partially obscured by strong background signals and spectral overlap, resulting in only subtle visible differences among samples.

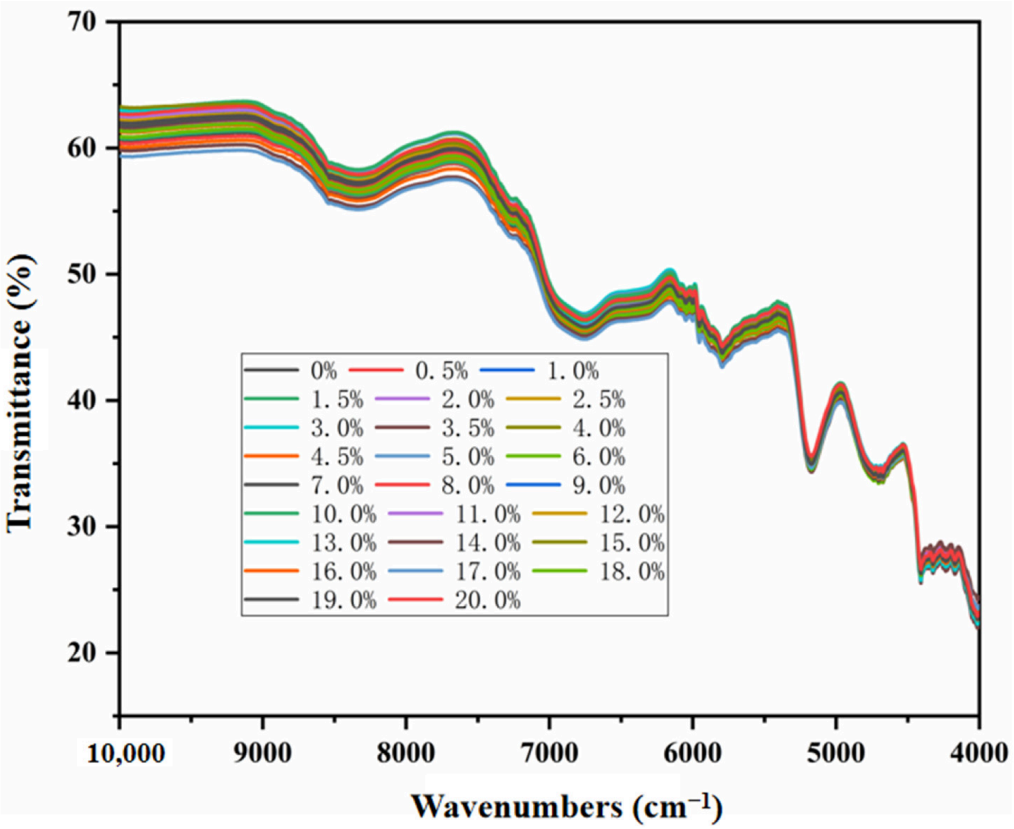


Figure 6. Near-infrared spectra of DMM, DML, and excipients mixture systems (from top to bottom: DML content 0.0–20.0%).

Considering the solid-state transformation between DMM and DML involving mesylate salt dissociation and changes in proton-transfer interactions, the observed spectral variations may be associated with alterations in hydrogen-bonding environments as well as overtone and combination vibrations related to N–H, O–H, and sulfonate-containing functional groups. Although these spectral differences are relatively subtle due to the low impurity content and spectral overlap within complex pharmaceutical matrices, they can still be effectively captured through appropriate spectral preprocessing and multivariate chemometric analysis.

3.3. NIR Spectral Pretreatment and PLSR Model Development

Raw NIR spectra obtained from pharmaceutical formulations are commonly affected by baseline drift, light scattering, particle-size heterogeneity, and instrumental noise, which may obscure subtle spectral differences associated with low-level crystalline impurities. Therefore, spectral pretreatment is essential to improve spectral quality and enhance extraction of chemically relevant information prior to chemometric analysis.

To optimize model performance, several preprocessing methods, including MSC, SNV, SG1, SG2, and wavelet denoising, were systematically evaluated individually and in combination. The corresponding preprocessing spectra are shown in Figures S1–S7. Among

these approaches, SG1 combined with wavelet denoising exhibited the most favorable overall analytical performance.

As shown in Figure 7, SG1 preprocessing effectively corrected baseline variation and enhanced subtle spectral features, while wavelet denoising reduced high-frequency noise and improved spectral smoothness without significant loss of chemical information. After preprocessing, the spectral differences associated with DML impurities became more distinguishable, facilitating subsequent multivariate analysis.

