Determination of Vitamin C Concentration Using NIR Spectroscopy and Chemometric Modeling

Anup Majumder, Timothy Andersen

November 2025

Abstract

Vitamin C (ascorbic acid) is an essential micronutrient that performs numerous physiological roles, such as providing antioxidant protection, regulating immune responses, and collagen synthesis. However, its natural instability and sensitivity to oxidative degradation make it hard to accurately quantify in food and biological samples. Some conventional methods like Iodometric titration and High-Performance Liquid Chromatography (HPLC) are common ways to find out how much of a substance is in a sample. These methods are accurate, but also destructive, take a long time, and need a lot of sample preparation. Recent improvements in Near-Infrared (NIR) spectroscopy and chemometric modeling have made it possible to quickly and accurately measure vitamin C levels without damaging the sample. Chemometrics allows for multivariate calibration and the interpretation of complex spectral data sets, enabling the accurate prediction of analyte concentrations without the use of chemical reagents. Some recent exploratory research demonstrates that regression-based models like Partial Least Squares Regression (PLSR), Multiple Linear Regression (MLR), and k-Nearest Neighbors (kNN), can achieve higher predictive result ($R^2 > 0.95$) when integrated with preprocessing techniques such as Multiplicative Scatter Correction (MSC), Standard Normal Variate (SNV) correction and Normalization Correction (NC). Both strategies work well to reduce spectral scattering and baseline drifts, which makes the model more stable and easier to reproduce. Additionally, supplementary research on nanoencapsulation support the influence of nanostructured carriers on vaccine stability and bioavailability. So, the combination of NIR spectroscopy and chemometric algorithms could be a useful way to find out how much vitamin C is in food and drugs without using any chemicals and in real time. This synthesis evaluates the current methodologies, conducts cross-comparisons of model accuracies, acknowledges dataset diversity and spectral interpretability constraints, and specifies the future research track aimed at the creation of universally applicable, field deployable spectro-chemometric models for vitamin C assessment.

1 Introduction

Vitamin C, also known as ascorbic acid, is an essential nutrient and powerful antioxidant naturally present in many foods such as fruits, vegetables, and juices [1]. Accurately measuring its concentration in different products is crucial for ensuring proper nutritional labeling, maintaining quality control, and supporting various areas of health research. Traditional laboratory techniques for quantifying vitamin C such as colorimetric assays, redox titration, high-performance liquid chromatography (HPLC), and ultraviolet (UV) spectrophotometry have been widely employed for decades. However, each method presents distinct analytical challenges. Classical titrimetric assays, while simple and inexpensive, often require relatively large sample volumes and are unsuitable for colored or turbid matrices because visual endpoint detection becomes unreliable [2]. In contrast, while HPLC provides outstanding sensitivity and specificity, it requires costly equipment, long analysis times, and complex sample preparation making it less practical for high-throughput testing or on-site use [3].

Similar challenges are also observed with other vitamins and bioactive compounds, as their instability and detection difficulties call for more innovative and reliable analytical approaches [4]. Infrared (IR) and Raman spectroscopy, in particular, enable non-destructive measurement of samples, producing spectral "fingerprints" that can be correlated with vitamin content [5] [6]. FTIR and NIR spectroscopy generate high-dimensional spectral data, which are effectively processed using multivariate calibration techniques such as PLS regression. PLS reduces data dimensionality and addresses collinearity, enabling robust prediction of chemical constituents, including vitamins, proteins, carbohydrates, fatty acids, carotenoids, and micronutrients in various foods [7] [8] [9] [10]. Partial least squares (PLS) regression and other multivariate algorithms can extract relevant information from broad spectral data, enabling accurate predictions of nutrient levels. Likewise, modern machine learning methods (support vector machines, k-nearest neighbors, etc.) have been introduced to enhance predictive performance from spectral measurements. Ensemble learning strategies, such as averaging or stacking multiple PLS models, further enhance predictive performance and automation [11].

