Research Square

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

# A Comparative Evaluation of Sample Selection Algorithms for Multivariate Calibration in Near-Infrared Spectroscopic Analysis of Pharmaceutical Formulations

**Amanita Sow**

aminasow100@gmail.com

University of Sciences, Techniques and Technologies of Bamako

**Harouna Sangaré**

University of Sciences, Techniques and Technologies of Bamako

**Fadaba Danioko**

Centre d'intelligence artificielle et de robotique (CIAR-Mali)

**Tidiane Diallo**

University of Sciences, Techniques and Technologies of Bamako

**Research Article**

**Keywords:** Near-infrared spectroscopy, chemometrics, multivariate calibration, sample selection algo rithms, Kennard–Stone algorithm, pharmaceutical analysis, Gaussian process regression

**Posted Date:** September 9th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-10490758/v1

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

**Additional Declarations:** No competing interests reported.

# A Comparative Evaluation of Sample Selection Algorithms for Multivariate Calibration in Near-Infrared Spectroscopic Analysis of Pharmaceutical Formulations

Aminata Sow[1,*], Harouna Sangaré[2], Fadaba Danioko[3], and Tidiane Diallo[4]

[1]Department of Physics, Faculty of Sciences and Techniques, University of Sciences, Techniques and Technologies of Bamako, Bamako, Mali
[2]Department of Mathematics and Computer Sciences, Faculty of Sciences and Techniques, University of Sciences, Techniques and Technologies of Bamako, Bamako, Mali
[3]Centre of Artificial Intelligence and Robotics of Mali (CIAR-Mali), Bamako, Mali
[4]Department of Drug Sciences, Faculty of Pharmacy, University of Sciences, Techniques and Technologies de Bamako, Bamako, Mali

[*]Corresponding author: Dr. Aminata Sow
Email: aminasow100@gmail.com; aminata.sow@usttb.edu.ml

Date: August 7, 2026

**Abstract**

The construction of robust multivariate calibration models for near-infrared (NIR) spectroscopic analysis necessitates careful partitioning of samples into training and validation sets. The selection strategy employed fundamentally influences model generalizability and predictive accuracy. This investigation presents a systematic comparative analysis of four established sample selection algorithms—Duplex, Honigs, Kennard–Stone, and Naes—applied to NIR spectral data acquired from 58 commercial paracetamol tablets. Gaussian process regression (GPR) served as the modeling framework, with model performance quantified through the coefficient of determination ($R^2$) and root mean square error of prediction (RMSEP). The Kennard–Stone algorithm employing the Mahalanobis distance metric demonstrated superior performance, yielding optimal validation statistics ($R^2 = 0.99999$, RMSEP $= 1.74 \times 10^{-6}$). Rigorous non-parametric statistical analysis employing Kruskal–Wallis and post-hoc Mann–Whitney U tests with Bonferroni correction confirmed significant performance differences among algorithms ($p < 0.001$), while revealing statistical equivalence between Kennard–Stone and Honigs methods. Systematic investigation of training set proportions (60–90%) elucidated the monotonic relationship between calibration set size and predictive accuracy. These findings provide evidence-based guidance for optimizing sample selection protocols in pharmaceutical NIR applications and underscore the critical importance of chemometric validation in spectroscopic method development.

**Keywords:** Near-infrared spectroscopy; chemometrics; multivariate calibration; sample selection algorithms; Kennard–Stone algorithm; pharmaceutical analysis; Gaussian process regression

**Graphical Abstract:** Systematic evaluation of four sample selection algorithms for NIR spectroscopic calibration demonstrates the superiority of Kennard–Stone methodology for pharmaceutical paracetamol analysis.

# 1 Introduction

Near-infrared (NIR) spectroscopy has emerged as a preeminent analytical technique for pharmaceutical quality control, offering rapid, non-destructive analysis with minimal sample preparation. When coupled with chemometric methodologies, NIR spectroscopy enables accurate quantification of active pharmaceutical ingredients in complex matrices. The development of reliable multivariate calibration models, however, presents several methodological challenges that must be systematically addressed to ensure robust analytical performance.