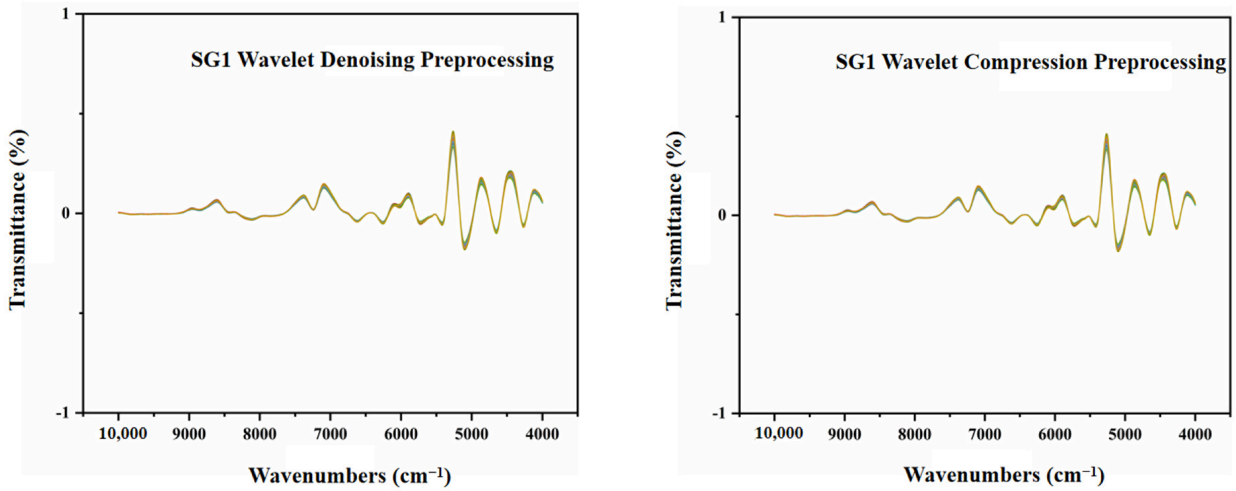


Figure 7. NIR spectra of DML after SG1 wavelet denoising (**left**) and SG1 wavelet compression (**right**) preprocessing.

Based on the optimized preprocessed spectra, a PLSR calibration model was established for quantitative analysis of DML. The calibration results are presented in Figure 8. The developed model demonstrated good linear correlation between predicted and reference values, indicating that impurity-related spectral variations could be effectively extracted from the complex formulation matrix containing DMM and excipients.

The predictive performances of different preprocessing strategies are summarized in Table 1. Models established using raw spectra exhibited relatively poor predictive ability, suggesting that baseline variation and spectral interference negatively affected quantitative analysis. Although certain preprocessing methods yielded lower calibration errors, these models generally required a substantially larger number of latent variables, indicating increased risk of overfitting and reduced model robustness.

Compared with other evaluated preprocessing approaches, the SG1 + wavelet denoising model achieved a more balanced compromise between predictive accuracy, model stability, and generalizability, with satisfactory RMSECV, RMSEP, and correlation coefficients. These results demonstrate that appropriate preprocessing plays a critical role in improving the reliability of NIR-based quantitative analysis for low-level crystalline impurities.

