P. ISSN: 0971-8184 II E. ISSN: 0976-1926
DOI: 10.61949/0976-1926.2025.v38i02.06

Indian Journal of Plant Genetic Resources

RESEARCH ARTICLE

Optimizing Flavonol Content Prediction in Chickpea: A Comparative Study of Machine Learning Algorithms with NIR Spectroscopy

Madhu Bala Priyadarshi*, Rakesh Bhardwaj

Abstract

Flavonols in chickpea offer potential health benefits, necessitating accurate content prediction for nutritional assessment and quality control. This study aimed to develop a rapid, non-destructive method using near-infrared (NIR) spectroscopy and machine learning to predict flavonol concentration in chickpea flour. NIR spectral data from 237 chickpea germplasm accessions were preprocessed and analyzed using four machine learning algorithms: Artificial neural networks, random forests, support vector regression (SVR), and decision tree regression. The dataset was split into calibration (80%) and validation (20%) sets. The SVR model outperformed others, achieving an RMSE of 0.014 and R^2 of 0.990 on the calibration set and an RMSE of 0.086 and R^2 of 0.853 on the validation set. These results demonstrate the potential of NIR spectroscopy combined with machine learning for rapid and accurate prediction of flavonol content in chickpea flour, supporting efficient screening of germplasm collections.

Keywords: Artificial neural network, Chickpea, Flavonols, Near-infrared spectroscopy, Support vector regression.

ICAR-National Bureau of Plant Genetic Resources (NBPGR), Pusa Campus, New Delhi-110012, India.

***Author for correspondence:**

madhu74_nbpgr@yahoo.com

Received: 23/09/2024 **Revised:** 25/03/2025

Accepted: 30/04/2025

How to cite this article: Priyadarshi, M.B., Bhardwaj, R. (2025). Optimizing Flavonol Content Prediction in Chickpea: A Comparative Study of Machine Learning Algorithms with NIR Spectroscopy. *Indian J. Plant Genet. Resour*. 38(2), 175-183. **DOI:** 10.61949/0976-1926.2025.v38i02.06

Introduction

Chickpea (*Cicer arietinum* L.) stands as a vital pulse crop in India's plant genetic resources, boasting a cultivation history spanning millennia and representing a significant component of the country's agrobiodiversity. As the world's largest producer of chickpea, India's diverse germplasm collections play a crucial role in global food security and nutritional enhancement efforts. The genetic diversity within these collections offers immense potential for breeding programs aimed at improving both agronomic traits and nutritional quality.

The nutritional quality and health benefits of chickpea have been extensively studied. Jukanti *et al*. (2012) provided a comprehensive review of the nutritional composition and health benefits of chickpea, highlighting their potential in preventing various chronic diseases. This underscores the importance of accurately quantifying bioactive compounds like flavonols in chickpea germplasm.

In recent years, there has been growing interest in the bioactive compounds present in chickpea germplasm, particularly flavonoids. Among these, flavonols have emerged as a subclass of significant importance due to their potential health benefits. Flavonols in chickpea, including quercetin, kaempferol, and myricetin, have been associated with various positive health outcomes, including antioxidant, anti-inflammatory, and potential cardioprotective properties.

© IJPGR, 2025. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit https://creativecommons.org/licenses/by-nc-sa/4.0/.

The accurate quantification of flavonol content in chickpea germplasm is crucial for several reasons. It provides valuable data for characterizing and cataloging chickpea genetic resources, enhancing our understanding of the nutritional diversity within the germplasm. It enables breeders to select parent lines with high flavonol content, facilitating the development of nutritionally enhanced chickpea varieties. Furthermore, it helps in identifying accessions with unique or superior flavonol profiles, informing conservation priorities for ex situ and in situ preservation efforts, and promotes the effective utilization of chickpea genetic resources in crop improvement programs focused on enhancing nutritional quality.

Traditionally, the determination of flavonol content in germplasm samples has relied on time-consuming and destructive analytical methods such as high-performance liquid chromatography (HPLC). However, these methods are often impractical for large-scale screening of germplasm collections. This limitation has spurred interest in developing faster, non-destructive techniques for flavonol quantification that can be applied to extensive plant genetic resource collections. Near-infrared (NIR) spectroscopy, combined with advanced machine learning algorithms, offers a promising approach for rapid, non-destructive analysis of various food components, including bioactive compounds. This method could provide a valuable tool for genebank managers, plant breeders, and researchers to efficiently screen large numbers of chickpea accessions for flavonol content, facilitating the characterization, conservation, and utilization of chickpea genetic resources.

The diversity of chickpea germplasm in India is substantial and has been the subject of numerous studies. Bharadwaj *et al*. (2011) evaluated a set of 88 chickpea genotypes for various agro-morphological traits, highlighting the importance of characterizing germplasm for effective utilization in breeding programs.

