RSC Advances

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *RSC Adv.*, 2026, **16**, 26513

Green simultaneous quantification of levodopa, carbidopa, and benserazide in anti-Parkinson's tablets by ATR-FTIR spectroscopy combined with hurdle modelling and machine learning

Manh Huy Nguyen, [ab] Thanh Dam Nguyen, [ab] Hong Anh Duong [ab] and Hung Viet Pham [*b]

Received 17th March 2026
Accepted 4th May 2026

DOI: 10.1039/d6ra02243j

rsc.li/rsc-advances

In this study, an integrated hurdle modelling workflow combining ATR-FTIR and machine learning was developed for the simultaneous quantification of levodopa (LD), carbidopa (CD), and benserazide (BZ) in anti-Parkinson's medication. A calibration set of 103 synthetic pellets was prepared with wide concentration ranges for the three active pharmaceutical ingredients (APIs), 0–80% w/w for LD, 0–40% w/w for BZ; and 0–8% w/w for CD, and analysed by ATR-FTIR and HPLC-DAD reference method. Nine pellets prepared from three commercial medications were used for external evaluation and calibration transfer. A two-stage hurdle model was employed for each analyte, including a classifier for presence/absence and a regressor fitted only on positive samples to predict concentration. The modelling workflow comprised four modules, in which the Savitzky–Golay first derivative, combined with the standard normal variate (SNV) on mean spectra, was the most suitable preprocessing method. Logistic regression for LD and BZ, and supported vector classification (SVC) for CD were selected as optimal classifiers, while random forest (RF) regression in the half spectral region (675–1800 cm^{-1}) was the best overall regressor for all APIs. When applied to commercial samples, the model classifiers correctly identified the presence/absence of all APIs, and calibration transfer using only six additional commercial pellets significantly reduced bias for all three APIs and achieved a high AGREE green score (0.75) at the same time. These results show that ATR-FTIR, combined with a carefully designed hurdle model workflow, can provide a rapid and green screening tool for multi-API in anti-Parkinson's medications.

1. Introduction

Parkinson's disease is a common chronic neurodegenerative disorder that affects millions of patients worldwide,[1,2] with over 6 million people affected globally and nearly 0.6 million in Vietnam. It is characterised by a progressive loss of dopaminergic neurons in the central nervous system and a resulting depletion of dopamine, leading to motor symptoms such as tremor, rigidity and bradykinesia, as well as several non-motor manifestations.[3,4] Levodopa (LD) remains the mainstay of symptomatic therapy because it can be converted to dopamine in the brain and thus partially compensates for this deficiency.[5] However, a large portion of LD is metabolised into dopamine in peripheral tissues by the DOPA decarboxylase before crossing the blood–brain barrier, thereby reducing the amount of dopamine produced in the central nervous system and contributing to more severe side effects at higher doses.[6–9] For this reason, LD is routinely co-formulated with peripheral decarboxylase inhibitors such as carbidopa (CD) or benserazide (BZ) to increase bioavailability and limit unwanted effects.[10] Fixed-dose combinations of LD/CD and LD/BZ therefore occupy a central position in the oral treatment of Parkinson's disease. Commercial preparations such as Madopar® (LD/BZ) and Syndopa® (LD/CD) are widely used in many countries, including Vietnam.

Because these medicines are administered chronically and often by vulnerable patients, strict quality control is required. This requirement is particularly important in low- and middle-income countries like Vietnam, where substandard and falsified medicines have been reported with increasing frequency and can lead to therapeutic failure or toxicity. A variety of analytical techniques have been proposed for the determination of LD, CD and BZ in pharmaceutical products. High-performance liquid chromatography (HPLC) is recommended in major pharmacopoeias as the reference method owing to its high sensitivity and

[a]Research Centre for Environmental Technology and Sustainable Development (CETASD), Faculty of Chemistry, VNU University of Science, Vietnam National University, Hanoi, 334 Nguyen Trai, Thanh Xuan, Hanoi 100000, Vietnam

[b]Key Laboratory of Analytical Technology for Environmental Quality and Food Safety Control (KLATEFOS), VNU University of Science, Vietnam National University, Hanoi, 334 Nguyen Trai, Thanh Xuan, Hanoi 100000, Vietnam. E-mail: vietph@vnu.edu.vn

selectivity.[11–17] HPLC-UV and LC-MS/MS methods can provide excellent trueness and precision, but these methods usually involve several steps of sample preparation and the use of noticeable volumes of organic solvents. Although efforts have been made to develop greener HPLC approaches using aqueous mobile phases,[18] analysis times remain relatively long, making such methods less suitable for rapid screening of large numbers of samples.

Recently, vibrational spectroscopy combined with chemometrics has attracted considerable attention as a route towards rapid, low-solvent and potentially greener analysis.[19] Among these techniques, attenuated total reflection Fourier transform infrared spectroscopy (ATR-FTIR) is particularly attractive for solid oral dosage forms due to non-destructive, fast measurements and minimal sample preparation.[20–24] As a result, ATR-FTIR is compatible with process analytical technology concepts and can, in principle, be integrated into manufacturing or quality-control workflows.[25] Several studies have shown that FTIR or ATR-FTIR, when coupled with multivariate calibration methods such as principal component analysis (PCA), partial least squares (PLS) or principal component regression (PCR), can be used for quantitative analysis of active ingredients and impurities in tablets, powders and biological samples.[26–31] In many cases, the analytical performance with ATR-FTIR has been sufficient for screening purposes and, in some situations, for partial replacement of conventional wet-chemical methods.

In the field of pharmaceutical quality control, ATR-FTIR and related techniques have already been applied to a variety of tasks. Rapid quantification of paracetamol in tablets and detection of substandard or falsified products have been reported using reflectance IR in combination with multivariate models.[32] Other studies have relied on FTIR/ATR-FTIR, combined with more advanced chemometric tools, to differentiate between expired and compliant tablets or to characterise counterfeit products containing undeclared active substances.[33] These examples confirm that ATR-FTIR, when combined with chemometrics, can provide a fast and non-destructive means of screening pharmaceutical products within a green-analysis framework.

For LD based pharmacotherapy, a range of analytical procedures alternative to HPLC has also been investigated, including UV-vis spectrophotometry,[34,35] fluorescence spectroscopy,[36] capillary electrophoresis,[18,37–39] and electroanalytical methods.[40–42] UV-vis methods combined with PLS or PCR, as well as voltametric procedures assisted by multivariate calibration, have been reported for binary and ternary mixtures containing levodopa and related drugs.[34,35] With respect to ATR-FTIR, simultaneous determination of LD and CD in LD–CD formulations has been described in solution using multivariate models built on selected wavenumber regions.[43] These works demonstrate that ATR-FTIR has potential for analysing LD/CD systems. However, they mainly concerned binary mixtures, relatively simple matrices or solution phase measurements, and did not address the transfer of calibration models from laboratory mixtures to commercial tablets with complex excipient compositions.

In parallel, a second line of development has focused on classification applications rather than quantitative analysis. ATR-FTIR and FTIR have been employed, often in combination with pattern recognition methods and modern classifiers, to distinguish compliant from non-compliant products, detect expired lots, or flag counterfeit medicines. Most of these studies focus on qualitative discrimination or on quantifying a single active ingredient. The simultaneous determination of several active substances in Parkinson's tablets, in the presence of complex excipient backgrounds and with clear differences between calibration and commercial samples, has not been studied in depth.

A further challenge arises when models constructed under controlled laboratory conditions are transferred to routine analysis. Differences in excipient type and ratio, particle size, compaction force, moisture content and contact conditions can all influence ATR-FTIR spectra of tablets. Calibration models built on synthetic mixtures with a single excipient profile may therefore lose predictive ability when they are applied to tablets from different manufacturers. This situation can be described as calibration transfer and domain mismatch. Recent work in pharmaceutical spectroscopy has emphasised the importance of model updating and calibration transfer strategies in order to mitigate these effects while maintaining the validity of predictions.

In view of these considerations, the present study was conducted to develop an integrated ATR-FTIR/machine learning workflow for simultaneous identification and quantification of LD, CD and BZ in Parkinson's tablets. A two-stage hurdle structure was used so that the presence/absence of each analyte and its concentration can be treated separately. The workflow included replicate level quality control, systematic comparison of preprocessing pipelines, nested cross-validation for model selection and assessment of calibration transfer from synthetic to commercial tablets. By addressing both the analytical and modelling aspects, the study aims to provide a framework that can be adapted to other multi-API tablet systems in pharmaceutical quality control.

2. Experimental

2.1. Chemicals

Levodopa (99.5%), carbidopa monohydrate (99.5%), and benserazide hydrochloride (99.5%) were obtained from Tokyo Chemical Industry (TCI, Japan) and used as reference standards for model development and HPLC validation. All standards were stored in amber glass vials at 4 °C in order to minimise degradation.