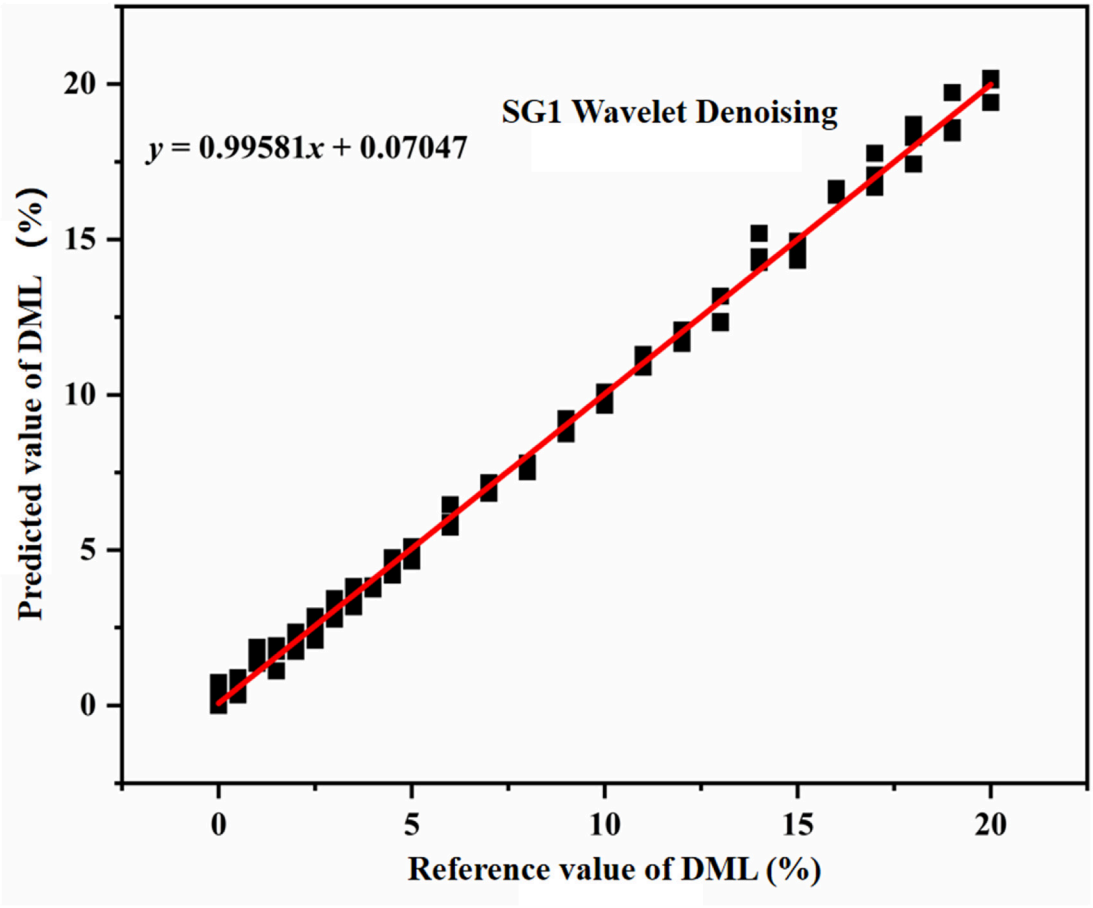


Figure 8. NIR calibration models for the quantitative analysis of DML using SG1 combined with wavelet denoising preprocessing.

Table 1. Summary of model parameters for the quantitative analysis of DML based on NIR spectroscopy.

Model	Number of Latent Variables	Cross-Validation Coefficient	RMSECV	RMSECP	RMSEC	R^2 (Calibration)	LOD	LOQ
Mean	7	−0.23908	0.00090	0.64647	0.80205	0.983	0.53937	1.63448
MSC	10	−0.04661	0.00626	0.07236	0.06364	0.998	0.04951	0.15003
SNV	11	0.01927	0.00741	0.07943	0.23711	0.999	0.05076	0.15384
SG1	18	0.05543	0.00100	0.20825	0.09813	1.000	0.00584	0.01771
SG2	17	0.01405	0.00138	0.30606	0.08734	1.000	0.01297	0.03930
Denoising	3	−0.05909	0.16217	3.99123	4.77439	0.412	39.49086	119.66928
Compression	3	−0.05902	0.16211	3.99109	4.77441	0.412	39.49046	119.66807
MSC + Denoising	11	−0.03541	0.01175	0.35449	0.32318	0.997	0.08494	0.25741
MSC + Compression	9	0.09486	0.01219	0.62216	0.54002	0.992	0.22311	0.67609
SNV + Denoising	11	−0.05399	0.01599	0.35311	0.32695	0.997	0.08969	0.27178
SNV + Compression	9	0.09457	0.01204	0.61687	0.54076	0.992	0.22454	0.68042
SG1 + Denoising	7	−0.27293	0.00315	0.51721	0.39714	0.996	0.23352	0.70766
SG1 + Compression	7	−0.24965	0.00305	0.48201	0.37697	0.995	0.22125	0.67046
SG2 + Denoising	5	0.03998	0.00314	0.79835	0.68863	0.988	0.49804	1.50921
SG2 + Compression	5	0.02495	0.00213	0.81944	0.67444	0.986	0.46715	1.41560

To further assess model predictive capability, the ratio of performance to deviation (RPD) was calculated using the standard deviation of reference concentrations and corresponding RMSECP values. The external prediction set covered DML concentrations ranging from 0.5% to 12.0%. The optimal SG1-wavelet model achieved an RPD value of 8.12, indicating excellent predictive performance according to commonly accepted chemometric criteria (RPD > 3: good prediction; RPD > 5: excellent prediction).

3.4. Validation and Application of the Optimal NIR Quantitative Model

To further independently verify the reliability of the developed NIR–PLSR model, an orthogonal PXRD-based quantitative approach was employed. Detailed methodology and validation results are provided in the Supporting Information (Figure S15 and Table S2). The PXRD analysis demonstrated acceptable specificity, linearity, precision and accuracy for low-level DML quantification. Compared with PXRD, the optimized NIR–PLSR model exhibited improved sensitivity while maintaining comparable analytical performance.

To evaluate the reliability and practical applicability of the developed near-infrared (NIR) quantitative model, a comprehensive method validation was performed using standard samples with known concentrations of DML in DMM. The performance of the optimal Partial Least Squares Regression (PLSR) model (SG1 combined with wavelet denoising) was systematically assessed in terms of linearity, precision, accuracy, sensitivity, and stability.

3.4.1. Linearity and Range

The optimal model exhibited excellent linearity over the concentration range of 0–20%, with a coefficient of determination (R^2) of 0.99569, indicating a strong correlation between predicted and reference values and satisfying the requirements for quantitative analysis.