Machine learning algorithms such as Support Vector Machines (SVM), Random Forests (RF), Artificial Neural Networks (ANN), and k-Nearest Neighbors (kNN) have been extensively utilized for the analysis of spectroscopic data as previously me. These approaches often demonstrate superior predictive performance compared to conventional Partial Least Squares (PLS) models, particularly when integrated with appropriate feature selection or dimensionality-reduction techniques [12] [13] [14]. These approaches yield high coefficients of determination ($R^2 > 0.90$) and low prediction errors for both macro- and micronutrient quantification in diverse food and plant matrices [8].

An overview of recent advancements in spectroscopic and chemometric methods for vitamin analysis is the objective of this synthesis work. It compiles and summarizes the findings from several key studies, each addressing different as-

pects of vitamin quantification particularly focusing on vitamin C, while also referencing other vitamins where relevant. The primary themes include: analytical method comparison (evaluating different spectroscopic techniques and algorithms), predictive modeling performance (accuracy and robustness of various computational models), and practical applications (implications for industry and research).

In order to achieve the objectives of this study, four seed papers on the qualitative and quantitative analysis of vitamins, particularly vitamin C were selected. These foundational papers served as the basis for further literature exploration to identify additional relevant studies. The four seed papers are mentioned in the table below.

Table 1: Seed Papers.

Paper Title(s)	Relevance	References
Rapid determination of vitamin C by NIR, MIR and FT-Raman techniques	Analysis of vitamin C using infrared and Raman spectroscopy	[15]
Rapid and non-destructive determination of vitamin C and antioxidant activity of intact red chilies using visible near-infrared spectroscopy and machine learning tools	Rapid and non-destructive analysis of vitamin C in foods	[16]
Toward Achieving Rapid Estimation of Vitamin C in Citrus Peels by NIR Spectra Coupled with a Linear Algorithm	Quantitative estimation of vitamin C using chemometric models	[17]
Co-encapsulation of multivitamins in micro & nano-sized starch: target release, capsule characterization and interaction studies	Stability enhancement of multivitamins using micro/nano encapsulation	[18]

The first study demonstrates a comparative analysis of multiple infrared and Raman spectroscopic techniques for rapid quantification of vitamin C. The second study employs visible and near-infrared (Vis–NIR) spectroscopy combined with machine learning algorithms to predict vitamin C concentration and antioxidant activity in intact chili peppers. The third study develops a near-infrared (NIR)-based linear modeling approach for estimating vitamin C levels in citrus peels, incorporating wavelength selection techniques to simplify and optimize the model. The fourth study takes a different perspective by focusing on the co-encapsulation of multiple vitamins (D, E, B_1, and B_2) in micro- and nano-sized starch particles to enhance stability, using spectroscopic characterization methods to verify encapsulation efficiency.

By synthesizing these studies, we can appreciate how computer-aided analytical methods are advancing vitamin C measurement and how these approaches extend to broader vitamins management. The following sections detail the methodologies, key results, comparative insights, future directions, and over-

all conclusions drawn from these works

2   Theories and Background

An overview of key background information will be presented in this section, covering essential topics such as spectroscopic principles, chemometric modeling techniques, wavelength selection methods, and advanced machine learning approaches applied to vitamin analysis. The purpose of this discussion is to provide the foundational understanding necessary to interpret how the selected studies, each employing different spectroscopic and analytical strategies have been synthesized to advance the field of rapid, non-destructive vitamin quantification.

2.1   Spectroscopic Fundamentals

Vibrational spectroscopy including near-infrared (NIR), mid-infrared (MIR), and FT-Raman relies on the interaction of electromagnetic radiation with molecular bonds, producing characteristic absorption or scattering patterns. These techniques are non-destructive, rapid, and require minimal sample preparation, making them ideal for food and pharmaceutical analysis. NIR and MIR are particularly effective for quantifying vitamin C and other bioactives due to their sensitivity to functional group vibrations [15] [17] [19] [20].

Spectroscopic techniques exploit the interaction between electromagnetic radiation and molecular vibrations to extract chemical and structural information about a substance. In the context of vitamin C analysis, infrared (IR) and Raman spectroscopy provide distinct yet complementary approaches for identifying and quantifying molecular species through their characteristic vibrational modes.