Contemporary chemometric practice recognizes three principal challenges in multivariate calibration development: (i) the rational division of available samples into calibration (training) and validation sets, (ii) the selection of informative spectral regions or wavelengths, and (iii) the identification of optimal regression or classification algorithms. Previous investigations by our group have addressed the latter two challenges, demonstrating the efficacy of Gaussian process regression for NIR-based paracetamol quantification and exploring artificial intelligence applications for spectral analysis. [1, 2, 3] The present investigation focuses specifically on the first challenge—sample set partitioning—which fundamentally constrains the generalizability and predictive accuracy of subsequent models.

The theoretical foundation for rational sample selection derives from the principle that calibration sets must adequately represent the variability present in the population to which the model will be applied. [4] Inadequate representation leads to model overfitting, where excellent calibration statistics mask poor predictive performance on novel samples. Conversely, unnecessarily large calibration sets increase analytical costs without proportional improvement in model accuracy. Several algorithms have been developed to address this optimization problem, including the Kennard–Stone algorithm, which selects samples maximizing the spectral space coverage; the Honigs method, which iteratively selects spectrally distinct samples; the Duplex algorithm, which simultaneously constructs balanced calibration and validation sets; and the Naes clustering-based approach, which ensures representation from all sample subtypes. [6, 5, 7, 8]

While previous comparative studies have examined these algorithms in various contexts, [9, 10] several gaps persist in the literature. First, rigorous statistical validation of observed performance differences remains uncommon, with many studies relying on descriptive comparisons rather than hypothesis testing. Second, systematic investigation of distance metric effects—particularly the comparison between Euclidean and Mahalanobis distances—has received limited attention despite its theoretical importance for correlated spectral data. Third, the influence of training set proportion on relative algorithm performance merits systematic examination across multiple split ratios.

The present investigation addresses these gaps through comprehensive evaluation of the four aforementioned algorithms applied to a pharmaceutical NIR dataset comprising 58 commercial paracetamol samples. Specific objectives include: (1) quantitative comparison of algorithm performance using GPR modeling with $R^2$ and RMSEP metrics; (2) statistical validation of observed differences through non-parametric hypothesis testing with appropriate multiple comparison corrections; (3) systematic examination of training set size effects across proportions from 60% to 90%; and (4) investigation of distance metric influence on Kennard–Stone and Honigs algorithm performance. By addressing these objectives, this work aims to provide methodologically sound guidance for practitioners developing NIR-based pharmaceutical assays.

# 2 Materials and Methods

## 2.1 Sample Collection and Reference Analysis

Fifty-eight commercial paracetamol tablets (500 mg nominal content) representing distinct batch numbers were randomly acquired from 29 community pharmacies in Bamako, Mali, ensuring representative sampling of the local pharmaceutical market. Reference paracetamol content determination followed validated pharmacopoeial methodology. [11] Briefly, each sample (20 tablets, finely triturated) was accurately weighed and dissolved in 0.1 N sodium hydroxide solution. Optical density measurements were performed using an Agilent Cary 60 UV-Vis spectrophotometer (Agilent Technologies, Santa Clara, CA, USA). The mean of ten replicate scans provided the optical density value, from which paracetamol content was calculated, yielding a vector of 58 reference concentration values.

## 2.2 NIR Spectral Acquisition

Powdered samples were passed through a 250 $\mu$m sieve to ensure particle size uniformity. Approximately 0.206 g of each sieved sample was subjected to NIR reflectance spectroscopy using an optical flame-NIR-INTSMAS25 spectrometer (Ocean Insight, Orlando, FL, USA) operating from 800–1700 nm with a diffuse reflectance probe. Final spectra, comprising the average of 10 individual scans, covered the wavelength range 930–1800 nm, generating a data matrix of dimensions 58 samples $\times$ 128 wavelength variables. Raw reflectance (R) measurements were transformed to pseudo-absorbance (A) via $A = \log_{10}(1/R)$.