3.4.2. Precision and Accuracy

Precision and accuracy were evaluated by performing six replicate measurements on samples containing 2.0% and 3.0% impurity levels (as shown in Figure 9). As summarized in Table 2, the relative standard deviations (RSDs) were 2.41% and 2.75%, respectively, both well below the acceptable limit of 15%. The corresponding accuracies, expressed as the ratio of predicted to actual values, were 102.01% and 100.77%, respectively, falling within the acceptable range of 80–120%. These results demonstrate that the proposed method provides satisfactory precision and accuracy for the quantitative determination of impurities.

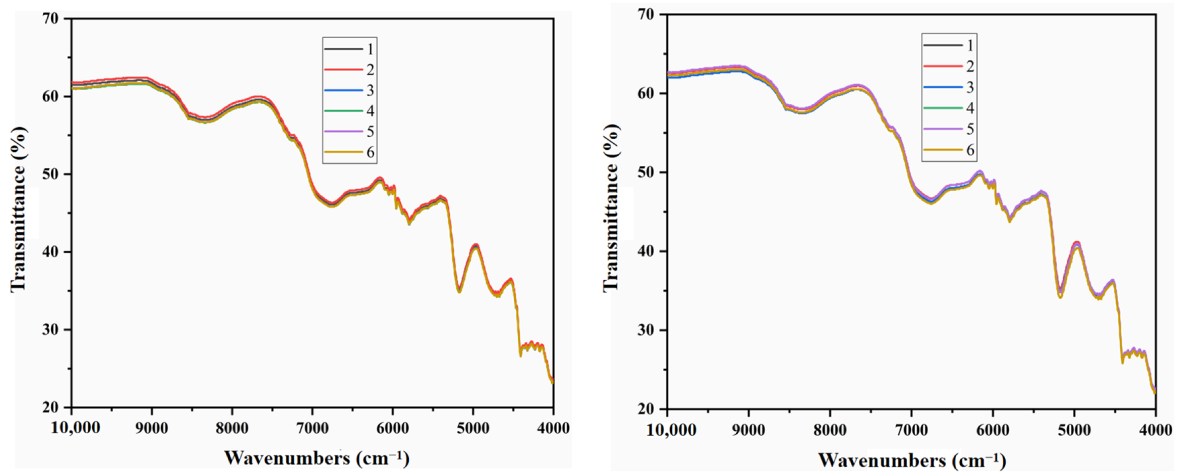


Figure 9. NIR spectra obtained from six replicate measurements for the repeatability assessment of DML content (left: 2.0%; right: 3.0%).

Table 2. Precision and accuracy results for DML based on the optimal model.

No.	2.0% Sample (%)	3.0% Sample (%)
1	2.0426	3.0362
2	2.0366	2.9935
3	1.9842	3.0928
4	2.0624	2.9935
5	1.9946	3.1282
6	2.1202	2.8936
Mean	2.0401	3.0230
Standard deviation	0.0492	0.0832
RSD %	2.41	2.75
Accuracy %	102.01	100.77

3.4.3. Limit of Detection and Quantification

The limit of detection (LOD) and limit of quantification (LOQ) of the optimal model were determined to be 0.234% and 0.708%, respectively. Compared with reported PXRD quantification ranges for certain polymorphic systems, the proposed NIR method exhibited improved sensitivity and demonstrated applicability for monitoring low-level impurities in pharmaceutical formulations.

However, according to ICH Q3B(R2), impurity reporting and identification thresholds for pharmaceutical products may range from approximately 0.1% to 0.2%, depending on the maximum daily dose. Therefore, although the developed model provides effective monitoring capability for low-level impurities, its current analytical sensitivity may not satisfy stringent regulatory requirements associated with impurity identification in all scenarios. In addition, ICH Q6A highlights the importance of controlling critical solid-state attributes when they may influence product quality, stability, or performance. Since the present work focuses on monitoring the solid-state transformation of DMM to DML during manufacturing and storage, the proposed method provides a practical analytical approach for evaluating such changes. Although further improvements in sensitivity may broaden its applicability, the developed NIR–PLSR model offers a rapid, non-destructive, and efficient tool for stability assessment, process monitoring, and pharmaceutical quality control.

Although the achieved LOQ may limit direct application for stringent regulatory impurity testing, further optimization of spectral preprocessing strategies and model design may improve analytical sensitivity in future studies.

The analytical performance achieved in this study compares favorably with previously reported NIR-based methods for solid-state pharmaceutical analysis. For example, da Silva et al. reported quantitative analysis of mebendazole polymorphs using NIR spectroscopy with correlation coefficients above 0.99 [22], while Heinz et al. demonstrated quantitative determination of indomethacin solid-state forms using NIR and Raman spectroscopy [13]. More recently, Liu et al. reported NIR–PLSR methods for low-content polymorphic impurity quantification in canagliflozin tablets [35,36]. Compared with these studies, the present work addresses a more challenging analytical scenario involving low-level crystalline impurities in a multicomponent pharmaceutical formulation containing active ingredient, impurity phase, and excipients. Despite the severe spectral overlap inherent to such systems, the developed model achieved satisfactory linearity, precision, and sensitivity, demonstrating the feasibility of NIR spectroscopy for rapid monitoring of solid-state transformations.