Common tablet excipients, including microcrystalline cellulose, mannitol, magnesium stearate, povidone, talc and maize starch, were supplied by the Vietnam National Institute of Drug Quality Control (Hanoi, Vietnam). These substances were used to prepare model mixtures that mimic the excipient matrix of commercial products. Hydrochloric acid (37%), sodium dihydrogen phosphate (NaH_2PO_4), phosphoric acid (85%) and ethanol (analytical grade) were purchased from Merck (Germany). Deionised water was produced by a Milli-Q system (Millipore, USA) and was used for the preparation of mobile phases and standard solutions in HPLC analyses.

2.2. Sample sets and reference analysis

2.2.1. Synthetic samples. A calibration set of 103 synthetic pellets was prepared in the laboratory. LD, BZ and CD were blended with a mixture of excipients, in which the content proportions of microcrystalline cellulose, mannitol, magnesium stearate, povidone, talc, and maize starch were fixed, and the mass fractions of the three APIs were varied systematically. For each synthetic tablet, 100 mg of a mixture of APIs and excipient was thoroughly mixed in an agate mortar, then 30 mg of the powder sample was accurately weighed and compressed into a pellet using a manual hydraulic press (PerkinElmer, USA). These pellets were then subjected to ATR-FTIR spectroscopy according to the procedure described in Section 2.2.3.

Four types of synthetic pellets were produced, including excipients-only (only 1 sample), single API formulations (17 samples, including 3 LD-only samples, 7 BZ-only samples, 7 CD-only samples), binary combinations (59 samples, including 25 LD/BZ samples, 25 LD/CD samples, and 9 BZ-CD samples), and ternary mixtures (26 samples). In this way, a wide concentration range was obtained within a controlled excipient profile, *i.e.*, LD: 0–80% w/w; BZ: 0–40% w/w; and CD: 0–8% w/w. These concentration ranges were chosen to ensure coverage of at least 50–150% of the API content in commercial tablet formulations. The nominal and HPLC-DAD measured API contents in these synthetic pellets are presented in Table S1 in the SI. For the synthetic pellets, a zero concentration of a given API corresponds to a formulation in which that API was not added to the blend; in these samples, the corresponding HPLC-DAD peak was below the lower limit of the calibration range and was therefore treated as true absence.

2.2.2. Commercial samples. Three different commercial fixed dose combinations were analysed, including Madopar® (Roche, Italy; 200 mg LD and 50 mg BZ per tablet), Syndopa® (Sun Pharma, India; 250 mg LD and 25 mg CD per tablet) and Evertogen® (Evertogen Life Sciences, India; 100 mg LD and 10 mg CD per tablet). All commercial products were obtained from common pharmaceutical distributors in Hanoi, Vietnam, in December 2025. In total, 45 tablets (five tablets per batch × three batches per brand × three brands) were included, including 30 tablets of LD/CD and 15 tablets of LD/BZ combinations. Five tablets from each batch were crushed, mixed thoroughly, and 30 mg of the powdered drug was weighed, pressed into a pellet, yielding a total of 9 commercial pellets. These pellets were then analysed by ATR-FTIR spectroscopy, as with the synthetic samples. The first three of these pellets (one per brand, S1–S3) were used to provide a spectra dataset as a fixed external evaluation set, while the remaining six (two per brand) were used for balanced calibration transfer experiments. For the commercial pellets, a zero concentration for a given API indicates that the corresponding tablet product does not contain this API according to the label claim and that no quantifiable peak was observed by HPLC-DAD at the respective detection wavelength.

2.2.3. ATR-FTIR measurements. ATR-FTIR measurements were carried out on a Nicolet iN10 MX FT IR microscope (Thermo Fisher Scientific, USA) equipped with a liquid nitrogen cooled mercury cadmium telluride (MCT) detector and a germanium (Ge) ATR crystal. Spectra were recorded in ATR mode over the range 675–4000 cm^{-1}. Before measurements, the MCT detector was cooled with liquid nitrogen for at least 30 min to stabilise the signal, and between measurements, the ATR crystal was cleaned with ethanol. For each spectrum, 64 scans were co-added at a resolution of 2 cm^{-1}. The acquisition time per spectrum was approximately 5 s, and the nominal contact pressure on the ATR crystal was 15 mbar. All measurements were performed at room temperature (25 °C) and at <40% relative humidity, controlled by an air conditioner and a humidifier, respectively.

Each pellet was analysed at four different positions on the surface, and two replicate spectra were recorded at each position. The same protocol was applied to synthetic and commercial pellets to ensure comparability between domains. In total, 824 spectra of synthetic pellets and 72 spectra of commercial pellets were acquired.

2.2.4. HPLC-DAD reference method. Reference values for LD, BZ and CD in both synthetic and commercial tablets were obtained by HPLC-DAD. The method was adapted from a previously described procedure[18] and was validated for the concentration range of interest. Analyses were performed on an HPLC 20A system (Shimadzu, Japan) consisting of an LC 20A pump, a SIL 20A autosampler, an SPD 20A photodiode array detector and a Shim pack VP ODS C18 column (150 × 4.6 mm i.d., 5 μm). The mobile phase was composed of 54.2 mM KH_2PO_4 buffer, adjusted to pH 3.8 with phosphoric acid and containing 121 μM sodium octane sulfonate as an ion pair reagent. The flow rate was 1.06 mL min^{-1}, the injection volume was 10 μL, and the column temperature was maintained at 35 ° C. Detection was carried out at 280 nm for LD and CD and at 210 nm for BZ.

The method showed linearity in the range 10–500 mg L^{-1}. Intra-day repeatability and inter-day precision, expressed as relative standard deviations of peak areas, were below 1% and 5%, respectively. Accuracy was evaluated by standard addition experiments in excipient matrices, with recoveries ranging from 92% to 105%.

For pellet analysis, an accurately weighed portion of each compressed sample was transferred into a volumetric flask and dissolved in 5 mL of 0.1 M H_3PO_4. The resulting solution was diluted to bring the concentrations within the calibration range and filtered through a 0.22 μm PTFE membrane before injection into the HPLC system. The concentrations measured by HPLC-DAD were used as reference values for all modelling steps.

2.3. Study design and modelling strategy

This work was designed to establish a unified chemometric and machine learning procedure for simultaneous identification and quantification of LD, BZ and CD in Parkinson's tablets based on ATR-FTIR spectra. A particular feature of the dataset was the presence of several formulations in which one or two of the actives were absent. At the level of each analyte, a substantial fraction of samples therefore had zero concentration. These structural zeros reflect the multi-product nature of the dataset and complicate the use of a single global regression model. In all cases, these zeros correspond to true absences rather than trace-level impurities, as explained in Sections 2.2.1 and 2.2.2.

To reflect this structure, a two-stage hurdle model was adopted. For each analyte, a classifier was first applied to distinguish presence (concentration $y > 0$) from absence ($y = 0$). A regression model was then fitted using only the samples with