## 2.3 Spectral Preprocessing

All preprocessing operations were implemented in R (version 4.2.1, R Foundation for Statistical Computing, Vienna, Austria) utilizing the `prospectr` [12] and `signal` packages. The following preprocessing techniques were systematically evaluated: Standard Normal Variate (SNV) correction for scatter removal; first and second derivatives computed via Savitzky–Golay algorithm (window size: 11 points, polynomial order: 2) for baseline correction and resolution enhancement; Multiplicative Scatter Correction (MSC) for physical light scattering compensation; and Savitzky–Golay smoothing (window size: 9 points, polynomial order: 2) for noise reduction. Consistent with previous findings, [2] smoothed spectra yielded optimal GPR performance and were consequently employed for all subsequent analyses.

## 2.4 Sample Selection Algorithms

Four established algorithms for partitioning spectral data into calibration and validation sets were implemented and systematically compared. Each algorithm was evaluated using both Euclidean and Mahalanobis distance metrics where applicable.

### 2.4.1 Kennard–Stone Algorithm

The Kennard–Stone algorithm [6] selects calibration samples maximizing multivariate space coverage. Initial selection identifies the two most distant samples based on Euclidean distance in the spectral space. Subsequent iterations select the sample exhibiting the maximum minimum distance to any previously selected sample. Selection continues until the predetermined calibration set size is achieved. This algorithm ensures comprehensive representation of the spectral variability.

### 2.4.2 Honigs Algorithm

The Honigs method [5] employs stepwise selection based on spectral distinctiveness. The algorithm iteratively selects the sample whose spectrum exhibits the largest absolute residual when projected onto the subspace defined by previously selected samples. This approach prioritizes spectrally unique samples that contribute maximal information to the calibration model.

### 2.4.3 Duplex Algorithm

The Duplex algorithm [7] extends the Kennard–Stone concept to simultaneously construct both calibration and validation sets. The two most distant samples are identified, with one assigned to calibration and the other to validation. Subsequent steps assign the sample farthest from existing calibration samples to the calibration set, and the sample farthest from existing validation samples to the validation set, continuing iteratively until desired set sizes are achieved.

### 2.4.4 Naes Algorithm

The Naes clustering-based method [8] first partitions the dataset into distinct clusters using k-means clustering on the spectral data. Representative samples—typically those nearest cluster centers—are then selected from each cluster for the calibration set, ensuring representation of all sample subtypes and minimizing extrapolation during prediction.

### 2.5 Gaussian Process Regression Modeling

Gaussian process regression (GPR) was selected as the modeling framework based on previous investigations demonstrating its superior performance for this dataset. [2] All models were implemented using the `kernlab` package [13] in R with a radial basis function (Gaussian) kernel. For each sample selection algorithm and training set proportion, GPR models were calibrated on the designated training set and validated on the corresponding test set.

Model performance was quantified using the coefficient of determination ($R^2$) and root mean square error (RMSE), calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{1}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{2}$$

where $y_i$ represents the reference concentration, $\bar{y}$ denotes the mean reference concentration, and $\hat{y}_i$ indicates the model-predicted concentration for the $i$-th sample. $N$ corresponds to the number of samples in the evaluated set.

### 2.6 Statistical Analysis

Rigorous statistical comparison of algorithm performance employed non-parametric methods to avoid distributional assumptions. Initial global comparison utilized the Kruskal–Wallis test on validation RMSE values obtained from multiple independent runs. Upon detecting significant global differences, post-hoc pairwise comparisons were conducted using Mann–Whitney U tests with Bonferroni correction for multiple testing. Statistical significance was defined as $p < 0.05$ after correction.