2.1.1   Near-Infrared (NIR) Spectroscopy

Near-infrared (NIR) spectroscopy has emerged as a rapid, non-destructive, and reagent-free technique for quantifying vitamin C (ascorbic acid) in foods and pharmaceuticals. By measuring the absorbance of NIR light (typically 700–2500 nm) and applying chemometric models, NIR enables efficient estimation of vitamin C content across diverse sample types.

Near-Infrared (NIR) spectroscopy emerged as a central analytical approach across all these four studies, demonstrating its versatility for both quantitative and structural applications in vitamin analysis. In vitamin C quantification, it first highlighted the power of NIR and FT-NIR spectroscopy as rapid, non-destructive techniques capable of producing highly accurate calibration models using Partial Least Squares (PLS) regression [15], achieving correlation coefficients above 0.98 for powder and solution samples. Building upon this foundation, it has been extended the use of visible–NIR (Vis–NIR) reflectance spectroscopy (381–1065 nm) to intact red chilies [16], showing that spectral informa-

tion combined with machine learning algorithms—particularly k-nearest neighbors (kNN)—could predict vitamin C and antioxidant activity with near-perfect accuracy ($R^2 \approx 0.999$).

Similarly, in the research work of [17] applied NIR spectroscopy (912–1667 nm) to citrus peels and refined the modeling process through normalization correction and wavelength-selection algorithms, enabling a simplified Multiple Linear Regression (MLR) model using only 15 key wavelengths without compromising predictive performance. Even in the encapsulation study by [18], while not focused on quantitative calibration, NIR-related analytical principles underpinned techniques such as ATR–FTIR and UV–Vis spectroscopy for monitoring encapsulated vitamin interactions and confirming molecular integrity. Collectively, these works establish NIR spectroscopy as a powerful, rapid, and eco-friendly tool for vitamin C assessment and related nutritional analyses—bridging analytical chemistry with modern data-driven modeling to enable real-time, non-destructive evaluation of food and pharmaceutical products.

Overall, NIR spectroscopy detects overtone and combination vibrations of molecular bonds most notably O–H, C–H, and N–H—that are abundant in vitamin C and its surrounding matrix [15]. These vibrations originate from higher-energy harmonics of fundamental infrared absorptions, producing broad and overlapping spectral features. Because of this spectral complexity, univariate approaches are insufficient; hence, multivariate calibration techniques such as Partial Least Squares (PLS) regression are widely employed to extract meaningful quantitative information from NIR data. Preprocessing techniques, including multiplicative scatter correction (MSC), standard normal variate (SNV), and first- or second-derivative transformations, are commonly applied to correct for baseline drift, light scattering, and particle-size effects. Together, these methods enhance signal quality, improve calibration robustness, and enable accurate prediction of vitamin C concentration in diverse sample types ranging from powders and liquids to intact biological matrices [21] [22] [23] [24].

2.1.2 FT-NIR (Fourier Transform Near-Infrared)

Fourier Transform Infrared (FTIR) spectroscopy, particularly when combined with chemometric modeling, has become a powerful tool for the quantification of vitamin C (ascorbic acid) in both pharmaceutical and food matrices. Across this four analyzed studies, FTIR was used in various modes like ATR (Attenuated Total Reflectance), DRIFTS (Diffuse Reflectance Infrared Fourier Transform Spectroscopy), and PAS (Photoacoustic Spectroscopy) to achieve accurate, rapid, and non-destructive measurements of vitamin concentrations and molecular interactions. Because FTIR spectra are complex, chemometric models, especially Partial Least Squares (PLS) regression are used to link spectral data with vitamin C concentration [15] . These models are calibrated and validated to ensure accuracy and reliability, while using the area under specific absorbance bands (like C=C or C=O stretching) helps improve quantification precision [25] [26].