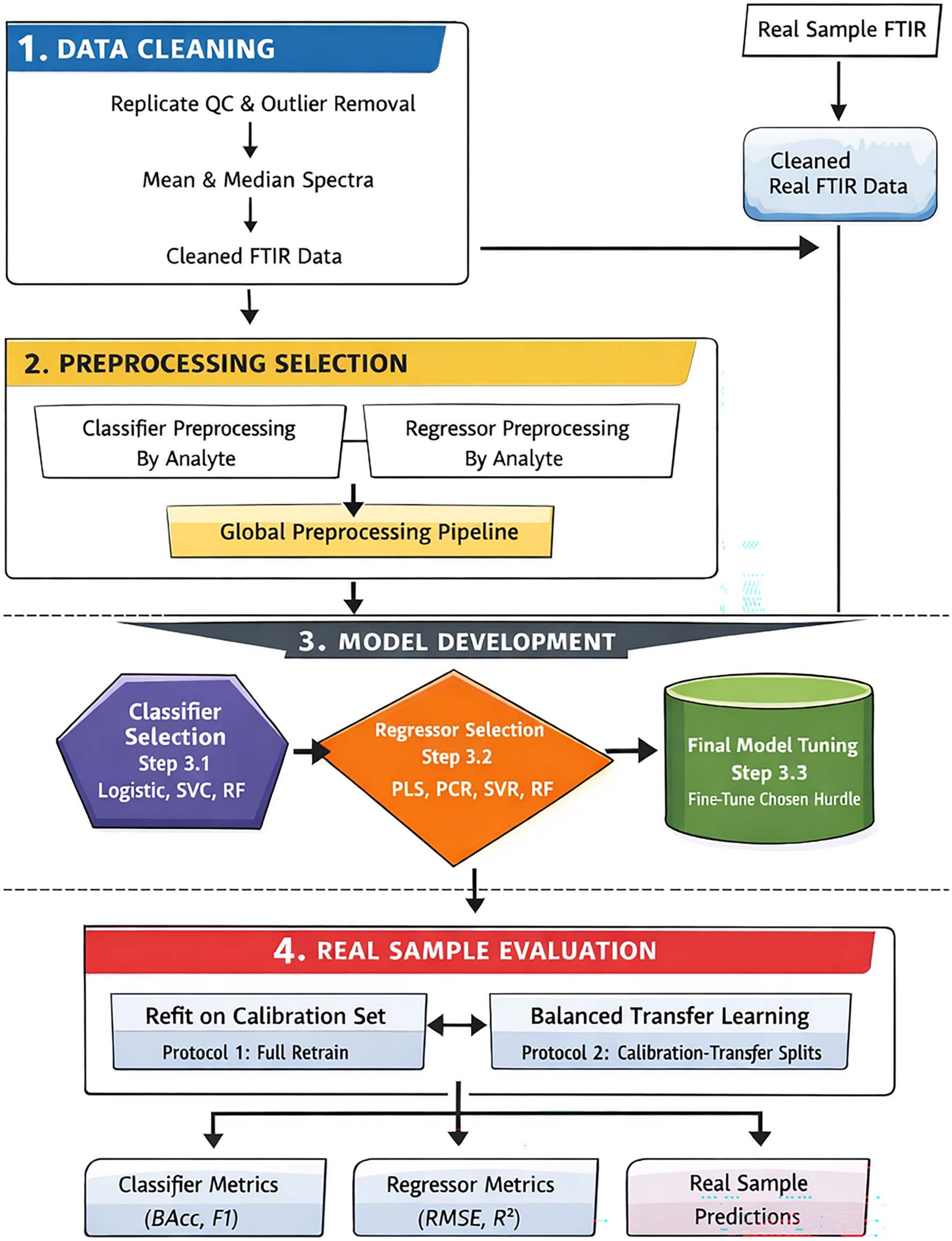

Fig. 1 Flowchart of the integrated ATR-FTIR and machine learning workflow for simultaneous quantification of LD, CD, and BZ in Parkinson's medication.

$y > 0$. During prediction, a zero value was assigned when the classifier indicated absence; otherwise, the output of the regressor was reported as the predicted concentration. In this way, classification and regression tasks were separated, and the effect of structural zeros was handled explicitly.

The overall modelling workflow was organised into four modules: (i) quality control of spectral replicates and aggregation of representative spectra for each sample; (ii) systematic screening of preprocessing pipelines; (iii) selection and optimisation of classifiers and regressors within the hurdle framework; and (iv) application of the final model to commercial tablets with and without calibration transfer (Fig. 1). Repeated nested cross-validation with grouping by sample ID (3 repeats × 5 outer folds) was employed in modules 2 and 3 to separate model selection from performance estimation and to prevent information leakage between replicates within the same tablet.

2.3.1. Module 1: replicate quality control and representative spectra. In module 1, individual spectral replicates were examined, and outliers were removed before representative spectra were computed. Raw spectra were first processed using a Savitzky–Golay (SG) filter (smooth, window = 11, poly = 2) combined with a standard normal variate (SNV) transformation. PCA was then performed on the preprocessed replicates. The number of principal components was selected so that at least 95% of the total variance was explained. Two statistics were employed for outlier detection: Hotelling's T^2, which reflects the position of each spectrum in the score space, and the Q residual, which measures the part of the variance not captured by the PCA model. Critical limits for T^2 (99% confidence) and Q (95% confidence) were calculated, and a replicate was rejected when either statistic exceeded its limit.

After outlier rejection, mean and median spectra were calculated for each pellet ID. These representative spectra, together with the corresponding reference concentrations, formed two cleaned datasets (mean based and median based) that were used in the subsequent evaluation of preprocessing pipelines. The mean-based and median-based dataset is provided in the SI data file.

2.3.2. Module 2: screening of preprocessing pipelines. In module 2, the influence of spectral preprocessing on hurdle model performance was systematically examined. Different combinations of SG filtering and scatter correction techniques, including SNV and multiplicative scatter correction (MSC), were compared. The following 8 pipelines were considered: raw + SNV, raw + MSC, SG smoothing (0th derivative) + SNV, SG smoothing + MSC, SG first derivative + SNV, SG first derivative + MSC, SG second derivative + SNV, and SG second derivative + MSC. SG smoothing and derivatives are widely used in FTIR analysis of solid pharmaceutical samples to correct baseline variations and reduce partially overlapping bands, while limiting distortion of band shapes.[44–46] SNV and MSC are popular single-spectrum scatter-correction techniques that compensate for variations in effective particle size and contact pressure in ATR measurements[47–49] and were therefore included systematically in the set of candidate pre-processing pipelines. Both mean based and median based representative spectra from module 1 were evaluated. For each pipeline and each analyte, a simplified hurdle structure was adopted. Logistic regression was used as the classifier for API presence/absence, and PLS regression was used to predict concentrations in positive samples.

Performance was assessed by repeated nested cross-validation with grouping by pellet ID (3 repeats × 5 outer folds). In the inner loop, the regularisation parameter C for logistic regression and the number of latent variables for PLS were tuned (Table S2). In the outer loop, classification and regression metrics were calculated. For the classifier, balanced accuracy (BAcc), $F1$ score and area under the ROC curve (AUC) were recorded. For the regression, three error measures were used, including $RMSE_0$ for truly zero samples, $RMSE_{pos}$ for all truly positive samples, and $RMSE_{pos,TP}$ for positive samples correctly classified as positive.

To compare pipelines, a rank was assigned to each configuration within each outer fold based on a composite criterion that combined the three classification metrics and the three regression errors, as described in SI Text S1. Mean ranks and their standard deviations were then calculated across the 15 outer tests for each pipeline. Pipelines with low mean rank and low variability were regarded as more favourable. The pipeline that offered the best compromise between classification and regression across the three analytes was selected for use in module 3.

2.3.3. Module 3: selection and optimisation of classifiers and regressors

2.3.3.1. Step 3.1 – selection of classifiers. In step 3.1, three spectral regions were studied: the fingerprint region (675–1200 cm^{-1}), a "half" region (675–1800 cm^{-1}) and the full region of recorded spectra (675–4000 cm^{-1}). For each spectral region and analyte, three classifiers were evaluated: logistic regression, support vector classification (SVC) with a radial basis function kernel, and random forest (RF). Hyperparameters for each classifier were tuned within the inner loop of the nested cross-validation. For logistic regression, the regularisation parameter C was optimised. For SVC, C and the kernel parameter γ were selected. For RF, the number of trees, maximum depth, minimum number of samples per leaf and the number of variables considered at each split were adjusted. All hyperparameter tuning was conducted using a random grid within predefined ranges (Table S3).

Classifier performance was evaluated in the outer folds using BAcc, $F1$ score and AUC. For each analyte region combination, a combined rank was calculated as a weighted sum of the ranks for BAcc and $F1$ score, with a higher weight assigned to balanced accuracy (Text S2). This calculation reflected the role of the classifier as the entry gate to the hurdle model, where balanced treatment of present and absent classes is important. The classifiers with the lowest overall rank for a given analyte and spectral region were retained for further consideration.

2.3.3.2. Step 3.2 – selection of regressors. In step 3.2, four regression approaches were examined: PLS, PCR, support vector regression (SVR) and RF regression. Both single output (one regressor per analyte) and multi output (one regressor for the three analytes simultaneously) modes were considered. As with classifier selection, the number of latent variables was optimised in the inner loop for PLS and PCR. For SVR, C, γ and the width of the ε insensitive zone were tuned. For RF regression, the number of trees, maximum depth, minimum number of samples per leaf and the number of variables considered at each split were adapted (Table S4).

The regressors were evaluated in the outer folds using $RMSE_0$, $RMSE_{pos}$ and $RMSE_{pos,TP}$. For each configuration, composite ranks were calculated from these three error measures, and the averages and standard deviations of the ranks were determined across outer folds (Text S3). Models were compared separately for each analyte and also at a global level. Special attention was paid to the difference between single output and multi output modes and to the stability of the rankings.

2.3.3.3. Step 3.3 – fine tuning of the final hurdle model. Once suitable classifiers and regressors had been identified, a final hurdle structure was defined. The SG first derivative + SNV pipeline on mean spectra and the half spectral region were fixed. For LD and BZ, logistic regression was retained as a classifier, whereas RF was selected for CD. RF regression in single output mode was used for all three analytes.

In this step, the combined model was optimised end to end. During the inner loop of the nested cross-validation, hyperparameters of both the classifier and regressor components, as well as the classification thresholds, were adjusted simultaneously according to a multi-criteria objective (Table S4). The latter included constraints on $RMSE_{pos}$ in addition to the classification metrics. For each parameter, a consensus value was obtained by taking the median (for continuous parameters) or the most frequent value (for categorical parameters) across the outer folds. These consensus settings were then used to refit the hurdle model on the full synthetic dataset before application to commercial samples.