3.4.4. Stability

The stability of the method was evaluated by analyzing samples containing 1.0% and 1.5% impurity levels over a 24 h period (as shown in Figure 10). The results (Table 3) show

that the RSD values were 2.57% and 4.75%, respectively, while the corresponding accuracies were 102.84% and 105.10%. These findings indicate that the spectral signals remained stable over time and that the model is robust against minor variations in sample conditions, supporting its suitability for routine analysis and stability studies.

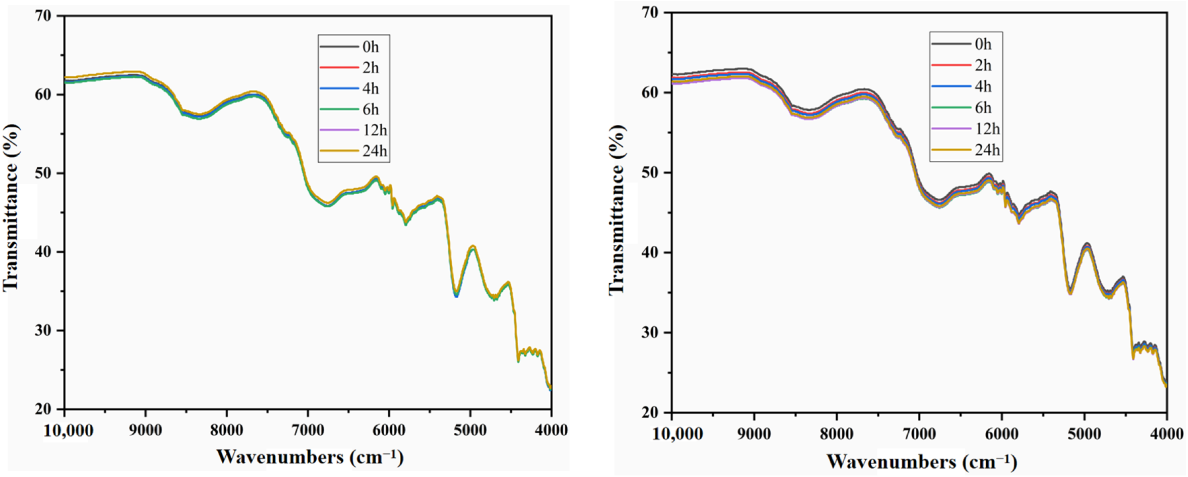


Figure 10. NIR spectra obtained during the stability study of DML content (left: 1.0%; right: 1.5%).

Table 3. Stability results for DML based on the optimal model.

Time (h)	1.0% Sample (%)	1.5% Sample (%)
0	1.0423	1.5623
4	1.0102	1.5042
6	1.0236	1.5609
8	0.9890	1.5032
12	1.0423	1.6362
24	1.0630	1.6923
Mean	1.0284	1.5765
Standard deviation	0.0264	0.0748
RSD %	2.57	4.75
Accuracy %	102.84	105.10

3.4.5. Application to Real Samples

The validated optimal NIR model was subsequently applied to evaluate authentic formulation samples, comprising six production batches (Lots 1–6) and three stability batches (Lots 7–9, subjected to long-term storage for 6 months). The results are shown in Table 4.

Table 4. Predicted content of the DML in different DMM formulation samples by NIR spectroscopy.

Lot	1	2	3	4	5	6	7	8	9
SG1 + Wavelet denois-ing	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.06754	1.29325	0.00000

No 0. and 1.29%, suggesting that transformation may occur during storage. This application demonstrates that the developed NIR method is not only suitable for rapid release testing but also serves as an effective tool for monitoring product stability and supporting formulation development, packaging selection, and shelf-life determination.

3.4.6. Model Specificity and Limitations

The developed model was established using a formulation-relevant three-component system comprising DMM, DML, and excipients. The results indicate that the preprocessing strategies combined with PLSR modeling enabled selective quantification of DML despite overlapping NIR spectral features.

Nevertheless, the present study only considered one impurity species and did not investigate simultaneous quantification in multi-impurity systems. In addition, variable importance analysis (e.g., VIP scores) and regression coefficient interpretation were not investigated in the present study. Future work may incorporate such analyses to improve mechanistic understanding of spectral contributions and enhance model interpretability. Therefore, the applicability of the proposed model to more complex impurity profiles requires further validation. Future work should explore extension toward multiple impurities and evaluate model robustness under broader compositional variability.

4. Conclusions

In this study, a NIR spectroscopy method combined with chemometric modeling was successfully developed for the quantitative determination of DML in DMM formulations. Systematic solid-state characterization confirmed distinct differences among DMM, DML and excipients, providing a foundation for subsequent NIR-based quantitative analysis. Through comprehensive comparison of preprocessing strategies, the combination of Savitzky–Golay first derivative and wavelet denoising coupled with PLSR was identified as the optimal modeling approach.