## 3 Results and Discussion

### 3.1 Comparative Algorithm Performance

Table I presents the GPR model performance metrics for each sample selection algorithm using an 80% training / 20% validation split, with random splitting included as a baseline comparator. The Kennard–Stone algorithm demonstrated superior predictive accuracy, yielding the lowest RMSEP ($1.74 \times 10^{-6}$) and highest validation $R^2$ (0.999992). The Honigs algorithm exhibited comparable performance ($R^2 = 0.999996$, RMSEP $= 1.76 \times 10^{-6}$), while Naes and Duplex algorithms produced notably higher validation errors. Random splitting yielded the poorest predictive performance, substantiating the value of rational sample selection methodologies.

Table 1: Comparative Performance Metrics for Sample Selection Algorithms (80% Training Proportion)

| Algorithm | Calibration | | Validation | |
|---|---|---|---|---|
| | $R^2$ | RMSE | $R^2$ | RMSEP |
| Kennard–Stone | 0.9999997 | $2.86 \times 10^{-7}$ | 0.999992 | $1.74 \times 10^{-6}$ |
| Honigs | 0.9999997 | $2.61 \times 10^{-7}$ | 0.999996 | $1.76 \times 10^{-6}$ |
| Naes | 1.0000000 | $1.45 \times 10^{-8}$ | 0.999984 | $2.16 \times 10^{-6}$ |
| Random | 0.9999915 | $2.11 \times 10^{-6}$ | 0.999966 | $2.30 \times 10^{-6}$ |
| Duplex | 1.0000000 | $1.65 \times 10^{-8}$ | 0.999932 | $2.81 \times 10^{-6}$ |

The elevated calibration performance coupled with degraded validation accuracy observed for Naes and Duplex algorithms suggests potential overfitting, wherein these methods select calibration samples that inadequately represent validation set variability despite excellent calibration statistics. This phenomenon underscores the importance of validation performance as the primary criterion for algorithm selection, as calibration metrics alone may prove misleading.

## 3.2 Statistical Validation of Performance Differences

The Kruskal–Wallis test applied to validation RMSE values revealed highly significant differences among algorithms ($\chi^2 = 54.32$, $p = 8.77 \times 10^{-9}$), rejecting the null hypothesis of equivalent performance. Post-hoc pairwise comparisons employing Mann–Whitney U tests with Bonferroni correction (Table II) elucidated the specific nature of these differences.

Table 2: Post-Hoc Pairwise Comparisons with Bonferroni Correction

| Comparison | $Z$ | Unadjusted $p$ | Adjusted $p$ |
|---|---|---|---|
| Duplex – Honigs | 5.399 | $6.69 \times 10^{-8}$ | $6.69 \times 10^{-7}$ |
| Duplex – Kennard–Stone | 5.307 | $1.11 \times 10^{-7}$ | $1.11 \times 10^{-6}$ |
| Honigs – Kennard–Stone | -0.092 | 0.927 | 1.000 |
| Duplex – Naes | 2.899 | $3.74 \times 10^{-3}$ | 0.037 |
| Honigs – Naes | -2.500 | 0.012 | 0.124 |
| Kennard–Stone – Naes | -2.408 | 0.016 | 0.160 |
| Duplex – Random | 1.733 | 0.083 | 0.830 |
| Honigs – Random | -3.666 | $2.46 \times 10^{-4}$ | $2.46 \times 10^{-3}$ |
| Kennard–Stone – Random | -3.574 | $3.51 \times 10^{-4}$ | $3.51 \times 10^{-3}$ |
| Naes – Random | -1.166 | 0.244 | 1.000 |

The post-hoc analysis reveals two principal findings. First, both Kennard–Stone and Honigs algorithms significantly outperformed Duplex, Naes, and Random methods ($p < 0.05$ after correction). Second, and notably, no statistically significant difference emerged between Kennard–Stone and Honigs algorithms ($p = 1.000$ after correction), indicating their statistical equivalence for this dataset despite numerical differences in point estimates. Figure 1 visually corroborates these findings, illustrating the RMSE distribution for each algorithm.