2.3.4. Module 4: application to commercial samples and calibration transfer. In module 4, the optimised hurdle model was applied to the synthetic samples for recovery tests and the commercial samples. Recovery experiments were carried out in order to further assess the ability of the ATR-FTIR/machine-learning models to quantify the APIs in the presence of common tablet excipients. Synthetic excipient pellets without intentionally added APIs were used as matrices, in which pellets containing only the fixed excipient mixture (microcrystalline cellulose, mannitol, magnesium stearate, povidone, talc and maize starch) were prepared as described in Section 2.2.1. Known amounts of LD, BZ and CD were then added to the excipient powder to obtain three spiking levels for each API (low, medium and high) covering approximately 80–120% of the typical label content range in commercial tablets. For each level, three independent spiked pellets were prepared, analysed by ATR-FTIR following the protocol in Section 2.2.3, and quantified using the final hurdle model. In parallel, HPLC-DAD reference values were obtained for the same pellets. Relative recoveries with respect to HPLC-DAD reference values were calculated as follows:

$$\text{Recovery } (\%) = 100 \times \frac{C_{\text{ATR-FTIR}}}{C_{\text{HPLC-DAD}}}$$

where $C_{\text{ATR-FTIR}}$ is the concentration predicted by the ATR-FTIR/hurdle-model workflow, $C_{\text{HPLC-DAD}}$ is the concentration determined by HPLC-DAD. Mean recovery and relative standard deviation (RSD) were calculated from triplicate pellets for each condition.

For commercial samples, two scenarios were examined. In the first scenario, the model was refitted on the full synthetic dataset and applied directly to the ATR-FTIR spectra of three selected commercial pellets (S1–S3) from three brands without any further adjustment. Before refitting, the classification thresholds were updated using out of fold predictions on the synthetic domain and by maximising Youden's J statistics.

In the second scenario, a calibration transfer strategy based on repeated product balanced splits was used. The ATR-FTIR spectra from six remaining commercial pellets (two additional batches per product brand) served as a small transfer set. 100 random splits were generated. In each split, one spectrum from each batch per product brand was assigned to the calibration transfer training subset (three spectra in total), and the remaining three batches were used as the calibration transfer test subset. The synthetic dataset (103 samples) and the three additional training samples were combined and used to update the hurdle model. The updated model was then applied to the three additional test samples and to the external evaluation samples.

For each analyte and each commercial batch, prediction errors with respect to HPLC UV reference values were calculated. For the direct application scenario, single point estimates were obtained for S1–S3. For the calibration transfer scenario, the distributions of predicted values across the 100 splits were summarised by the median and the interquartile range. Recall, precision and the confusion matrix of the classifier component were also examined to verify that errors in presence/absence assignment did not dominate overall prediction performance.

2.4. Software

All data preprocessing, statistical analysis, and machine learning workflows were implemented in Python 3.14.2. Code development and execution were performed in Visual Studio Code (Microsoft, USA). Numerical array operations were handled using NumPy 2.3.4, tabular data processing was conducted with pandas 2.3.3, and additional scientific computing functions were provided by SciPy 1.16.3. Model training, hyperparameter tuning, and validation were carried out using scikit-learn 1.7.2. Greenness of the analytical methods were assessed through the web version of AGREE calculator tool (https://agree-index.anvil.app/).

3. Results and discussion

3.1. ATR-FTIR spectral characteristics of active ingredients and excipients

The chemical structures of LD, CD, and BZ are showed in Fig. S1 and the ATR-FTIR spectra of these compounds and the excipient mixture are presented in Fig. 2. Measurements were recorded over 675–4000 cm^{-1}, although the most informative

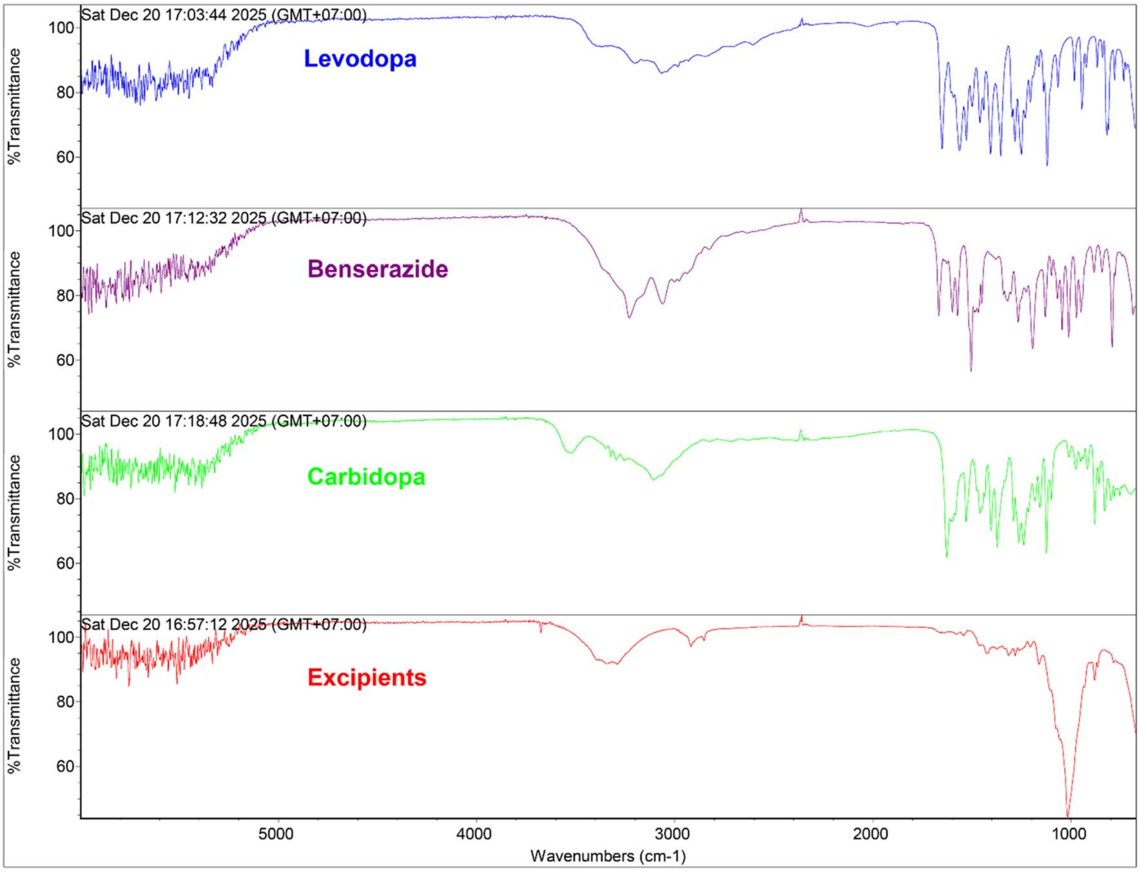

Fig. 2 ATR-FTIR spectra of LD, BZ, CD and excipient mixture.

region was found to be 900–1800 cm^{-1}, where characteristic vibrations of the main functional groups occurred. In the LD spectrum, strong absorption bands were observed around 1600–1650 cm^{-1} (C═O stretching), 1515 cm^{-1} (N–H bending), 1120–1260 cm^{-1} and 720–870 cm^{-1} (C–O stretching and out of plane C–H bending). A broad band near 3300 cm^{-1} was attributed to O–H stretching of the phenolic group. On the other hand, the CD spectrum showed a similar overall shape, but with shifts to slightly lower wavenumbers. Prominent peaks were seen at approximately 1705 cm^{-1} (C═O), 1620 cm^{-1} (N–H) and 1220 cm^{-1} (C–O). Moreover, several intense bands were present in the BZ spectrum, including 1660 cm^{-1} (amide C═O stretching), 1600 cm^{-1} (aromatic C═C stretching) and 1450 cm^{-1} (C–N–H bending). A broad absorption between 3200 and 3400 cm^{-1} arose from overlapping O–H and N–H stretching vibrations, which indicated strong hydrogen bonding. In the excipient mixture, characteristic bands were located at ~3400 cm^{-1} (O–H stretching of cellulose and starch), 2850–2920 cm^{-1} (C–H stretching) and strong peaks near 1450 cm^{-1} and 1050 cm^{-1} (C–O–C and C–O–H vibrations). These spectral features confirmed that the fingerprint and half regions contain the main signals that distinguish the three APIs from the excipient background.

Overall, there was substantial spectral overlap, particularly between LD and CD in the 1100–1600 cm^{-1} region, arising from their similar hydroxyl, amine, and carboxyl groups. Moreover, excipients, mainly cellulose-based materials, exhibit strong absorptions within 900–1200 cm^{-1}, overlapping with key vibrational regions of the APIs and introducing additional background interference. Such complexity limits the applicability of conventional univariate approaches based on single-peak intensity measurements. Therefore, advanced chemometric and machine learning strategies are required to resolve overlapping information and capture subtle multivariate relationships embedded in the full spectral dataset.