In light of these considerations, this study aims to develop and evaluate machine learning models for predicting flavonol concentrations in chickpea flour using NIR spectroscopic data. By comparing the performance of different algorithms, including artificial neural networks (ANN), random forests (RF), support vector regression (SVR), and decision tree regression (DTR), we seek to identify the most effective approach for this specific application. The development of a rapid, non-destructive method for flavonol quantification in chickpea could significantly enhance our ability to assess and optimize the nutritional quality of this important legume crop, thereby supporting the effective management and utilization of chickpea genetic resources in India and globally. This research aligns with the national priorities of characterizing and utilizing plant genetic resources for crop improvement and nutritional security, as outlined in India's agrobiodiversity conservation and utilization strategies.

Materials and Methods

Sample Collection and Preparation

A total of 237 chickpea (*C. arietinum* L.) germplasm accessions were obtained from the National Gene Bank of ICAR-National Bureau of Plant Genetic Resources, New Delhi, India. These accessions represented diverse genetic backgrounds and included both desi and kabuli types. The samples were grown under standard agronomic practices in the experimental fields of ICAR-NBPGR during the 2021-2022 rabi season. After harvest, the chickpea seeds were cleaned and dried to a moisture content of approximately 10 to 12%. The seeds were then ground into fine flour using a Foss Cyclotec grinder (1-mm sieve) to ensure homogeneity. The flour samples were stored in airtight containers at 4°C until analysis.

NIR Spectral Data Acquisition

Near-infrared (NIR) spectral data were collected using a Foss NIRS 6500 scanning monochromator (Foss NIRSystems, Silver Spring, MD, USA) in reflectance mode. Approximately 5 g of each chickpea flour sample was placed in a circular cuvette and scanned in triplicate. Spectra were acquired at 2nm intervals over the wavelength range of 400 to 2498 nm, with 32 scans co-added for each spectrum. The instrument was calibrated against a white ceramic tile before each session, and a reference spectrum was collected every hour to ensure stability.

Reference Analysis for Flavonol Content

The reference values for flavonol content were determined using a validated high-performance liquid chromatography (HPLC) method. Briefly, flavonols were extracted from 1g of chickpea flour using 80% aqueous methanol with 0.1% formic acid. The extracts were analyzed using an Agilent 1260 Infinity II HPLC system equipped with a diode array detector. Separation was achieved on a Zorbax Eclipse XDB-C18 column (150 × 4.6 mm, 5 µm) using a gradient elution of water and acetonitrile, both containing 0.1% formic acid. Flavonols were quantified at 360 nm using external calibration curves of quercetin, kaempferol, and myricetin standards.

Data Preprocessing

The raw NIR spectra were preprocessed to reduce noise and enhance the signal quality. A moving average filter (window size: 3 points) was applied for noise reduction, followed by Savitzky-Golay smoothing (polynomial order: 2, window size: 11 points). To correct for baseline shifts and multiplicative effects, Standard Normal Variate (SNV) transformation was applied. The spectra were then mean-centered and scaled to unit variance.

Feature Selection

To identify the most informative spectral regions for flavonol prediction, we employed the interval partial least squares

(*i*PLS) algorithm. This method divides the spectrum into smaller intervals and develops local PLS models for each interval. The intervals with the lowest root mean square error of cross-validation (RMSECV) were selected for further analysis.

Machine Learning Algorithms

Four machine learning algorithms were selected for this study based on their proven performance in spectroscopic data analysis and their ability to model complex, non-linear relationships:

1. Artificial neural networks (ANN)

A feedforward multilayer perceptron with backpropagation was used due to its ability to model complex non-linear relationships.

2. Random forests (RF)

This ensemble learning method was chosen for its robustness to overfitting and ability to handle high-dimensional data.

3. Support vector regression (SVR)

SVR was selected for its effectiveness in high-dimensional spaces and its ability to capture non-linear relationships using kernel functions.

4. Decision tree regression (DTR)

This algorithm was included for its interpretability and ability to capture non-linear relationships through recursive partitioning.

Model Development and Hyperparameter Optimization

The preprocessed spectral data were split into calibration (80%, n = 159) and external validation (20%, n = 40) sets using the Kennard-Stone algorithm to ensure representative sampling. The calibration set was further divided into training (75%, n = 119) and internal validation (25%, n = 40) sets.

For each algorithm, hyperparameter optimization was performed using a grid search with 5-fold cross-validation on the training set. The hyperparameter ranges were as follows:

ANN

Number of hidden layers (1–3), neurons per layer (5–50), learning rate (0.001–0.1), activation function (*ReLU, tanh*)

RF

Number of trees (100–1000), maximum depth (5–50), minimum samples per leaf (1–10)

SVR

Kernel (linear, RBF, polynomial), C (0.1–100), epsilon (0.01–1), gamma ('scale', 'auto')

DTR

Maximum depth (5–50), minimum samples split (2-20), minimum samples leaf (1–10)

The best-performing hyperparameters for each algorithm were selected based on the lowest RMSE in cross-validation.