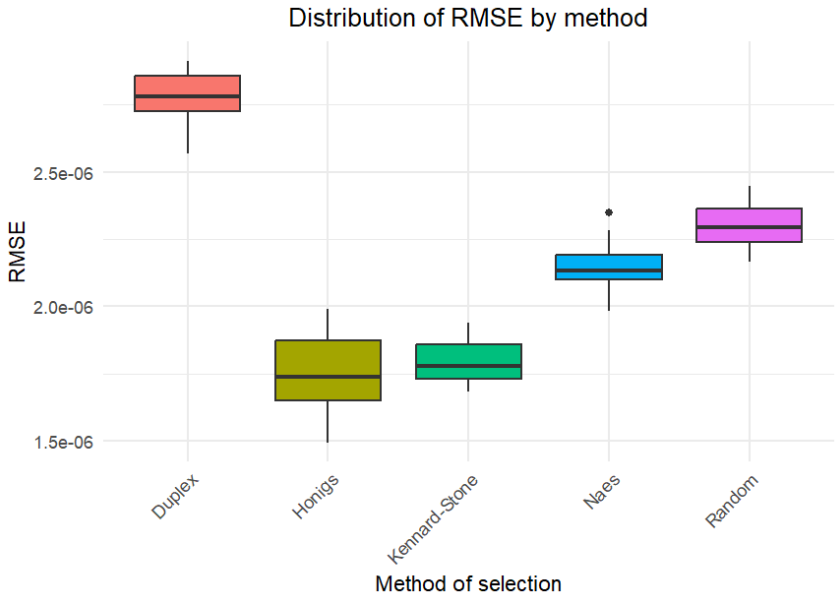



Figure 1: Distribution of validation RMSE values by sample selection algorithm, demonstrating the superior central tendency and reduced dispersion for Kennard–Stone and Honigs methods.

## 3.3 Influence of Training Set Size

Systematic variation of training set proportions from 60% to 90% enabled investigation of size effects on model performance for the two top-performing algorithms (Kennard–Stone and Honigs). Table III presents the complete results, while Figure 2 illustrates the observed trends.

Table 3: Training Set Size Effects on Model Performance

| Algorithm | Training (%) | Calibration | | Validation | |
|---|---|---|---|---|---|
| | | $R^2$ | RMSE | $R^2$ | RMSEP |
| Kennard–Stone | 90 | 0.9999997 | $2.55 \times 10^{-7}$ | 0.9999985 | $1.08 \times 10^{-6}$ |
| | 80 | 0.9999997 | $2.86 \times 10^{-7}$ | 0.9999921 | $1.74 \times 10^{-6}$ |
| | 75 | 0.9999996 | $3.19 \times 10^{-7}$ | 0.9999879 | $2.09 \times 10^{-6}$ |
| | 70 | 0.9999996 | $3.36 \times 10^{-7}$ | 0.9999619 | $2.60 \times 10^{-6}$ |
| | 65 | 0.9999996 | $3.54 \times 10^{-7}$ | 0.9999276 | $3.45 \times 10^{-6}$ |
| | 60 | 0.9999995 | $4.08 \times 10^{-7}$ | 0.9998787 | $4.38 \times 10^{-6}$ |
| Honigs | 90 | 0.9999997 | $2.53 \times 10^{-7}$ | 0.9999991 | $1.49 \times 10^{-6}$ |
| | 80 | 0.9999997 | $2.61 \times 10^{-7}$ | 0.9999957 | $1.76 \times 10^{-6}$ |
| | 75 | 0.9999996 | $2.97 \times 10^{-7}$ | 0.9999908 | $2.14 \times 10^{-6}$ |
| | 70 | 0.9999996 | $3.06 \times 10^{-7}$ | 0.9999823 | $2.50 \times 10^{-6}$ |
| | 65 | 0.9999996 | $3.25 \times 10^{-7}$ | 0.9999699 | $2.93 \times 10^{-6}$ |
| | 60 | 0.9999995 | $3.66 \times 10^{-7}$ | 0.9999078 | $4.99 \times 10^{-6}$ |