3.2. Replicate quality control and outlier rejection

Spectral replicates were subjected to quality control in module 1 before representative spectra were created. After preprocessing with Savitzky–Golay smoothing and SNV, PCA was applied to all 824 spectral replicates, and then Hotelling's T^2 and Q residuals were calculated and used together to identify outliers.

As a result, 54 replicates (6.55%) were rejected, corresponding to 36 of 103 tablet IDs from which at least one replicate was removed, whereas 67 IDs retained all replicates. Most affected IDs lost only one or two replicates, and among them, 22 (40.74%) exceeded the T^2 limit, 20 (37.04%) exceeded the Q limit, and 12 (22.22%) exceeded both limits. This pattern indicates that the two outlier criteria capture complementary types of spectral abnormalities. After outlier removal, mean and median spectra were computed for each pellet ID. These representative spectra were used as input for module 2.

3.3. Screening of preprocessing pipelines

Eight preprocessing pipelines were compared systematically in module 2. The mean and median spectra from module 1 were evaluated separately for each analyte using a simplified hurdle model (a logistic classifier followed by a PLS regressor). Performance was assessed by repeated nested cross-validation (3 repeats × 5 outer folds), and global ranks and their standard deviations were calculated from classification metrics (BAcc, $F1$-score, AUC) and regression errors ($RMSE_0$, $RMSE_{pos}$, $RMSE_{pos,TP}$).

On average across pipelines, mean spectra outperformed median spectra with mean global rank was 6.71 and 10.29, respectively (Table S5). Both SNV and MSC were evaluated in combination with the different SG filter options, and the final choice of SG first derivative followed by SNV on mean spectra reflects a compromise between baseline correction, band-shape preservation and robustness of quantitative predictions under the present ATR-FTIR conditions. This observation suggested that mean spectra retain more chemical information after effective outlier removal. Besides, pipelines based on SG second derivatives showed superior classification metrics in several cases but resulted in substantially higher regression errors, particularly for LD. This trade-off was consistent with the known tendency of higher-order derivatives to emphasise shape differences at the expense of noise amplification.

At the analyte level, SG first-derivative + SNV was optimal for LD and BZ, yielding the lowest $RMSE_{pos,TP}$ (9.377 and 3.384, respectively) and good classification metrics (BAcc > 0.92, $F1$ > 0.95, and AUC > 0.98 for both analytes). For CD, raw + MSC performed marginally better than SG first-derivative + SNV ($RMSE_{pos,TP} = 1.269$ *vs.* 1.337, respectively), but the difference was small compared to the advantages of the latter pipeline for LD and BZ. Finally, the pipeline SG first derivative + SNV applied to mean spectra yielded the lowest global rank (3.80) and acceptable stability (SD of 1.253); therefore, it was adopted as the standard preprocessing for module 3.

3.4. Classifier selection

Three spectral regions (fingerprint, half, and full) and three classifiers (logistic, SVC, and RF) were compared in step 3.1 of module 3. Ranks were derived from outer-fold metrics using a weighted combination that favoured BAcc. Sensitivity analysis with different weightings ($rank_{BAcc}/rank_{F1} = 0.5/0.5$, 0.6/0.4, 0.7/0.3, and 0.8/0.2) showed that the optimal classifier remained unchanged in 8 of 9 analyte–region combinations, with only LD in the half region showing equivalent rankings between logistic regression and RF when a weight ratio of 0.5/0.5 was used. This analysis confirmed that classifier selection was robust with respect to the ranking rule. In this context, BAcc is equal to the average of sensitivity and specificity. For completeness, the corresponding sensitivity and specificity values for all analyte–region combinations can be found in Table 1.

The obtained results, as shown in Table 1, indicated that no single classifier was optimal across all analytes and regions. For LD, RF was the most reasonable classification model across the fingerprint and full regions, with BAcc ≈ 0.94–0.95 and AUC close to 1.0, whereas in the half region, logistic regression achieved the best rank, with slightly higher BAcc (0.955) despite RF having a little better $F1$ and AUC. This outcome suggested that, under the chosen preprocessing, the decision boundary for LD present/absent in the half region was approximately linear, while the two other regions appeared to benefit more from the ability to capture additional nonlinear interactions by RF. For BZ, logistic regression was clearly the foremost classifier, achieving perfect performance (AUC = $F1$ = BAcc = 1.000) in the half region and also the best in the full region, although RF performed only marginally better in the full region (BAcc = 0.9875 *vs.* 0.9808). This result suggested that BZ was the easiest analyte to classify, with the presence/absence classes well separated after preprocessing, allowing a simple linear classifier to suffice.

In contrast, CD was the most challenging analyte for classification. In the fingerprint region, SVC was the optimal classifier (BAcc = 0.9084, $F1$ = 0.9401, AUC = 0.9611), while RF was the best-performing model in the half- and full-region with BAcc of 0.905 and 0.898, respectively. This finding demonstrated that SVC could explain a locally non-linear decision structure present in the CD fingerprint spectral region, but when a larger number of variables was included, interactions across multiple spectral bands were better captured by RF. The lower AUC and BAcc values observed for CD were attributed to its lowest concentration and to greater overlap of its spectral bands with those of LD, the largest content. All nine best region-classifier combinations per API were used to select the regressors.

3.5. Regressor selection and final model

For each APIs, four regression methods were evaluated in step 3.2 of module 3 under single output and multi output

Table 1 Best classifier by active ingredient and spectral region

Analyte	Spectral region	Best classifier	Sensitivity	Specificity	BAcc	$F1$	AUC
LD	Fingerprint	RF	0.963	0.920	0.942	0.968	0.993
LD	Half	Logistic	0.951	0.960	0.955	0.967	0.993
LD	Full	RF	0.987	0.920	0.953	0.981	1.000
BZ	Fingerprint	Logistic	0.986	0.971	0.979	0.986	0.998
BZ	Half	Logistic	1.000	1.000	1.000	1.000	1.000
BZ	Full	RF	1.000	0.975	0.988	0.992	1.000
CD	Fingerprint	SVC	0.972	0.844	0.908	0.940	0.961
CD	Half	RF	0.933	0.876	0.905	0.942	0.958
CD	Full	RF	0.919	0.876	0.898	0.934	0.956

conditions across three spectral regions, using the best optimal classifier for each region determined in step 3.1. The aggregated results (Table S6) indicated that across all 72 evaluated configurations (3 APIs × 4 regressors × 2 output mode × 3 classifier-region), the mean global rank was substantially lower for single-output models than for multi-output models (10.93 *vs.* 14.07), indicating that analyte-specific information was better maintained when separate regressors were built. The RF single-output model in the half region was the best overall configuration, as reflected by the lowest composite rank (6.40) and good stability (1.25). This result suggested that, unlike the classification, the spectral–concentration relationship in quantification was not fully linear, particularly for LD and especially BZ, such that the ability of RF was advantageous. PCR also emerged as a relatively robust alternative in the half- and full-region, showing competitive global ranks (8.45 and 9.72, respectively). By contrast, SVR showed weaker performance, with both single-output and multi-output SVR configurations appearing in the lower half of the ranking table, suggesting that its sensitivity to hyperparameter tuning outweighed any benefit of non-linear flexibility under the limited effective sample size of positive cases.

At the analyte level, however, different modelling behaviours were observed. For LD, the best performance was obtained with RF single-output in the half region ($RMSE_{pos,TP}$ = 8.635), although PLS single-output in the fingerprint region produced a comparable result ($RMSE_{pos,TP}$ = 8.655), suggesting that both linear and non-linear components contributed to the spectral–concentration relationship. For BZ, RF was clearly dominant, with the three best configurations all corresponding to RF single-output models in the half, full, and fingerprint regions ($RMSE_{pos,TP}$ = 3.013, 3.103, and 3.144, respectively), thereby confirming the strong advantage of non-linear modelling for this API. In contrast, CD was best modelled by latent-variable linear methods, with the lowest $RMSE_{pos,TP}$ obtained for PLS single-output in the full region (1.353), whereas the globally best RF configuration gave a slightly higher value (1.479). Taken together, the results of step 3.2 identified RF single-output in the half region as the most reasonable choice, while also demonstrating that the full hurdle model should be fine-tuned in step 3.3.