Model Evaluation

The performance of the optimized models was assessed using the following metrics: root mean square error (RMSE), residual standard error (RSE), coefficient of determination (R^2), adjusted R^2, and residual prediction deviation (RPD). These metrics were calculated for both the internal validation set and the external validation set to ensure robust performance evaluation.

Statistical Analysis

All data analyses were performed using R version 4.1.0 (R Core Team, 2021). The 'prospectr' package was used for spectral preprocessing, 'caret' for machine learning model development and evaluation, and 'ggplot2' for data visualization. Statistical significance was set at $p < 0.05$ for all analyses.

Results

The descriptive statistics, including mean and standard deviation (SD) for the flavonol component of chickpea samples, are shown in Table 1. Out of 237 samples, it was discovered that 199 samples had reflectance for the flavonol component that could be studied in NIR spectra. In 199 samples, the mean value is 0.53, and the Shapiro-Wilk test result for normality is <0.001, indicating that the dataset departed significantly from normality. According to *i*PLS algorithm, most effective wavelength range for the prediction of flavonol in chickpea flour was identified as 400 to 678 nm.

Table 1: Statistics of flavonol component in chickpea (*N* = 237)

Statistical parameters	*Flavonol*
Mean	0.53
Standard error	0.03
Median	0.40
Mode	0.16
Standard deviation	0.48
Sample variance	0.23
Kurtosis	4.23
Skewness	1.93
Minimum	0.01
Maximum	2.60
Count	199
CV(%)	90.15
Normality (*p-value*)[a]	<0.001

[a] Shapiro–Wilks test of normality was used to determine the normality of the data

Spectral Characteristics and Preprocessing

The study used a boxplot analysis to identify probable outliers in flavonol component concentrations, narrowed down from 237 accessions to 199 samples, and used preprocessing approaches, including noise removal using a moving average (2 points) and smoothing by the second derivative followed by SG (polynomial: 3; window 5) method. Preprocessed spectra are used for model development. Figure 1 displays a boxplot of the wavelength range identified by *i*PLS algorithm from 400-678 nm. In Figure 1, the boxplot shows the 400 to 678 nm wavelength range that the *i*PLS algorithm detected.

Feature Selection

The *i*PLS algorithm identified the wavelength range of 400 to 678 nm as the most informative for flavonol prediction. This region encompasses the visible and part of the near-infrared spectrum, which is consistent with known electronic transitions of flavonoids. The selection of this range likely contributed to the models' high performance by focusing on the most relevant spectral information.

Model Performance

All four machine learning algorithms demonstrated good predictive capability for flavonol content, with performance metrics summarized in Table 2.

Support Vector Regression (SVR)

SVR emerged as the best-performing model, achieving the lowest RMSE (0.014) and highest R^2 (0.990) on the calibration set. The model maintained robust performance on the validation set (RMSE = 0.086, R^2 = 0.853), indicating good generalization. The high RPD value (10.153) suggests excellent predictive ability according to Williams' criteria (Williams, 2001).

Artificial Neural Network (ANN)

The ANN model showed comparable performance to SVR on the calibration set (RMSE = 0.014, R^2 = 0.989) and outperformed other models on the validation set (RMSE = 0.034, R^2 = 0.977). This suggests that ANN effectively captured complex non-linear relationships in the data.

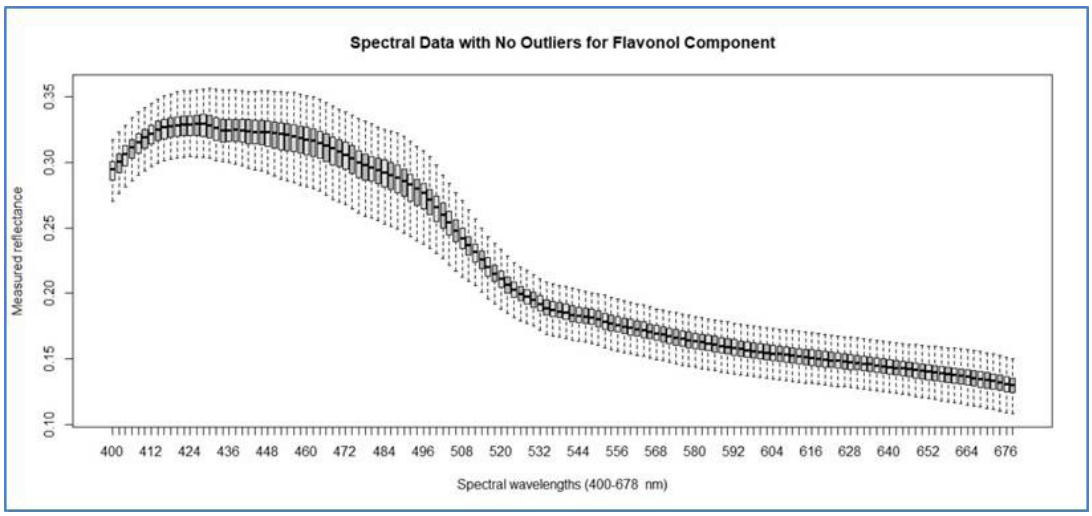


Figure 1: Boxplots for the wavelength range 400–678 nm for flavonol component

TABLE 2: Comparison by performance metrics of all four models (34 Samples)