It has established of FTIR's quantitative reliability in predicting vitamin

C concentration with near-perfect accuracy, especially using the ATR configuration [15]. In other side, it used FTIR primarily for functional verification, ensuring that the NIR/Vis-NIR models were chemically meaningful [16] [17]. Then, it expanded FTIR's scope to nano-encapsulation characterization, where peak shifts directly reflected hydrogen bonding and encapsulation efficiency [18]. FTIR techniques show outstanding performance in vitamin C analysis, with correlation coefficients up to 0.999, minimal prediction errors, and excellent reproducibility. Recovery rates often exceed 100%, and relative standard deviations remain below 4.6%, confirming the method's accuracy, consistency, and suitability for quality control of both raw materials and finished products [15] [25]. In my word it can be state that, FTIR proved to be a versatile, non-destructive, and rapid analytical approach that bridges classical chemistry and modern chemometrics. It not only quantifies vitamin C accurately but also provides mechanistic insights into molecular interactions, structural stability, and matrix integration, reinforcing its vital role in both nutritional analytics and materials science.

2.1.3   Vis–NIR Spectroscopy (Visible–Near Infrared Spectroscopy)

Visible and near-infrared (Vis–NIR) spectroscopy is a rapid, non-destructive analytical technique that measures light absorption by a sample to determine its chemical composition [15]. In this spectral range (381–1065 nm), absorption features arise primarily from electronic transitions of color pigments and overtone or combination bands of molecular vibrations such as O–H, C–H, and N–H [16]. Because these absorptions are directly related to the concentration of various biochemical constituents, Vis–NIR spectroscopy has become widely employed in the food industry for quantifying nutrients, moisture, and antioxidants without requiring sample destruction or chemical reagents [15] [16].

In [16] it has been expanded the application of visible–near-infrared (Vis–NIR) spectroscopy for non-destructive vitamin C and antioxidant analysis in intact red chilies. Spectral data were collected using a NirVana AG410 spectrometer in the 381–1065 nm wavelength range. Each chili was scanned at three distinct positions (north, center, and south) to account for natural heterogeneity, and the spectra were recorded in diffuse reflectance mode and converted to absorbance values for analysis. This experiment proved that Vis–NIR spectroscopy can serve as a reliable, low-cost tool for rapid, non-invasive screening of vitamin C and antioxidant levels in agricultural products, eliminating the need for destructive assays.

In general, Vis–NIR spectroscopy is used for the rapid assessment of freshness, adulteration, and quality parameters (e.g., soluble solids, acidity, firmness) in products such as meat, tomatoes, fruits, and coconut powder. Models often achieve high accuracy ($R^2 > 0.85$) and enable discrimination between varieties or detection of contaminants [27] [28] [29].

2.1.4   ATR-FTIR Spectroscopy

Attenuated Total Reflectance Fourier Transform Infrared (ATR-FTIR) spectroscopy has emerged as a key vibrational technique for molecular fingerprinting, functional group identification, and interaction analysis. By recording mid-infrared absorbance directly from the sample surface without extensive preparation, ATR-FTIR enables rapid, non-destructive evaluation of chemical bonds and compositional changes [30]. Here, in my studies ATR-FTIR played two major roles like quantitative correlation between infrared absorbance and vitamin C concentration and other one is qualitative structural characterization of encapsulated vitamins and carrier matrices.

In [15], ATR–FTIR along with NIR and Raman spectroscopy was used to quantify vitamin C in powdered starch mixtures and aqueous solutions. Using a ZnSe ATR crystal over the range 4000–650 cm^{-1}, spectra were recorded at a resolution of 16 cm^{-1} with 256 scans. Key absorption bands for C=O stretching (around 1750 cm^{-1}) and O–H bending (around 3400 cm^{-1}) showed strong linear correlation with ascorbic acid concentration. In the micro/nano-encapsulation study [18], ATR–FTIR was integral to characterizing the interaction between starch carriers and encapsulated vitamins (B_1, B_2, D, and E). Measurements were taken over the range 4000–650 cm^{-1} using an Agilent Cary 630 FTIR spectrometer. Here, ATR-FTIR in this context provided molecular-level insight into binding mechanisms that underpinned the physical and chemical stability of co-encapsulated multivitamins.