The fine-tuning results in step 3.3 showed that the outer-fold performance of the final model changed only marginally relative to that of the best configuration identified in step 3.2. Specifically, the $RMSE_{pos,TP}$ for LD remained essentially unchanged (8.624 *vs.* 8.635), while only minor variations were observed for BZ and CD. This finding indicates that the configuration selected in step 3.2 was already close to optimal, and that step 3.3 primarily served as a final stage of parameter stabilisation and threshold normalisation rather than as a source of substantial improvement in predictive accuracy. A further notable observation was the analyte-dependent difference in threshold and hyperparameter stability. The greatest threshold fluctuation across folds was observed for LD, whereas CD showed markedly greater stability. At the same time, the optimal RF regression hyperparameters differed among analytes. For example, larger values of min_samples_leaf were generally selected for LD, indicating a greater need for regularisation, whereas for CD, larger values of max_features were favoured, suggesting that its quantitative information was distributed across a broader range of spectral variables (Table 2). These findings further supported the use of a single-output regression structure, as they indicated that, despite being processed within a common analytical framework, the three analytes retained different predictive characteristics.

3.6. Performance on commercial samples

3.6.1. Recovery in excipient matrices. Recovery experiments were conducted to further assess the selectivity of the ATR-FTIR/hurdle-model workflow for the target APIs in the presence of common tablet excipients. The results obtained for the synthetic excipient pellets are summarised in Table S7. In the synthetic excipient matrix, the mean recoveries of LD, BZ, and CD at the three spiking levels were found to range from 94% to 106%, with the exception of samples containing low CD levels. However, relatively large variations in RSD were observed for all APIs, ranging from 5.9% to 12.7% (Table S7). These fluctuations likely reflected heterogeneity among the pellets, which may be attributed to insufficiently uniform mixing of the API and excipient components. This observation suggests that multiple pellets should be measured for each drug tablet to obtain more reliable results. Such heterogeneity appeared to be

Table 2 The consensus settings for two-stage hurdle model for simultaneous quantification of LD, CD, and BZ in Parkinson's medication

				Classifier		Regressor	
Analyte	Preprocessing pipeline	Spectral type	Spectral region	Model	Hyperparameter	Model	Hyperparameter
BZ	SG first derivative + SNV	Mean	Half	Logistic	$C = 6.639$	RF	max_depth = 21.5 max_features = 0.5 min_samples_leaf = 2.0 n_estimators = 426.0
CD	SG first derivative + SNV	Mean	Half	RF	max_depth = 21.5 max_features = 0.2 min_samples_leaf = 4.0 n_estimators = 1135.0	RF	max_depth = 35.0 max_features = 0.8 min_samples_leaf = 2.0 n_estimators = 531
LD	SG first derivative + SNV	Mean	Half	Logistic	$C = 13.163$	RF	max_depth = 25.0 max_features = 0.3 min_samples_leaf = 6.0 n_estimators = 406

less pronounced in the commercial tablet samples. In addition, paired *t*-test analysis indicated that no statistically significant difference was present between the contents determined by ATR-FTIR and those obtained by HPLC-DAD for the samples used in the recovery evaluation ($p > 0.05$ for all three APIs). Collectively, these findings confirm that, when the excipient profile is held constant, the ATR-FTIR/machine-learning workflow is capable of capturing API-related spectral variation rather than relying solely on differences in the sample matrix.

3.6.2. Direct application. In the direct application scenario, the optimal hurdle model selected in step 3.3 was refitted using the full synthetic-domain dataset (103 samples), and the classifier threshold was re-optimised from out-of-fold predictions using the Youden *J* statistic before direct application to the three commercial tablet samples. Under these conditions, the classification component showed excellent performance, with recall and precision equal to 1.0 for positive cases, indicating that no incorrect presence/absence assignments were made for the commercial products (Fig. 3).

For quantification, the predicted LD content ranged from 37.37 to 58.44% w/w, corresponding to errors of −1.61% to 14.74% relative to HPLC-DAD results (Table 3). For BZ, two products were correctly identified as BZ-free, whereas the remaining sample was predicted to be 10.56% w/w, with an error of 10.58%. In the two LD/CD products, the predicted CD contents were 4.14 and 4.16% w/w, corresponding to errors of −17.69% and −17.79%, respectively. Overall, acceptable agreement was obtained for LD/BZ formulations, whereas larger deviations were observed for LD/CD formulations, with LD tending to be overestimated and CD underestimated. This pattern was likely related to closer matrix similarity between the synthetic calibration set and the LD/BZ commercial products than between the synthetic samples and the LD/CD products. Accordingly, although the model showed promising performance for screening, the observed errors remained large for direct replacement of HPLC-DAD, particularly in formulations containing LD and CD.

Table 3 Concentrations of APIs in commercial medications predicted directly by the hurdle model by the HPLC-DAD method

Analyte	Sample	S1	S2	S3
LD	Labelled content[a] (% w/w)	50.89	36.38	48.48
	HPLC-DAD (% w/w)	51.97	37.98	48.71
	Direct model (% w/w)	58.44	37.37	55.89
	Bias[b] (%)	12.45	−1.61	14.74
CD	Labelled content[a] (% w/w)	4.85		5.09
	HPLC-DAD (% w/w)	5.03	0	5.06
	Direct model (% w/w)	4.14	0	4.16
	Bias[b] (%)	−17.69	—	−17.79
BZ	Labelled content[a] (% w/w)		9.10	
	HPLC-DAD (% w/w)	0	9.55	0
	Direct model (% w/w)	0	10.56	0
	Bias[b] (%)	—	10.58	—

[a] Labelled content = mass of API claimed on label/mass of a tablet × 100 (%). [b] Bias = (concentration predicted by model − concentration measured by HPLC-DAD)/concentration measured by HPLC-DAD × 100 (%).

3.6.3. Calibration transfer. When the repeated product-balanced 3/3 calibration transfer strategy was applied, quantitative performance was improved for all three APIs. After transfer, the median predicted LD contents across 100 repeated splits decreased in all three products and ranged from 34.41 to 55.40% w/w, corresponding to errors of −9.39% to 13.74% relative to HPLC-DAD results (Table 4). The predicted BZ content was 9.97% w/w, differing by only 4.41% from the reference value, whereas the greatest improvement was observed for CD, for which the errors in the two LD/CD products were reduced to −10.88% and −10.56%, corresponding to predicted concentrations of 4.48 and 4.53% w/w, respectively. The agreement between ATR-FTIR/hurdle-model predictions and HPLC-DAD reference values for the three commercial products (S1–S3), before and after calibration transfer, is illustrated in the parity plots provided in Fig. 4. These findings suggest that including a small number of real samples in the calibration set, while maintaining balance across product lines,

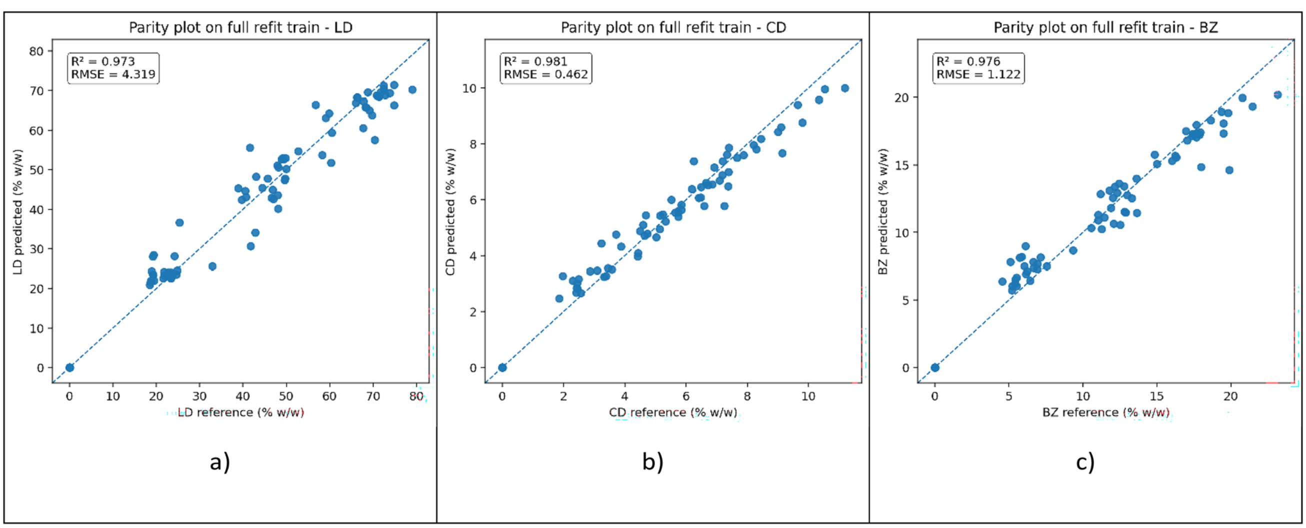

Fig. 3 Parity plots of (a) LD, (b) CD, and (c) BZ on full calibration set (103 samples) at the optimal conditions of the two-stage hurdle model.