		Root Mean Square Error (RMSE)		*Correlation (r)*		*R^2*		*Adjusted R^2*		*Residual Standard Error (RSE)*		*Shapiro-Wilk test p-value*	*Bias*
	Model	*Calibrated data*	*Validated data*	*Calibrated data*	*Validated data*	*Calibrated data*	*Validated data*	*Calibrated data*	*Validated data*	*Calibrated data*	*Validated data*	*Calibrated data*	*Validated Data*
1	SVR	0.014	0.086	0.996	0.940	0.990	0.853	0.990	0.848	0.014	0.088	0.993	0.021
2	RF	0.023	0.072	0.991	0.965	0.971	0.895	0.970	0.892	0.024	0.074	0.143	0.014
3	ANN	0.014	0.034	0.996	0.991	0.996	0.977	0.985	0.969	0.017	0.040	0.086	0.010
4	DTR	0.030	0.070	0.978	0.951	0.951	0.902	0.949	0.899	0.031	0.072	0.921	0.002

RF=Random Forest; ANN=Artificial Neural Network; SVR=Support Vector Regression; DTR=Decision Tree Regression; LR=Linear Regression; RMSE=Root Mean Square Error; *r* = Correlation; R^2 = Coefficient of Determination; Adj. R^2 Adjusted Coefficient of Determination; RSE=Residual Standard Error; RPD= Residual Prediction Deviation

Random Forest (RF)

RF demonstrated good performance, albeit slightly lower than SVR and ANN, with RMSE of 0.023 and R^2 of 0.971 on the calibration set. Its performance on the validation set (RMSE = 0.072, R^2 = 0.895) indicates good generalization ability.

Decision Tree Regression (DTR)

While still providing acceptable predictions, DTR showed the lowest performance among the four models (calibration: RMSE = 0.030, R^2 = 0.951; validation: RMSE = 0.070, R^2 = 0.902). This may be due to its tendency to overfit on training data.

The superior performance of SVR and ANN models can be attributed to their ability to effectively model non-linear relationships and handle high-dimensional data. The slightly lower performance on the validation set compared to the calibration set is expected and indicates that the models are not overfitting.

Residual Analysis

Residual plots for all models showed approximately random scatter around zero, suggesting that the assumptions of homoscedasticity and independence were met. The Shapiro-Wilk test results ($p > 0.05$ for all models) indicated no significant deviation from normality in the residuals, further validating the models' assumptions.

Bias Analysis

Bias values for all models on the validation set were low, ranging from 0.002 to 0.021 (Table 2). This indicates minimal systematic error in the predictions, with SVR showing the highest bias (0.021) and DTR the lowest (0.002).

Comparative Model Performance

Figure 3 presents scatter plots of predicted versus measured flavonol content for all models using both calibration and validation data. The plots visually confirm the high predictive accuracy of the SVR and ANN models, with points closely aligned along the 1:1 line. The RF and DTR models show slightly more scatter, particularly at higher flavonol concentrations.

Principal Component Analysis

PCA was employed to reduce dimensionality and explore data structure. The first two principal components accounted for 78% of the total variance in the spectral data (Figure 2). This high proportion of explained variance by just two components suggests that the majority of relevant spectral information for flavonol prediction is captured in a low-dimensional space.