In my observation, ATR-FTIR consistently served as a bridge between spectral measurement and molecular understanding. By integrating ATR-FTIR with chemometric modeling, thermal profiling, and microscopy, its analysis collectively demonstrated how infrared spectroscopy can unify composition analysis, model prediction, and structural validation forming a holistic framework for vitamin detection, stability assessment, and formulation design. Its adaptability across solid, liquid, and encapsulated systems positions it as a cornerstone tool for next-generation spectro-chemometric research on vitamins and bioactive compounds.

2.1.5   FT-Raman spectroscopy

Fourier Transform Raman (FT-Raman) spectroscopy is a vibrational spectroscopic technique that complements infrared (IR) analysis by probing molecular vibrations through inelastic light scattering rather than absorption. Because Raman scattering involves the change in molecular polarizability, FT-Raman provides structural and compositional information about non-polar bonds and symmetric functional groups that may be weak or invisible in FTIR spectra [31]. FT-Raman spectroscopy utilizes near-infrared (NIR) laser excitation most commonly at 1064 nm to induce Raman scattering, which is subsequently processed using Fourier transform algorithms. This configuration minimizes fluorescence interference, a major limitation in conventional Raman spectroscopy, making FT-Raman particularly suitable for analyzing biological and food sam-

ples [15] [32] [33]. The technique requires minimal sample preparation and allows direct measurement of samples in their native state, including through packaging or in aqueous environments, due to its weak interaction with water [32] [33].

In [15], FT–Raman spectroscopy was integrated alongside NIR, FT–NIR, and FTIR to assess vitamin C concentrations in both solid starch-based mixtures and aqueous solutions. Here, FT–Raman provided a robust alternative for non-destructive vitamin C quantification where infrared absorption was limited or confounded by moisture interference. In [18], the study of micro/nano-encapsulation employed FT–Raman spectroscopy as a confirmatory structural tool to elucidate the intermolecular interactions between starch matrices and encapsulated vitamins (B_1, B_2, D, and E). Here, Combining with ATR-FTIR, XRD, and DSC, the FT-Raman data established a multi-technique verification of encapsulation efficiency, molecular bonding, and stability enhancement in the nano-capsules. Raman spectra thus played a vital role in confirming chemical integrity and interaction dynamics at the molecular level.

FT-Raman spectroscopy excels in resolving overlapping infrared bands, detecting non-polar molecular vibrations, and providing sharper spectral resolution without interference from water absorption. When integrated with chemometric models such as Partial Least Squares Regression (PLSR), Multiple Linear Regression (MLR), or k-Nearest Neighbors (kNN), Raman spectral data can achieve quantitative performance comparable to that of NIR spectroscopy [34]. The technique's sensitivity to molecular symmetry, hydrogen bonding, and conformational changes makes it especially valuable in studies of encapsulation, stability, and material interactions. Collectively, these works demonstrate that FT-Raman, when synergized with NIR and FTIR, forms a robust triad for multi-modal, non-destructive vitamin analysis—integrating quantitative estimation, structural verification, and computational modeling into a unified spectro-analytical framework. Its non-destructive nature, resistance to water interference, and ability to probe non-polar functional groups establish FT-Raman as a powerful complementary tool to ATR-FTIR and NIR spectroscopy in modern chemometric and material characterization workflows [15].

3   Methodology Analysis

The methodology and analysis section aims to provide an in-depth evaluation of the research methods, spectral data acquisition techniques, chemometric modeling approaches, and analytical best practices presented in the selected seed papers. It critically examines the similarities, differences, and limitations of the methodologies employed for vitamin quantification, highlighting how each study integrates spectroscopic and computational tools to improve analytical accuracy. Furthermore, this section discusses the potential, challenges, and future directions of using advanced spectroscopic and machine learning techniques for rapid, non-destructive vitamin analysis across diverse food and pharmaceutical matrices. Also this synthesis provides an overview of the common chemometric models, sample matrices analyzed, spectral preprocessing methods, model

training and validation strategies, performance metrics, and a comparison of the advantages and limitations relative to conventional assays.