The observed monotonic relationship between training set size and predictive accuracy aligns with theoretical expectations from statistical learning theory. As calibration sets encompass greater sample variability, model generalizability proportionally improves. Notably, both algorithms exhibited asymptotic performance approaching the 90% training proportion, suggesting diminishing returns beyond this threshold.

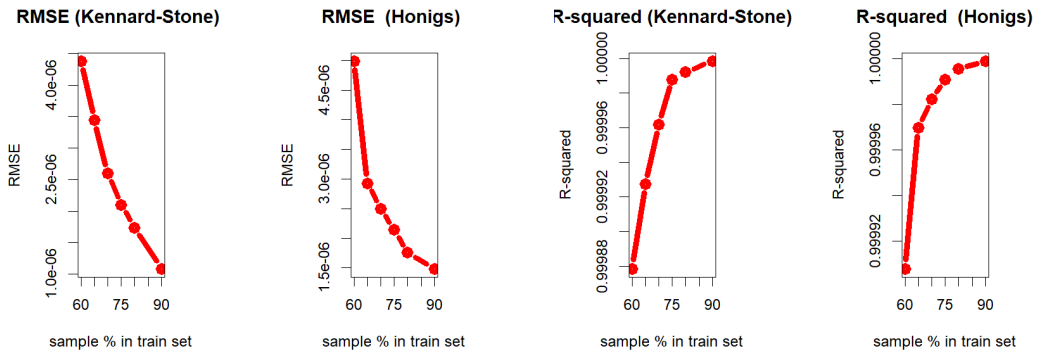



Figure 2: Relationship between training set proportion and validation performance metrics: (a) RMSEP demonstrating monotonic decrease with increasing training size; (b) $R^2$ exhibiting corresponding improvement. Red traces represent Kennard–Stone algorithm; green traces represent Honigs algorithm.

Figure 3 provides direct algorithm comparison across all training proportions, visually confirming the statistical equivalence established through hypothesis testing. The near-coincident performance trajectories substantiate the conclusion that both algorithms represent viable options for this application domain.

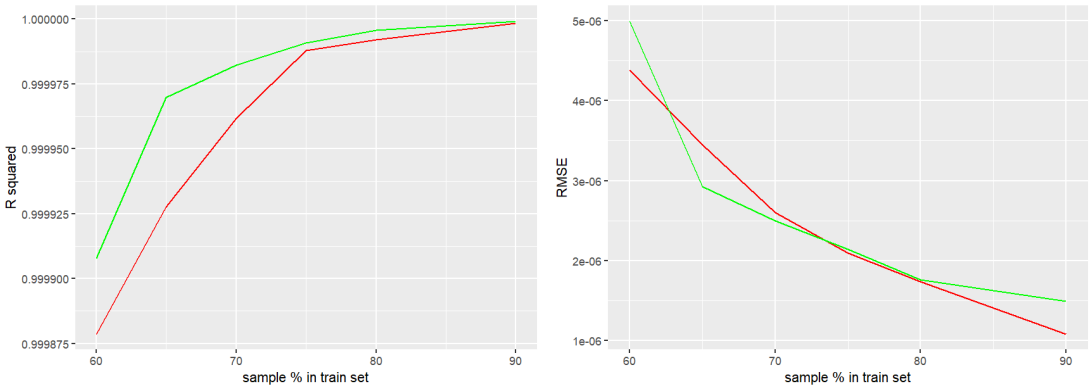



Figure 3: Direct comparison of Kennard–Stone (red) and Honigs (green) algorithms across training set proportions: (a) validation $R^2$; (b) validation RMSEP. The overlapping trajectories confirm statistical equivalence.