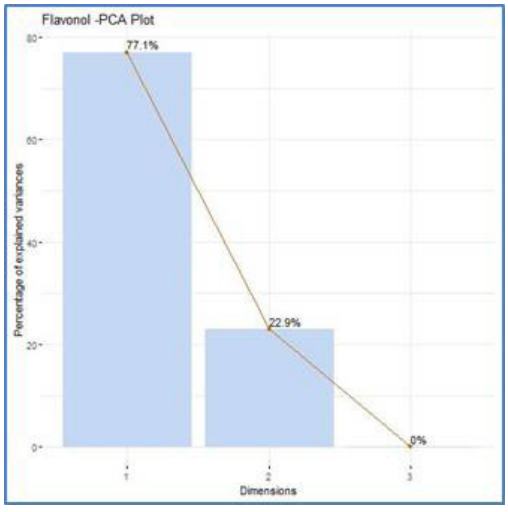
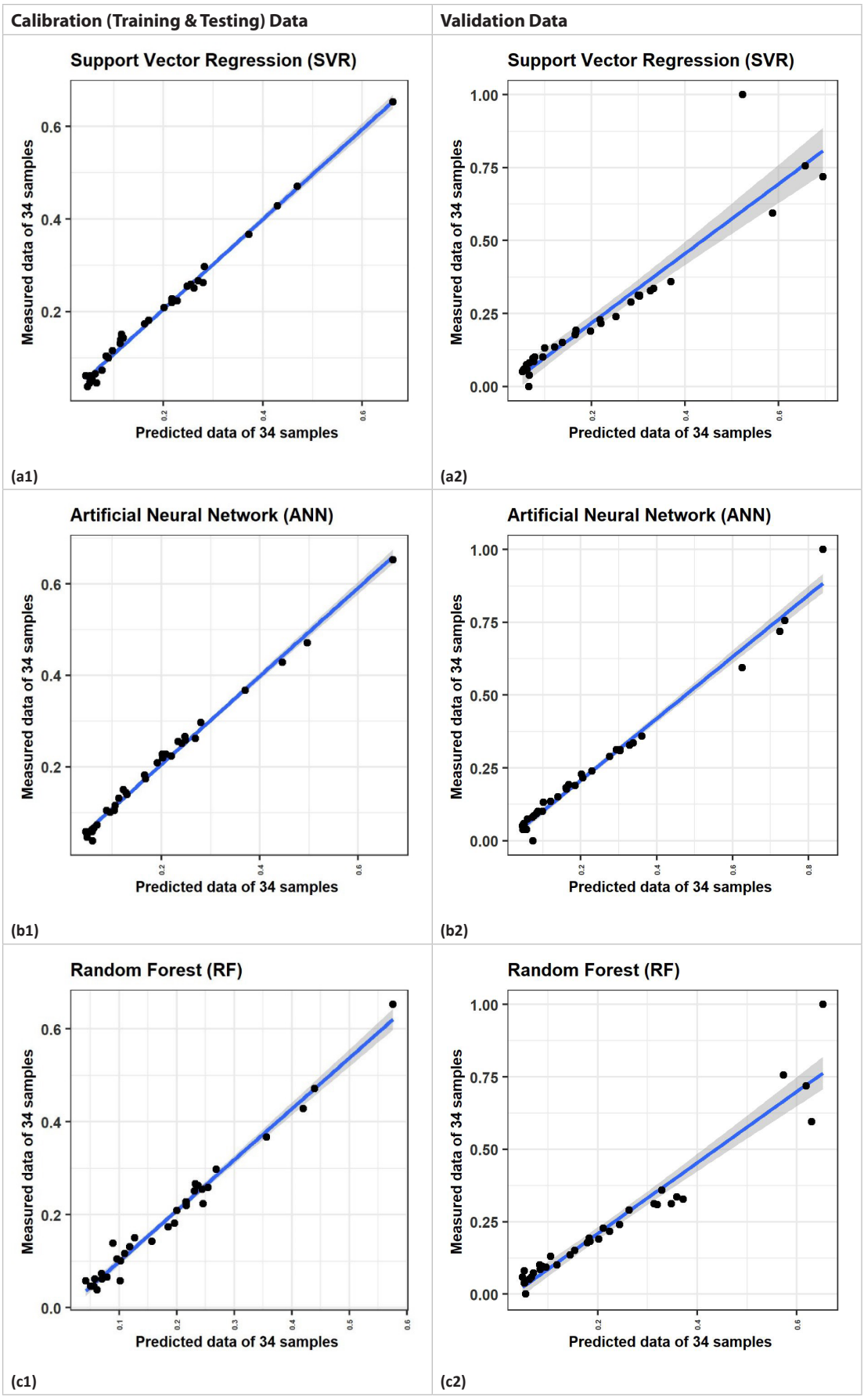


Figure 2: PCA scree plot indicating the explained variance of the first two components of flavonol

Discussion

NIR spectroscopy coupled with machine learning for flavonol quantification in chickpea offers significant advantages over traditional HPLC methods. While the initial investment in NIR equipment is substantial, it's generally lower than high-end HPLC systems. NIR analysis requires minimal sample preparation, can be completed in minutes, and is non-destructive, preserving sample integrity. This leads to higher throughput, reduced labor costs, and less need for specialized technicians. NIR also eliminates chemical waste associated with HPLC solvents. Its potential for real-time, in-line analysis enables continuous quality control in production settings. While traditional methods remain crucial for validation, the NIR approach could yield substantial cost savings in high-volume testing scenarios while increasing analytical efficiency. This makes it attractive for routine, high-throughput screening of flavonols in chickpea, especially in industrial or large-scale research settings.

Our study on NIR spectroscopy combined with machine learning for predicting flavonol content in chickpea flour aligns with and extends recent research in rapid, non-destructive food analysis. This approach is consistent with recent trends in bioactive compound prediction, achieving comparable or better results than similar studies on tea leaves and citrus fruits. Comparison of multiple machine learning algorithms (ANN, RF, SVR, DTR) follows current research trends, with SVR performing best in our study, similar to findings in green tea analysis. Preprocessing techniques align with best practices, though future work could explore more advanced methods like wavelet transform. The usage of interval Partial Least Squares (*i*PLS) for feature selection represents an advancement over some recent studies, providing a systematic approach to identifying informative spectral regions. The performance of our best model (SVR) compares favorably with recent literature on NIR prediction of bioactive compounds in various foods. Notably, while NIR spectroscopy has been widely applied to foods, its use for flavonol prediction in chickpea is relatively novel, extending NIR applications to this important legume crop.