These four studies demonstrate that chemometric modeling transforms spectroscopy into a predictive, high-throughput analytical platform. Together, these methodologies establish a multi-modal chemometric framework combining physical spectroscopy, statistical modeling, and computational intelligence to quantify, validate, and interpret vitamin C with precision, speed, and minimal sample preparation. Unlike traditional wet-chemical methods (e.g., HPLC or titration), Near-Infrared (NIR) spectroscopy requires minimal sample preparation and no chemical reagents, allowing for rapid and high-throughput analysis [20] [35]. In NIR analysis, broad spectral features corresponding to O–H, C–H, and N–H bond overtones and combination bands are correlated with vitamin C concentration using multivariate calibration models [36]. Recent studies, particularly within the past five years, have employed advanced chemometric techniques such as Partial Least Squares (PLS) regression, Support Vector Machines (SVM), and Artificial Neural Networks (ANN) to enhance the accuracy and robustness of vitamin C quantification in both food and pharmaceutical matrices.

In [15] the methodology for this study we can see in fig. 1 involved preparing both powdered and aqueous vitamin C samples with varying concentrations. These samples were then measured using six different hesspectroscopic techniques: NIR, FT-NIR, FTIR-ATR, DRIFTS, FTIR-PAS, and FT-Raman. To reduce noise and normalize the spectral data, Multiplicative Scatter Correction (MSC) was applied during data preprocessing. The dataset was split into 75% calibration and 25% validation sets. A PLS regression model was developed to relate the spectral data to vitamin C concentration, with optimal factors determined through PRESS minimization. Model evaluation involved calculating R^2, SECV, and SEE to assess accuracy for both calibration and validation datasets. The performance of all methods was compared, with FTIR-ATR achieving the highest R^2.

In [16] the methodology for this study we can see in fig. 2 involved the collection of 18 chili varieties, prepared under controlled conditions to ensure sample consistency. Vis–NIR spectroscopy (381–1065 nm) was employed to capture chemical information from the samples by measuring light absorption. To obtain actual vitamin C and antioxidant activity data, iodometric titration and DPPH assays were conducted as reference measurements. Spectral data were then preprocessed using Standard Normal Variate (SNV) to normalize the spectra and reduce noise, enhancing the model's robustness. The next step involved developing prediction models using PLS regression (PLSR), Support Vector Machine Regression (SVMR), and k-Nearest Neighbors (kNN) to correlate the spectral data with the reference measurements. Model evaluation was performed by comparing calibration and prediction accuracy through R^2, RMSE, and RPD. Here, Vis–NIR spectroscopy combined with machine learning provides a fast, non-destructive, and accurate method for predicting vitamin C and antioxidant content, making it highly suitable for application in the food industry.