### 3.4 Distance Metric Effects

Investigation of Euclidean versus Mahalanobis distance metrics for Kennard–Stone and Honigs algorithms revealed differential effects (Table IV). For the Kennard–Stone algorithm, Mahalanobis distance yielded superior validation performance (RMSEP $1.74 \times 10^{-6}$) compared to Euclidean distance (RMSEP $1.84 \times 10^{-6}$). This finding accords with theoretical considerations: Mahalanobis distance accounts for inter-wavelength correlations inherent in NIR spectral data, providing more meaningful dissimilarity measures in high-dimensional correlated spaces. For the Honigs algorithm, both metrics performed comparably, with Euclidean distance showing a marginal advantage.

Table 4: Distance Metric Effects on Algorithm Performance (80% Training)

| Algorithm | Distance Metric | Calibration | | Validation | |
|---|---|---|---|---|---|
| | | $R^2$ | RMSE | $R^2$ | RMSEP |
| Kennard–Stone | Euclidean | 0.9999997 | $2.55 \times 10^{-7}$ | 0.999978 | $1.84 \times 10^{-6}$ |
| | Mahalanobis | 0.9999997 | $2.86 \times 10^{-7}$ | 0.999992 | $1.74 \times 10^{-6}$ |
| Honigs | Euclidean | 0.9999997 | $2.61 \times 10^{-7}$ | 0.999996 | $1.76 \times 10^{-6}$ |
| | Mahalanobis | 0.9999997 | $2.84 \times 10^{-7}$ | 0.999987 | $1.52 \times 10^{-6}$ |

### 3.5 Methodological Implications

The findings presented herein carry several implications for chemometric practice in pharmaceutical NIR analysis. First, the demonstrated superiority of rational selection algorithms over random splitting mandates incorporation of systematic sample selection protocols in method development workflows. Second, the statistical equivalence of Kennard–Stone and Honigs algorithms provides practitioners with flexibility in algorithm selection based on computational considerations or implementation convenience. Third, the demonstrated importance of distance metric selection—particularly the advantage of Mahalanobis distance for Kennard–Stone—highlights the need for careful attention to this methodological detail.

The observed overfitting tendency for Naes and Duplex algorithms warrants particular attention. While both methods produced excellent calibration statistics, their degraded validation performance suggests inadequate representation of validation set variability. This finding reinforces the principle that calibration metrics alone provide insufficient evidence of model quality and underscores the necessity of external validation.

# 4 Conclusions

This investigation provides a systematic, statistically rigorous comparison of four sample selection algorithms for NIR spectroscopic calibration in pharmaceutical analysis. The principal findings may be summarized as follows:

1. Rational sample selection algorithms significantly outperform random partitioning, with Kennard–Stone and Honigs methods demonstrating superior predictive accuracy for paracetamol quantification.

2. Statistical analysis employing non-parametric methods with multiple comparison corrections reveals equivalence between Kennard–Stone and Honigs algorithms, while confirming their superiority over Duplex, Naes, and random approaches.

3. Training set size exhibits a monotonic positive relationship with model performance, with optimal results obtained at 80–90% training proportions for this dataset.

4. Distance metric selection influences algorithm performance: Mahalanobis distance proves advantageous for Kennard–Stone, while Honigs performs comparably with either metric.

5. The observed overfitting in Naes and Duplex algorithms underscores the critical importance of validation-based performance assessment in chemometric model development.

For practitioners developing NIR-based pharmaceutical assays, the Kennard–Stone algorithm employing Mahalanobis distance represents a well-validated, statistically supported approach to sample set partitioning. Future investigations should extend these findings to diverse pharmaceutical product classes and explore advanced optimization techniques such as genetic algorithm-based sample selection for further performance enhancement.