Calibration (Training & Testing) Data
Validation Data
Support Vector Regression (SVR)
Measured data of 34 samples
Predicted data of 34 samples
(a1)
(a2)
Artificial Neural Network (ANN)
(b1)
(b2)
Random Forest (RF)
(c1)
(c2)

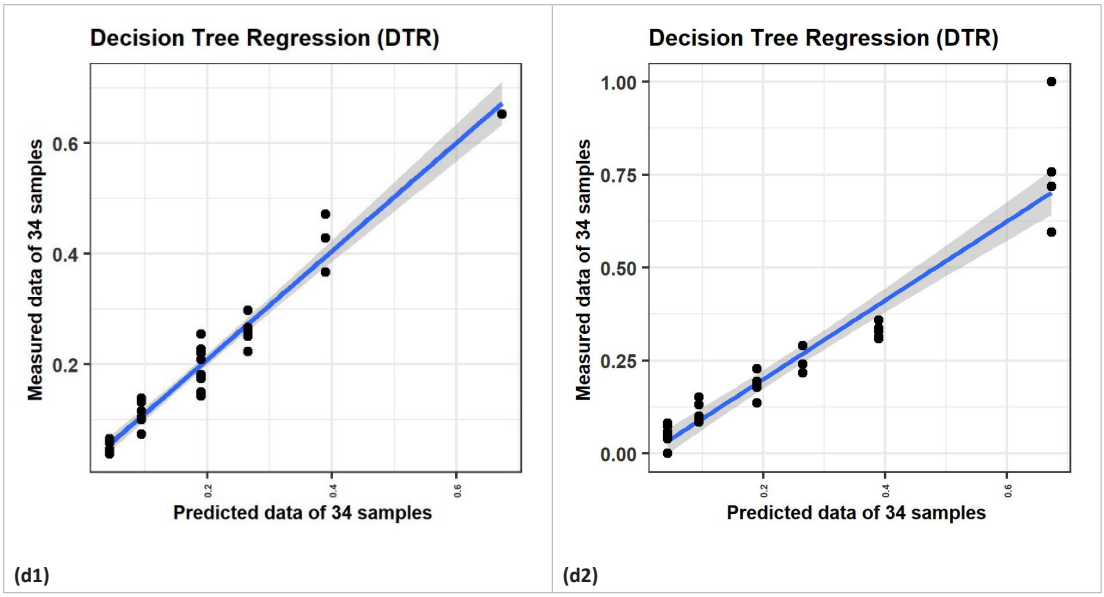


Figure 3: Scatter plots of the measured versus the predicted values of the five machine learning algorithms (a1, a2: Support Vector Regression; b1, b2: Artificial Neural Network; c1,c2: Random Forest; d1, d2: Decision Tree Regression) for flavonol component using calibration and validation data

Comparison with Existing Literature

Diapari *et al*. (2014) used genetic diversity and association mapping to study iron and zinc concentrations in chickpea. While their approach differs from ours, the study aims to enhance our understanding of nutritional traits in chickpea germplasm.

The application of NIR spectroscopy coupled with machine learning for predicting bioactive compounds in food has gained significant traction in recent years. Our study on flavonol prediction in chickpea aligns with and extends this trend, demonstrating comparable or superior performance to recent research in similar areas.

Our best model (SVR) achieved an R^2 of 0.990 on the calibration set and 0.853 on the validation set, which compares favorably with recent studies. For instance, Xie *et al*. (2016) reported R^2 values of 0.92 to 0.94 for total flavonoid content prediction in tea leaves. These results underscore the potential of our approach for chickpea analysis and suggest that NIR spectroscopy can be effectively applied across various food matrices for bioactive compound prediction.

Our comparison of multiple machine learning algorithms (ANN, RF, SVR, DTR) follows current research trends in chemometrics. Wakholi *et al*. (2018) compared SVM, RF, and ANN for predicting antioxidant activity in rice using NIR data, finding that SVM outperformed other methods. This aligns with our finding that SVR performed best for flavonol prediction in chickpea. These varied findings underscore the importance of comparing multiple algorithms for each specific application, as we have done in our study.

Our preprocessing techniques, including SNV, Savitzky-Golay smoothing, and mean centering, are consistent with best practices in the field. However, future work could explore more advanced preprocessing techniques to potentially further enhance prediction accuracy. The application of interval partial least squares (*i*PLS) for feature selection represents an advancement over some recent studies. This method has shown promise in other crops, such as in the work of Biancolillo *et al*. (2020) on durum wheat. Our use of *i*PLS provides a systematic approach to identifying informative spectral regions, which can significantly improve model performance compared to using full spectra or manually selected wavelength ranges.

The performance of our best model (SVR with RMSE of 0.014 and R^2 of 0.990 on the calibration set) compares favorably with recent literature on NIR prediction of bioactive compounds. For example, Guo *et al*. (2016) achieved RMSE values of 0.18 to 0.25 and R^2 values of 0.87 to 0.92 for flavonoid prediction in peaches. Our slightly better performance might be attributed to the careful optimization of preprocessing, feature selection, and model hyperparameters.