Table 4 Concentration and interquartile range (IQR) of APIs in commercial medications predicted by calibration transfer approaches

Analyte	Sample	S1	S2	S3
LD	Calibration transfer (% w/w)	54.10	34.41	55.40
	IQR (% w/w)	53.21–54.33	33.2–34.47	51.93–57.15
	Bias (%)	4.09	−9.39	13.74
CD	Calibration transfer (% w/w)	4.48	0	4.53
	IQR (% w/w)	4.38–4.52	0–0	4.34–4.79
	Bias (%)	−10.88	0	−10.56
BZ	Calibration transfer (% w/w)	0	9.97	0
	IQR (% w/w)	0–0	9.97–10.59	0–0
	Bias (%)	0	4.41	0

produced a sufficiently strong domain-anchoring effect to reduce mismatch between the synthetic and commercial domains, particularly for LD and CD.

Although most errors were substantially reduced, some extreme deviations remained, especially for LD in sample S3, indicating that formulation-specific domain mismatch was not fully captured with only three transfer samples per split. However, the considerable reduction in CD bias after transfer indicates that this analyte benefited particularly from recalibration with real samples. Taken together, these results show that calibration transfer has improved cross-domain robustness and brought the ATR-FTIR/machine learning approach closer to a level suitable for use as a screening or secondary method to support HPLC-DAD in the quality control of Parkinson's disease tablets containing LD/BZ or LD/CD.

3.7. Greenness assessment of developed ATR-FTIR/machine learning approach

In order to quantify the greenness in line with recent recommendations for greenness assessment,[50–52] the ATR-FTIR/machine-learning workflow and the HPLC-DAD reference method were evaluated using the AGREE metric, which summarises the 12 principles of green analytical chemistry into a single score and pictogram. Because the proposed ATR-FTIR/machine learning workflow required a relatively large number of samples for model development and calibration transfer (103 synthetic samples and 6 real samples; $n = 109$), comparison of method greenness based only on analysis of the target samples would have favoured HPLC-DAD unfairly, since fewer calibration points are typically required for that method. The greenness of the ATR-FTIR/machine learning and HPLC-DAD approaches was therefore evaluated under a strategy that explicitly included both calibration and analysis of three real samples. For this purpose, sample size, waste volume, and analysis time were entered into the AGREE tool as average values calculated for three pellet samples, based on totals of 109 calibration samples + 3 real samples for ATR-FTIR and 7 calibration samples + 3 real samples for HPLC-DAD. The detailed AGREE scoring parameters for ATR-FTIR/hurdling model is showed in Table S8 in SI.

Under these conditions, both methods yielded AGREE scores above 0.5, whereas the ATR-FTIR approach showed clearly superior greenness relative to HPLC-DAD (0.75 *vs.* 0.57, Fig. 5). For ATR-FTIR, criterion 3 received a score of 0 because measurements were performed in an off-line configuration rather than *in situ*, although this limitation could be mitigated with portable FTIR systems. Criterion 8 also received a score of 0 because inclusion of the 109 calibration samples increased the average analysis time for the three evaluated samples to 150 min; however, this penalty is expected to decrease as sample throughput increases, given that the actual measurement time per sample was only 4 min. The HPLC-DAD score differed only slightly from that reported previously[18] (0.57 *vs.* 0.58), mainly because the present assessment was based on analysis of a 30 mg pellet rather than a 500 mg tablet and explicitly incorporated the calibration curve. Taken together, these

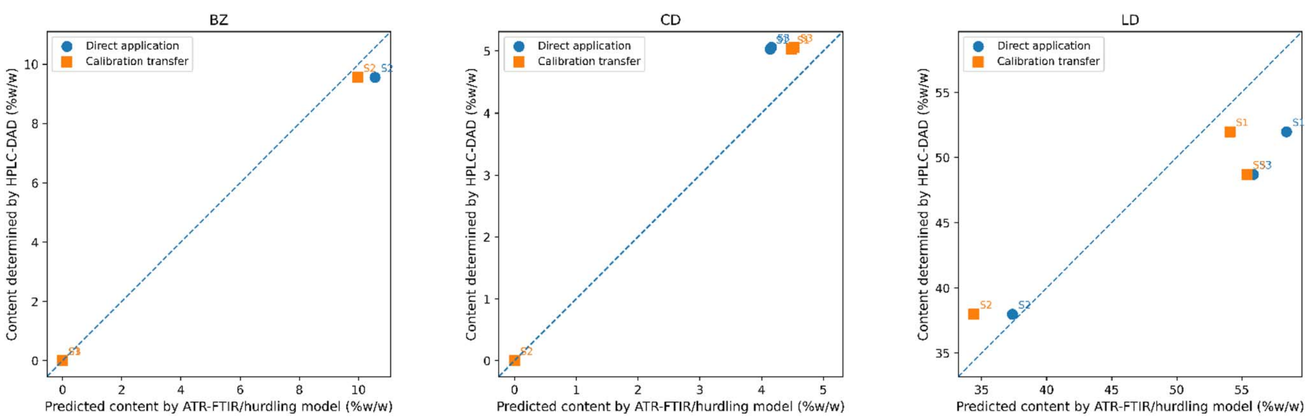

Fig. 4 Correlation plot between the analytical results of ATR-FTIR/hurdling model and HPLC-DAD method for a set of three tablet samples.

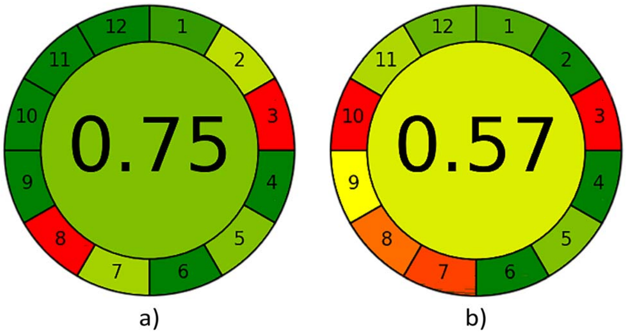

Fig. 5 Greenness picogram of (a) ATR-FTIR/machine learning and (b) HPLC-DAD methods for simultaneous determination of LD, CD, and BZ scored by AGREE calculator tool.

results indicate that, despite the relatively large sample requirements for model construction, the ATR-FTIR/machine learning strategy developed in the present study remains a rapid and environmentally profitable approach for pharmaceutical screening and quality-control applications.

3.8. Limitations and future development

Despite the promising performance of the proposed ATR-FTIR/machine learning workflow, several limitations of this study should be recognised. First, the commercial validation set remained small, with only nine samples included for calibration transfer and external validation. Although the dataset was sufficient to demonstrate the value of calibration transfer and to demonstrate different performance patterns for LD, BZ, and CD, it was not large enough to support a general claim of robustness across the wider range of commercially available anti-Parkinson's formulations. Second, the background matrix of synthetic samples was generated from a fixed excipient profile, which may differ in excipient composition and ratios from those of the commercial products. Such variation increased the risk of domain mismatch and likely contributed to the systematic overestimation of LD and underestimation of CD in the direct application scenario. Recovery experiments in a synthetic excipient matrix have been carried out, but additional formulation series with more diverse excipient profiles will still be required to fully characterise API selectivity and robustness of the models across different manufacturing conditions. A further limitation lies in the scope of the modelling. Although nested group cross-validation was applied to avoid overfitting, the model space remained largely restricted to conventional chemometric and machine learning algorithms, namely PLS/PCR, SVM, and RF. More advanced approaches, including XGBoost and 1D-CNN, were not examined, partly because the current sample size was insufficient to support them.

Future development should therefore focus on expanding the calibration domain by adding synthetic tablets with various excipient profiles, as well as additional commercial brands and production batches. With such extensions, the present workflow could evolve into a broadly applicable framework for multi-API quantitative solid pharmaceutical quality control.

4. Conclusions

In this work, an integrated ATR-FTIR/machine learning workflow was established for the simultaneous identification and quantification of levodopa, carbidopa and benserazide in solid oral Parkinson's medicines. A calibration set of 103 synthetic pellets and a small but representative set of commercial pellets were used to evaluate both the internal performance and the behaviour of the models under domain shift between laboratory mixtures and real products. A two-stage hurdle structure was adopted to handle the structural zeros inherent to multi product data, separating presence/absence decisions from quantitative prediction for positive samples. Systematic comparison of eight preprocessing pipelines and several classifier–regressor families under repeated nested group cross-validation showed that SG first derivative with SNV on mean spectra in the 675–1800 cm^{-1} region, combined with single-output RF regression and appropriate classifiers, offered the most favourable compromise between prediction error and stability.

Direct application of the final model to commercial pellets yielded excellent classification of API presence/absence, whereas quantitative performance varied by analyte and formulation. When a repeated-product balanced calibration transfer strategy was implemented with only six additional commercial pellets, median biases were reduced for all three APIs, with a particularly pronounced improvement for CD. The results demonstrated that ATR-FTIR, combined with a carefully validated hurdle model framework, can provide a rapid and green screening tool for multi-API Parkinson's medications and suggested a general strategy that may be extended to other complex solid dosage forms once broader calibration domains and more extensive calibration transfer schemes are implemented.