# Acknowledgments

The authors express their gratitude to the University of Sciences, Techniques and Technologies of Bamako and the Centre of Artificial Intelligence and Robotics of Mali (CIAR-Mali) for institutional support and research facilities.


# Author Contributions

**Aminata.SOW.**: Writing – Original Draft, Investigation (data acquisition and experimentation), Formal analysis (data analysis), Software (coding with Harouna.SANGARE.). **Harouna.SANGARE.**: Software (coding with Aminata.SOW.), Methodology. **Fadaba.DANIOKO.**: Writing – Review & Editing. **Tidiane.DDIALLO.**: Investigation (data acquisition with Aminata.SOW.), Writing – Review & Editing, Supervision. All authors reviewed and approved the final manuscript.

# Clinical Trial Registration

NOT APPLICABLE. This study does not involve a clinical trial.

# Consent to Publish

Consent to Publish declaration: not applicable.

# Ethics and Consent to Participate

Ethics and Consent to Participate declarations: not applicable.

# Declaration of Conflicting Interests

The authors declare no potential conflicts of interest with respect to the research, authorship, or publication of this article.

## Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## Supplemental Material

Supplemental materials, including R code for all analyses and preprocessing operations, are available from the corresponding author upon reasonable request.

## Data Availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

# References


[1] A. Sow, I. Traore, T. Diallo, A. Ba. "Analysis of Local Samples of Paracetamol at Bamako by Reflectance Near-Infrared Spectroscopy". *Sci. J. Chem.* 2022. 10(6): 202–210.

[2] A. Sow, I. Traore, T. Diallo, M. Traore, A. Ba. "Comparison of Gaussian process regression, partial least squares, random forest and support vector machines for a near infrared calibration of paracetamol samples". *Results Chem.* 2022. 4: 100508.

[3] A. Sow, T. Diallo. "Techniques of Artificial Intelligence Applied to Near-Infrared Spectra". arXiv preprint arXiv:2510.10638 [physics.optics]. 2025.

[4] E. Lopez, J. Etxebarria-Elezgarai, J.M. Amigo, A. Seifert. "The importance of choosing a proper validation strategy in predictive models. A tutorial with real examples". *Anal. Chim. Acta.* 2023. 1275: 341532.

[5] D.E. Honigs, G.M. Hieftje, H.L. Mark, T.B. Hirschfeld. "Unique-sample selection via near-infrared spectral subtraction". *Anal. Chem.* 1985. 57(12): 2299–2303.

[6] R.W. Kennard, L.A. Stone. "Computer Aided Design of Experiments". *Technometrics.* 1969. 11(1): 137–148.

[7] R.D. Snee. "Validation of Regression Models: Methods and Examples". *Technometrics.* 1977. 19(4): 415–428.

[8] T. Naes. "The design of calibration in near infra-red reflectance analysis by clustering". *J. Chemom.* 1987. 1(2): 121–134.

[9] X. Fu, Y. Ying, D. Yang. "A comparative study of representative subset selection for NIR model updating". In: *2011 Louisville, Kentucky, August 7-10, 2011*. American Society of Agricultural and Biological Engineers; 2011.

[10] R.K.H. Galvao, M.C.U. Araujo, G.E. Jose, M.J.C. Pontes, E.C. Silva, T.C.B. Saldanha. "A method for calibration and validation subset partitioning". *Talanta.* 2005. 67(4): 736–740.

[11] British Pharmacopoeia Commission. *British Pharmacopoeia 2020*. London: The Stationery Office; 2020.

[12] A. Stevens, L. Ramirez-Lopez. "An introduction to the prospectr package". *R Package Vignette, R Package Version 0.2.6*. 2022.

[13] A. Karatzoglou, A. Smola, K. Hornik, A. Zeileis. "kernlab – An S4 Package for Kernel Methods in R". *J. Stat. Softw.* 2004. 11(9): 1–20.