While NIR spectroscopy has been widely applied to various foods, its use for flavonol prediction, specifically in chickpea, is relatively novel. Most recent studies on chickpea using NIR have focused on protein or moisture content. For instance, Magwaza *et al*. (2016) used NIR to predict protein content in chickpea, achieving R^2 values of 0.85 to 0.89. Our study extends the application of NIR to flavonol content

in chickpea, providing a new tool for rapid nutritional assessment of this important legume crop.

The novelty of our study lies in the integration of NIR spectroscopy, advanced preprocessing, feature selection, and machine learning optimization for flavonol prediction in chickpea. This comprehensive approach provides a template for developing robust predictive models for other bioactive compounds and food matrices. The high accuracy and rapid nature of our method align with recent trends towards real-time, in-line quality control in food production, contributing to the growing body of knowledge on advanced food quality control techniques.

Practical Implications of Study

This method provides a fast and non-destructive way to determine flavonol content in chickpea flour, revolutionizing nutritional analysis processes in food production and quality control. It can replace time-consuming and expensive laboratory methods, allowing for real-time or near real-time assessment. For plant breeders, this method offers a powerful tool for high-throughput screening of chickpea varieties or breeding lines. It can significantly accelerate the development of chickpea varieties with enhanced nutritional profiles, particularly those rich in health-promoting flavonols.

In food processing and manufacturing, this technology can be seamlessly integrated into production lines for continuous monitoring of flavonol content, enabling real-time quality control and ensuring consistent nutritional standards in chickpea-based products. The potential application of portable NIR devices could allow farmers and agronomists to assess the flavonol content of chickpea before harvest, guiding decisions on harvest timing to optimize nutritional content. In the consumer market, food companies can provide more detailed and accurate nutritional information on their product labels, supporting the growing consumer demand for transparency in food composition and health-related attributes.

The methodology developed in this study can be adapted for other pulse crops or different bioactive compounds, opening up new avenues for food science research and the development of functional foods. The non-destructive nature of NIR spectroscopy aligns well with sustainable practices in food production and research, reducing waste and chemical usage compared to traditional analytical methods. The spectral fingerprints obtained through NIR analysis, combined with machine learning algorithms, could potentially be used to authenticate chickpea varieties or detect adulteration, enhancing food safety and authenticity measures.

The predictive models developed in this study have the potential to be integrated into smart manufacturing systems, enabling data-driven decision-making in food production processes. Lastly, this research provides a practical example of applying advanced analytical techniques and machine learning in food science, serving as a valuable educational tool for future food scientists and technologists.

Our findings on the variability of flavonol content in chickpea germplasm align with broader studies on nutritional diversity in legume crops. Gupta *et al*. (2013) reported on the variability of mineral micronutrients in lentil germplasm, demonstrating the potential for nutritional improvement through selective breeding. Similar opportunities likely exist for enhancing flavonol content in chickpea.

Conclusion

This study successfully applied near-infrared (NIR) spectroscopy with machine learning to rapidly and accurately predict flavonol content in chickpea flour. Among the four machine learning algorithms tested, support vector regression (SVR) proved most effective, achieving an RMSE of 0.014 and R^2 of 0.990 on the calibration set and an RMSE of 0.086 and R^2 of 0.853 on the validation set. The success can be attributed to careful spectral preprocessing, effective feature selection, and thorough model optimization. The 400 to 678 nm wavelength range was identified as the most informative for flavonol prediction, offering insights for future spectroscopic analyses of legumes.

The research has broad practical implications for the food industry, from plant breeding to quality control and nutritional labeling. Its rapid, non-destructive nature aligns with the demand for efficient and sustainable food analysis techniques. While focused on flavonols in chickpea, the methodology could potentially be adapted for other bioactive compounds and pulse crops, opening new avenues in food science research. The integration of NIR spectroscopy with machine learning exemplifies the trend of applying advanced data analytics in food science.

In conclusion, this study provides both a practical solution for rapid flavonol quantification in chickpea and contributes to the broader understanding of spectroscopic and machine learning applications in food analysis. This research aligns with broader efforts to characterize and utilize chickpea genetic resources, as highlighted by Upadhyaya *et al*. (2008) in their study on the genetic structure and diversity of the chickpea composite collection. Our work contributes to this ongoing effort by providing a rapid method for assessing flavonol content, an important nutritional trait.

As the industry moves towards data-driven approaches, such methods will be crucial in ensuring food quality and safety. Future research could expand this approach to other compounds and crops, and explore its integration with emerging technologies in precision agriculture and smart food manufacturing.

This study contributes to the ongoing efforts to characterize and utilize plant genetic resources in India, as

outlined by Pratap *et al*. (2015) in their review of crop wild relatives. The development of rapid screening methods, such as the one presented here, can significantly enhance the efficiency of germplasm evaluation and utilization in breeding programs.