The validated model demonstrated excellent linearity (R^2 = 0.99569), high precision (RSD < 3%), satisfactory accuracy (100.77–102.01%), and acceptable sensitivity for low-level impurity monitoring (LOD = 0.234%, LOQ = 0.708%). The developed method showed acceptable analytical performance under the tested conditions and exhibited potential utility for rapid screening and monitoring of low-level impurities.

Furthermore, application to real samples confirmed the capability of the method to detect impurity formation during stability studies, highlighting its potential value in pharmaceutical quality control. In summary, this work presents a rapid, non-destructive, and reliable analytical tool requiring minimal sample preparation for monitoring low-level impurities, with potential applicability in process analytical technology (PAT) and quality-by-design (QbD) strategies.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/separations13060170/s1, Figure S1: NIR spectra of DML after mean preprocessing; Figure S2: NIR spectra of DML after MSC (left) and SNV (right) preprocessing; Figure S3: NIR spectra of DML after SG1 (left) and SG2 (right) preprocessing; Figure S4: NIR spectra of DML after wavelet denoising (left) and wavelet compression (right) preprocessing; Figure S5: NIR spectra of DML after MSC wavelet denoising (left) and MSC wavelet compression (right) preprocessing; Figure S6: NIR spectra of DML after SNV wavelet denoising (left) and SNV wavelet compression (right) preprocessing; Figure S7: NIR spectra of DML after SG2 wavelet denoising (left) and SG2 wavelet compression (right) preprocessing; Figure S8: NIR calibration models for the quantitative analysis of DML using mean preprocessing (left) and MSC preprocessing (right); Figure S9: NIR calibration models for the quantitative analysis of DML using SNV preprocessing (left) and SG1 preprocessing (right); Figure S10: NIR calibration models for the quantitative analysis of DML using SG2 preprocessing (left) and wavelet denoising preprocessing (right); Figure S11: NIR calibration models for the quantitative analysis of DML using wavelet compression preprocessingg (left) and MSC combined with wavelet denoising preprocessing (right); Figure S12: NIR calibration models for the quantitative analysis of DML using MSC combined with wavelet compression preprocessingg (left) and SNV combined with wavelet denoising preprocessing (right); Figure S13: NIR calibration

models for the quantitative analysis of DML using SNV combined with wavelet compression preprocessingg (left) and SG1 combined with wavelet compression preprocessing (right); Figure S14: NIR calibration models for the quantitative analysis of DML using SG2 combined with wavelet denoising preprocessingg (left) and SG2 combined with wavelet compression preprocessing (right); Figure S15: PXRD patterns of DML, DMM, excipients and formulation samples showing characteristic diffraction peaks selected for quantitative analysis; Figure S16: NIR spectra of different DMM solid forms and Corresponding Vibration Assignments; Table S1: DML, DMM and Excipients Mixing Ratios; Table S2: Comparative analytical performance of PXRD and the optimized NIR–PLSR model.

Author Contributions: Conceptualization, R.G. and M.L. (Maolin Li); Methodology, R.G., M.L. (Maolin Li) and L.Z.; Validation, M.L. (Maolin Li) and Q.Y.; Investigation, R.G. and S.W.; Writing—original draft, R.G. and X.L.; Writing—review and editing, X.L. and M.L. (Mingdi Liu); Supervision, M.L. (Mingdi Liu) and S.W.; Project administration, L.Z. and Q.Y.; Funding acquisition, S.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Central Government-Guided Local Science and Technology Development Fund Project (254Z3001G), Special Project for High-Level Talent Team Construction in Hebei Provincial Science and Technology Plan (253A7626D), Tianjin Youth Science and Technology Talent Project (QN20230220) and Tianjin Natural Science Foundation (23JCYBJC01450). And The APC was funded by the Central Government-Guided Local Science and Technology Development Fund Project (254Z3001G).