In [17], they developed a methodology we can see in fig. 3 that, 249 citrus peel samples from 50 different varieties were prepared and cleaned to ensure
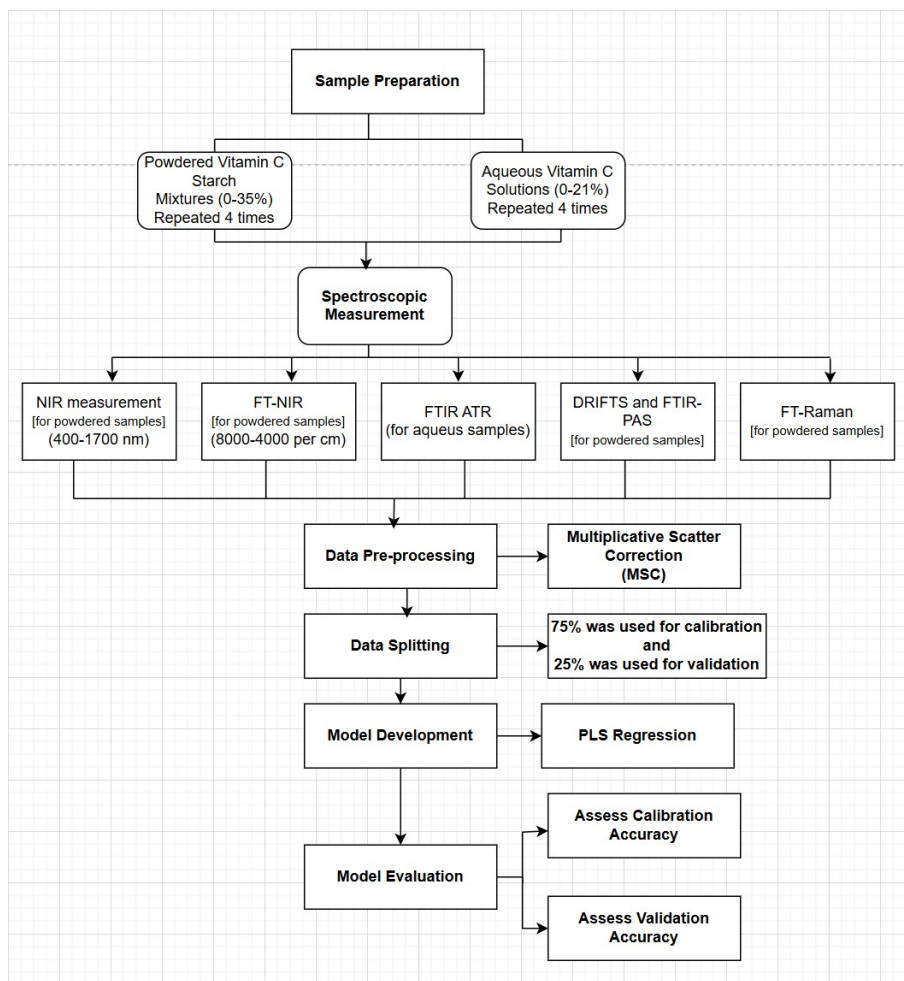
9

Figure 1: Methodology for Rapid determination of vitamin C by NIR, MIR and FT-Raman techniques

both variety and sample integrity. Near-infrared (NIR) spectroscopy was then employed to collect spectral data within the 912–1667 nm range using an Isuzu Optics spectrometer, capturing the chemical fingerprints of the samples. Various spectral preprocessing techniques including Savitzky–Golay smoothing, normalization correction (NC), multiple scattering correction (MSC), and derivative transformations were evaluated, with NC identified as the most effective for reducing noise and light scattering.

In this methodology, Partial Least Squares (PLS) regression modeling was performed using full-spectrum NC-preprocessed data, achieving a strong correlation. To refine model efficiency, four wavelength selection algorithms namely
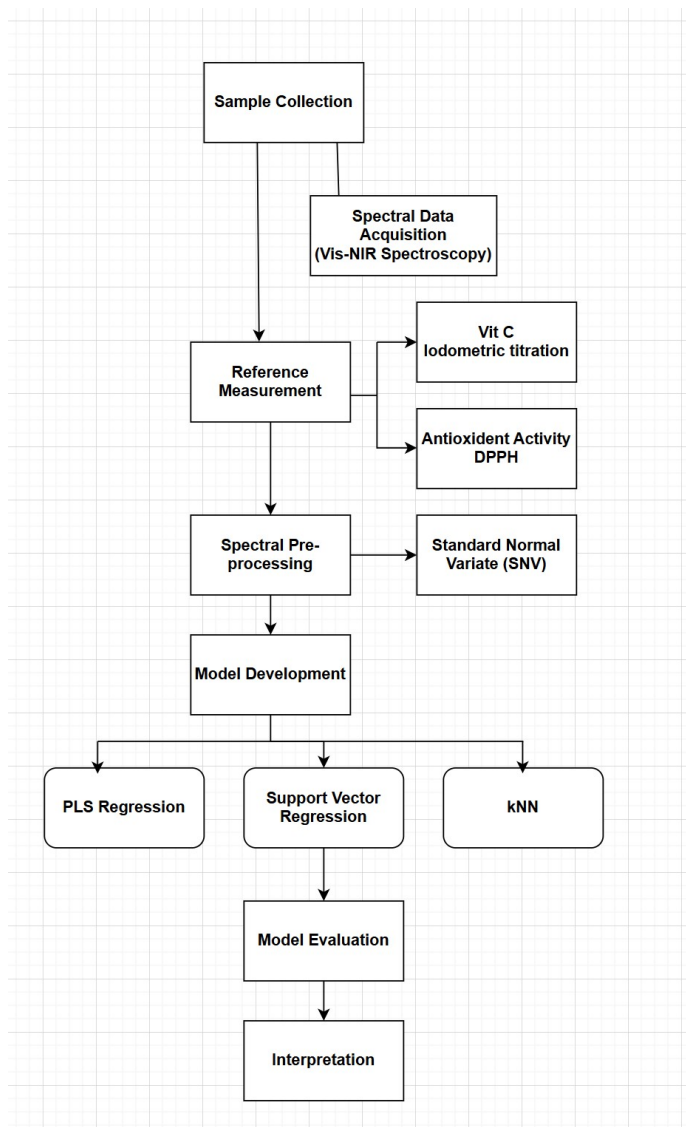
Figure 2: Methodology for determination of vitamin C and antioxidant activity of intact red chilies using visible near-infrared spectroscopy