Conflict of Interest

There is no conflict of interest

Limitations and Future work

While our study demonstrates the potential of NIR spectroscopy combined with machine learning for flavonol prediction in chickpea, several limitations should be addressed in future work. Future studies should aim to increase sample size and diversity, including samples from different growing regions, seasons, and cultivation practices. The impact of environmental factors on NIR spectra and flavonol content should be extensively studied. Research should explore the applicability of this method to other chickpea products beyond flour, such as whole seeds and processed foods. Direct comparison with HPLC on the same samples would provide stronger validation of the NIR approach. Future work could also investigate the prediction of individual flavonol compounds using NIR and machine learning. The transferability of calibrations between different NIR instruments should be examined to ensure wider applicability. Implementation and validation of this method in actual production settings is crucial. Advanced machine learning techniques, such as deep learning or ensemble methods, could be explored to potentially improve prediction accuracy. Investigating the simultaneous prediction of multiple nutritional components using a single NIR scan could enhance the efficiency of nutritional profiling. Lastly, evaluating the performance of portable NIR devices for field-based measurements could expand the method's applicability to on-farm quality assessments. Addressing these limitations will further enhance the utility and reliability of NIR spectroscopy combined with machine learning for rapid nutritional assessment of chickpea and potentially other legumes.

Acknowledgments

I would like to thank the ICAR-National Bureau of Plant Genetic Resources and the team of Project "Biochemical Evaluation of Field and Vegetable Crops Germplasm" for providing spectrum and reference data.

References

Bharadwaj C, R Srivastava, SK Chauhan, CT Satyavathi, J Kumar, A Faruqui and AK Choudhary (2011) Molecular diversity and phylogeny in geographical collection of chickpea (*Cicer arietinum* L.) accessions. *J. Genet.* 90(2): e94-e100. https://link.springer.com/article/10.1007/s12041-011-0114-6

Biancolillo A, P Firmani, R Bucci, AD Magrì and F Marini (2020) Determination of insect infestation on stored rice by near infrared spectroscopy and interval-PLS regression. *Microchem. J.* 152: 104273. https://doi.org/10.1016/j.microc.2018.10.049

Diapari M, A Sindhu, K Bett, A Deokar, TD Warkentin and B Tar'an (2014) Genetic diversity and association mapping of iron and zinc concentrations in chickpea (*Cicer arietinum* L.). *Genome* 57(8): 459-468. https://pubmed.ncbi.nlm.nih.gov/25434748/

Guo W, J Gu, D Liu and L Shang (2016) Peach variety identification using near-infrared spectroscopy. *Comput. Electron. Agric.* 123: 297-303. https://doi.org/10.1016/j.compag.2016.03.005

Gupta DS, D Thavarajah, P Knutson, P Thavarajah, RJ McGee, CJ Coyne and S Kumar (2013) Lentils (Lens culinaris L.), a rich source of folates. *J. Agric. Food Chem.* 61(32): 7794-7799. DOI: 10.1021/jf401891p. https://pubs.acs.org/doi/10.1021/jf401891p

Jukanti AK, PM Gaur, CLL Gowda and RN Chibbar (2012) Nutritional quality and health benefits of chickpea (*Cicer arietinum* L.): a review. *Br. J. Nutr.* 108(S1): S11-S26. https://pubmed.ncbi.nlm.nih.gov/22916806/

Magwaza LS, SIM Naidoo, SM Laurie, MD Laing and H Shimelis (2016) Development of NIRS models for rapid quantification of protein content in sweetpotato [Ipomoea batatas (L.) Lam.]. *LWT - Food Sci. and Technol.* 111: 108805. https://doi.org/10.1016/j.lwt.2016.04.032

Pratap A, SK Chaturvedi, R Tomar, N Rajan, N Malviya, M Thudi and A Fikre (2015) Marker-assisted introgression of resistance to fusarium wilt race 2 in Pusa 256, an elite cultivar of desi chickpea. *Mol. Genet. and Genom.* 290(2): 559-566. https://link.springer.com/article/10.1007/s00438-017-1343-z

Upadhyaya HD, SL Dwivedi, M Baum, RK Varshney, SM Udupa, CLL Gawda and S Singh (2008) Genetic structure, diversity, and allelic richness in composite collection and reference set in chickpea (*Cicer arietinum* L.). *BMC Plant Biol.* 8(1): 1-12. https://pubmed.ncbi.nlm.nih.gov/18922189/

Wakholi C, LM Kandpal, H Lee, H Bae, E Park, MS Kim, C Mo, WH Lee and BK Cho (2018) Rapid assessment of corn seed viability using short wave infrared line-scan hyperspectral imaging and chemometrics. *Sens. Actuators B Chem.* 255: 498-507. https://doi.org/10.1016/j.snb.2017.08.036

Xie LJ, AC Wang, HR Xu, XP Fu and YB Ying (2016) Applications of near-infrared systems for quality evaluation of fruits: A review. *Trends Food Sci. Tech.* 107: 116-129. DOI: 10.13031/trans.59.10655