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: Author Xiaogang Lian was employed by the company Hansoh Pharma (China). The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Bellur Atici, E.; Karliga, B. Quantitative determination of two polymorphic forms of imatinib mesylate in a drug substance and tablet formulation by X-ray powder diffraction, differential scanning calorimetry and attenuated total reflectance Fourier transform infrared spectroscopy. *J. Pharm. Biomed. Anal.* **2015**, *114*, 330–340. [CrossRef]
2. Guo, C.; Luo, X.; Zhou, X.; Shi, B.; Wang, J.; Zhao, J.; Zhang, X. Quantitative analysis of binary polymorphs mixtures of fusidic acid by diffuse reflectance FTIR spectroscopy, diffuse reflectance FT-NIR spectroscopy, Raman spectroscopy and multivariate calibration. *J. Pharm. Biomed. Anal.* **2017**, *140*, 130–136. [CrossRef]
3. Lee, E.H. A practical guide to pharmaceutical polymorph screening & selection. *Asian J. Pharm. Sci.* **2014**, *9*, 163–175. [CrossRef]
4. Li, Y.; Chow, P.S.; Tan, R.B. Quantification of polymorphic impurity in an enantiotropic polymorph system using differential scanning calorimetry, X-ray powder diffraction and Raman spectroscopy. *Int. J. Pharm.* **2011**, *415*, 110–118. [CrossRef]
5. Otsuka, M.; Kato, F.; Matsuda, Y. Comparative evaluation of the degree of indomethacin crystallinity by chemoinfometrical Fourier-transformed [corrected] near-infrared spectroscopy and conventional powder X-ray diffractometry. *AAPS PharmSci* **2000**, *2*, E9. [CrossRef] [PubMed]
6. Pindelska, E.; Sokal, A.; Kolodziejski, W. Pharmaceutical cocrystals, salts and polymorphs: Advanced characterization techniques. *Adv. Drug Deliv. Rev.* **2017**, *117*, 111–146. [CrossRef] [PubMed]
7. Testa, C.G.; Prado, L.D.; Costa, R.N.; Costa, M.L.; Linck, Y.G.; Monti, G.A.; Cuffini, S.L.; Rocha, H.V.A. Challenging identification of polymorphic mixture: Polymorphs I, II and III in olanzapine raw materials. *Int. J. Pharm.* **2019**, *556*, 125–135. [CrossRef]
8. Chernyshev, V.V. Structural Characterization of Pharmaceutical Cocrystals with the Use of Laboratory X-ray Powder Diffraction Patterns. *Crystals* **2023**, *13*, 640. [CrossRef]
9. Qiu, J.B.; Li, G.; Sheng, Y.; Zhu, M.R. Quantification of febuxostat polymorphs using powder X-ray diffraction technique. *J. Pharm. Biomed. Anal.* **2015**, *107*, 298–303. [CrossRef] [PubMed]
10. Zappi, A.; Maini, L.; Galimberti, G.; Caliandro, R.; Melucci, D. Quantifying API polymorphs in formulations using X-ray powder diffraction and multivariate standard addition method combined with net analyte signal analysis. *Eur. J. Pharm. Sci.* **2019**, *130*, 36–43. [CrossRef]