Author contributions

Manh Huy Nguyen: writing – original draft, writing – review and editing, project administration, funding acquisition, investigation, formal analysis, data curation. Thanh Dam Nguyen: writing – original draft, writing – review and editing, methodology, visualisation, software, conceptualisation. Hong Anh Duong: writing – review and editing, validation, resources. Hung Viet Pham: writing – review and editing, funding acquisition, supervision, resources.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data supporting this article have been included as part of the supplementary information (SI). Supplementary information is available. See DOI: https://doi.org/10.1039/d6ra02243j.

Acknowledgements

This research has been done under the research project QG.23.11 of Vietnam National University, Hanoi.

References

1 World Health Organization, https://www.who.int/news-room/fact-sheets/detail/parkinson-disease, accessed March 2025.
2 T. A. Zesiewicz, *Continuum*, 2019, **25**, 896–918.
3 S. Chapuis, L. Ouchchane, O. Metz, L. Gerbaud and F. Durif, *Mov. Disord.*, 2005, **20**, 224–230.
4 S. Ramesh and A. S. P. M. Arachchige, *AIMS Neurosci.*, 2023, **10**, 200.
5 C. W. Olanow, *Mov. Disord.*, 2019, **34**, 812–815.
6 D. B. Calne, *N. Engl. J. Med.*, 1993, **329**, 1021–1027.
7 S. Fahn, in *Parkinson's Disease*, CRC Press, 2007, pp. 69–90.
8 J. A. Obeso, C. W. Olanow and J. G. Nutt, *Trends Neurosci.*, 2000, **23**, S2–S7.
9 A. Di Stefano, P. Sozio and L. S. Cerasa, *Molecules*, 2008, **13**, 46–68.
10 S.-P. Khor and A. Hsu, *Curr. Clin. Pharmacol.*, 2007, **2**, 234–243.
11 USP-NF, https://www.uspnf.com/sites/default/files/usp_pdf/EN/USPNF/revisions/carbidopa-and-levodopa-ert-pending-nitr.pdf, accessed March 2025.
12 D. Sravanthi, M. Anusha, S. Madhavi, S. Firdose and B. N. Nalluri, *J. Chem. Pharm. Res.*, 2013, **5**, 422–428.
13 F. Belal, F. Ibrahim, Z. Sheribah and H. Alaa, *J. Chromatogr. B*, 2018, **1091**, 36–45.
14 S. M. Mahgoub, M. R. Mahmoud, A. Y. Binsaleh, M. A. Almalki, M. A. Mohamed and H. F. Nassar, *Sustainable Chem. Pharm.*, 2023, **36**, 101291.
15 F. Bugamelli, C. Marcheselli, E. Barba and M. A. Raggi, *J. Pharm. Biomed. Anal.*, 2011, **54**, 562–567.
16 İ. Bulduk and S. Gökçe, *Hacettepe J. Biol. Chem.*, 2021, **49**, 413–422.
17 K. V. Özdokur, E. Engin, Ç. Yengin, H. Ertaş and F. N. Ertaş, *Anal. Lett.*, 2018, **51**, 73–82.
18 T. D. Nguyen, T. H. Tu, T. M. T. Nguyen, M. H. Nguyen, H. A. Duong and H. V. Pham, *J. Chromatogr. A*, 2025, 466443.
19 A. Biancolillo and F. Marini, *Front. Chem.*, 2018, **6**, 576.
20 Y. Song, Y. Cong, B. Wang and N. Zhang, *Expert Opin. Drug Delivery*, 2020, **17**, 551–571.
21 T. R. Devi and S. Gayathri, *Int. J. Pharm. Sci. Rev. Res.*, 2010, **2**, 106–110.
22 A. A. Bunaciu, H. Y. Aboul-Enein and S. Fleschin, *Appl. Spectrosc. Rev.*, 2010, **45**, 206–219.
23 S. G. Kazarian, K. A. Chan and F. H. Tay, *Infrared and Raman Spectroscopic Imaging*, 2009, pp. 347–375.
24 G. Lawson, J. Ogwu and S. Tanna, *PLoS One*, 2018, **13**, e0202059.
25 L. Christie, S. Rutherford, D. S. Palmer, M. J. Baker and H. J. Butler, *Front. Bioeng. Biotechnol.*, 2024, **12**, 1349473.
26 A. Cherniienko, R. Lesyk, L. Zaprutko and A. Pawełczyk, *J. Pharm. Anal.*, 2024, **14**, 100951.
27 N. L. Calvo, T. S. Kaufman and R. M. Maggio, *J. Pharm. Biomed. Anal.*, 2015, **107**, 419–425.
28 L. Miloudi, F. Bonnier, K. Barreau, D. Bertrand, X. Perse, F. Yvergnaux, H. J. Byrne, I. Chourpa and E. Munnier, *Analyst*, 2018, **143**, 2377–2389.
29 J. Tashkhourian, M. Hormozi-Nezhad and J. Khodaveisi, *Spectrochim. Acta, Part A*, 2011, **82**, 25–30.
30 M. Chamsaz, A. Safavi and J. Fadaee, *Anal. Chim. Acta*, 2007, **603**, 140–146.
31 K. M. Ali, M. H. Mashhadizadeh, A. M. Mazloum and N. Sahraei, *J. Food Drug Anal.*, 2008, **16**, 9.
32 M. M. Said, S. Gibbons, A. C. Moffat and M. Zloh, *Int. J. Pharm.*, 2011, **415**, 102–109.
33 D. Custers, T. Cauwenbergh, J. Bothy, P. Courselle, J. De Beer, S. Apers and E. Deconinck, *J. Pharm. Biomed. Anal.*, 2015, **112**, 181–189.
34 T. Madrakian and M. Mohammadnejad, *Chem. Pharm. Bull.*, 2007, **55**, 865–870.
35 M. F. Abdel-Ghany, L. A. Hussein, M. F. Ayad and M. M. Youssef, *Spectrochim. Acta, Part A*, 2017, **171**, 236–245.
36 W. H. Kim, M. M. Karim and S. H. Lee, *Anal. Chim. Acta*, 2008, **619**, 2–7.
37 Y. X. Zhang, Z. J. Zhang and F. Yang, *Chin. J. Chem.*, 2008, **26**, 489–494.
38 P. T. T. Ha, A. Van Schepdael, T. Hauta-aho, E. Roets and J. Hoogmartens, *Electrophoresis*, 2002, **23**, 3404–3409.
39 A. M. Zeid, J. J. M. Nasr, F. F. Belal, S. Kitagawa, N. Kaji, Y. Baba and M. I. Walash, *RSC Adv.*, 2016, **6**, 17519–17530.
40 C. Zapata-Urzúa, M. Pérez-Ortiz, M. Bravo, A. Olivieri and A. Álvarez-Lueje, *Talanta*, 2010, **82**, 962–968.
41 A. Babaei and M. Babazadeh, *Electroanalysis*, 2011, **23**, 1726–1735.
42 M. Mazloum-Ardakani, B. Ganjipour, H. Beitollahi, M. K. Amini, F. Mirkhalaf, H. Naeimi and M. Nejati-Barzoki, *Electrochim. Acta*, 2011, **56**, 9113–9120.
43 M. Khanmohammadi, E. Mobedi, A. B. Garmarudi, H. Mobedi and K. Kargosha, *Pharm. Dev. Technol.*, 2007, **12**, 573–580.
44 B. Zimmermann and A. Kohler, *Appl. Spectrosc.*, 2013, **67**, 892–902.
45 S. M. Eid, S. S. Soliman, M. R. Elghobashy and O. M. Abdalla, *Vib. Spectrosc.*, 2020, **106**, 102995.
46 A. P. Fellows, M. T. L. Casford and P. B. Davies, *Appl. Spectrosc.*, 2020, **74**, 597–615.
47 M. S. Dhanoa, S. J. Lister, R. Sanderson and R. J. Barnes, *J. Near Infrared Spectrosc.*, 1994, **2**, 43–47.
48 Å. Rinnan, F. van den Berg and S. B. Engelsen, *TrAC, Trends Anal. Chem.*, 2009, **28**, 1201–1222.
49 C. Mas, L. Rubio, L. Valverde-Som, L. A. Sarabia and M. C. Ortiz, *Chemom. Intell. Lab. Syst.*, 2020, **201**, 104006.
50 F. Pena-Pereira, W. Wojnowski and M. Tobiszewski, *Anal. Chem.*, 2020, **92**, 10076–10082.
51 M. Sajid and J. Plotka-Wasylka, *Talanta*, 2022, **238**, 123046.
52 L. Yin, L. Yu, Y. Guo, C. Wang, Y. Ge, X. Zheng, N. Zhang, J. You, Y. Zhang and M. Shi, *J. Pharm. Anal.*, 2024, **14**, 101013.