11. Zhao, Y.-M.; Zheng, Z.-B.; Li, S. Quantification of flupirtine maleate polymorphs using X-ray powder diffraction. *Chin. Chem. Lett.* **2016**, *27*, 1666–1672. [CrossRef]
12. Farias, M.; Carneiro, R. Simultaneous quantification of three polymorphic forms of carbamazepine in the presence of excipients using Raman spectroscopy. *Molecules* **2014**, *19*, 14128–14138. [CrossRef] [PubMed]
13. Heinz, A.; Savolainen, M.; Rades, T.; Strachan, C.J. Quantifying ternary mixtures of different solid-state forms of indomethacin by Raman and near-infrared spectroscopy. *Eur. J. Pharm. Sci.* **2007**, *32*, 182–192. [CrossRef]
14. Hennigan, M.C.; Ryder, A.G. Quantitative polymorph contaminant analysis in tablets using Raman and near infra-red spectroscopies. *J. Pharm. Biomed. Anal.* **2013**, *72*, 163–171. [CrossRef]
15. Hu, Y.; Erxleben, A.; Ryder, A.G.; McArdle, P. Quantitative analysis of sulfathiazole polymorphs in ternary mixtures by attenuated total reflectance infrared, near-infrared and Raman spectroscopy. *J. Pharm. Biomed. Anal.* **2010**, *53*, 412–420. [CrossRef]
16. Rath, M.; Bērziņš, K.; Boyd, B.J.; Heinz, A. Spatially-offset Raman spectroscopy for the non-invasive, real-time characterization of in situ skin implant formation and drug release kinetics. *J. Control. Release* **2025**, *387*, 114182. [CrossRef] [PubMed]
17. Rodà, F.; Picciolini, S.; Mangolini, V.; Gualerzi, A.; Seneci, P.; Renda, A.; Sesana, S.; Re, F.; Bedoni, M. Raman Spectroscopy Characterization of Multi-Functionalized Liposomes as Drug-Delivery Systems for Neurological Disorders. *Nanomaterials* **2023**, *13*, 699. [CrossRef]
18. Bruni, G.; Capsoni, D.; Milanese, C.; Cardini, A. Polymorphic quantification of dexketoprofen trometamol by differential scanning calorimetry. *J. Therm. Anal. Calorim.* **2022**, *148*, 1949–1958. [CrossRef]
19. Guo, W.; Li, C.; Du, P.; Wang, Y.; Zhao, S.; Wang, J.; Yang, C. Thermal properties of drug polymorphs: A case study with felodipine form I and form IV. *J. Saudi Chem. Soc.* **2020**, *24*, 474–483. [CrossRef]
20. Ostrowska-Ligęza, E.; Wirkowska-Wojdyła, M.; Brzezińska, R.; Piasecka, I.; Synowiec, A.; Gondek, E.; Górska, A. Application of Differential Scanning Calorimetry and Thermogravimetry for Thermal Analysis of Dark Chocolates. *Appl. Sci.* **2024**, *14*, 9502. [CrossRef]
21. Bobály, B.; Fleury-Souverain, S.; Beck, A.; Veuthey, J.-L.; Guillarme, D.; Fekete, S. Current possibilities of liquid chromatography for the characterization of antibody-drug conjugates. *J. Pharm. Biomed. Anal.* **2018**, *147*, 493–505. [CrossRef]
22. da Silva, V.H.; Goncalves, J.L.; Vasconcelos, F.V.; Pimentel, M.F.; Pereira, C.F. Quantitative analysis of mebendazole polymorphs in pharmaceutical raw materials using near-infrared spectroscopy. *J. Pharm. Biomed. Anal.* **2015**, *115*, 587–593. [CrossRef]
23. Peng, X.; Yu, X.; Lu, L.; Ye, X.; Zhong, L.; Hu, W.; Chen, S.; Song, Q.; Cai, Y.; Yin, J. Application of handheld near infrared spectrometer in quality control of traditional Chinese medicine: Rapid screening and quantitative analysis of Lonicerae Japonicae Flos adulteration. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2025**, *326*, 125215. [CrossRef]
24. Salzer, R. Practical Guide to Interpretive Near-Infrared Spectroscopy. By Jerry Workman, Jr. and Lois Weyer. *Angew. Chem. Int. Ed.* **2008**, *47*, 4628–4629. [CrossRef]
25. Xie, Q.; Wu, R.; Zhong, X.; Dong, Y.; Fan, Q. Real-time simultaneous detection of microbial contamination and determination of an ultra low-content active pharmaceutical ingredient in tazarotene gel by near-infrared spectroscopy. *RSC Adv.* **2018**, *8*, 27037–27044. [CrossRef]
26. Jiao, H.; Wang, Y.; Lian, H.; Bu, R.; Bo, S.; Bai, W. Rapid and nondestructive authentication of Glehniae Radix and detection of adulteration using near-infrared spectroscopy and chemometrics. *Results Chem.* **2026**, *25*, 103275. [CrossRef]
27. Kachrimanis, K.; Rontogianni, M.; Malamataris, S. Simultaneous quantitative analysis of mebendazole polymorphs A-C in powder mixtures by DRIFTS spectroscopy and ANN modeling. *J. Pharm. Biomed. Anal.* **2010**, *51*, 512–520. [CrossRef]
28. Li, J.; Jiang, Y.; Fan, Q.; Chen, Y.; Wu, R. Simultaneous determination of the impurity and radial tensile strength of reduced glutathione tablets by a high selective NIR-PLS method. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2014**, *125*, 278–284. [CrossRef]
29. Sun, J. A correlation principal component regression analysis of NIR data. *J. Chemom.* **2005**, *9*, 21–29. [CrossRef]
30. Liu, M.; Shi, P.; Wang, G.; Wang, G.; Song, P.; Liu, Y.; Wu, S.; Gong, J. Quantitative analysis of binary mixtures of entecavir using solid-state analytical techniques with chemometric methods. *Arab. J. Chem.* **2021**, *14*, 103360. [CrossRef]
31. Næs, T.; Martens, H. Principal component regression in NIR analysis: Viewpoints, background details and selection of components. *J. Chemom.* **2005**, *2*, 155–167. [CrossRef]
32. Yuan, H.; Tang, S.; Luo, Q.; Xiao, T.; Wang, W.; Ma, Q.; Guo, X.; Wu, Y. Micro-FTIR spectroscopy and partial least-squares regression for rapid determination of moisture content of nanogram-scaled heat-treated wood. *J. Wood Sci.* **2020**, *66*, 1. [CrossRef]
33. Dhanoa, M.S.; Lister, S.J.; Sanderson, R.; Barnes, R.J. The Link between Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) Transformations of NIR Spectra. *J. Near Infrared Spectrosc.* **1994**, *2*, 43–47. [CrossRef]
34. Liu, M.; Fu, R.; Xu, G.; Dong, W.; Qi, H.; Dong, P.; Song, P. Simultaneous Quantitative Analysis of Polymorphic Impurities in Canagliflozin Tablets Utilizing Near-Infrared Spectroscopy and Partial Least Squares Regression. *Molecules* **2026**, *31*, 230. [CrossRef]

35. Liu, M.; Liu, J.; Wang, Q.; Song, P.; Li, H.; Sun, Z.; Shi, C.; Dong, W. Quantitative analysis of low-content impurity crystal forms in canagliflozin tablets by NIR solid-state analysis technique. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2024**, *311*, 124000. [CrossRef]
36. Liu, M.; Liu, J.; Wang, Q.; Song, P.; Li, H.; Wu, S.; Gong, J. Quantitative analysis of low content polymorphic impurities in canagliflozin tablets by PXRD, NIR, ATR-FITR and Raman solid-state analysis techniques combined with stoichiometry. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2023**, *293*, 122458. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.