Regression Coefficients (RC), Stepwise Regression (SR), Successive Projections Algorithm (SPA), and Coherent Anti-Stokes Raman Scattering (CARS) were tested, with SR identified as optimal, selecting 15 key wavelengths. A simplified multiple linear regression (SR-NC-MLR) model was then developed, demonstrating rapid and accurate vitamin C prediction. Statistical validation using F-test and t-test confirmed no significant difference between predicted and ref-
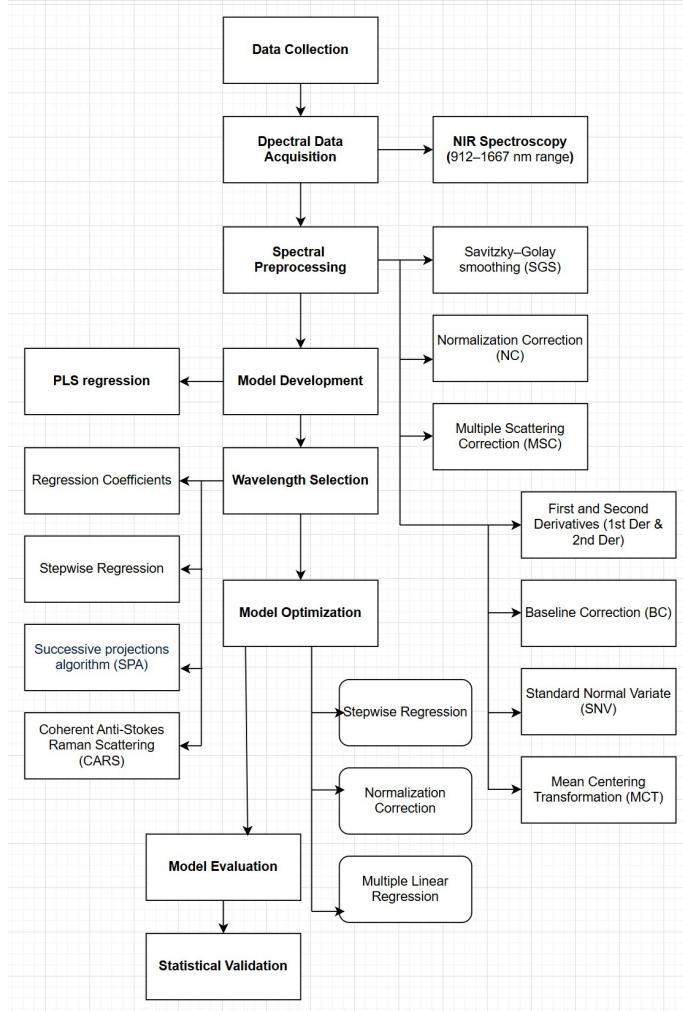
Figure 3: Methodology for Rapid Estimation of Vitamin C in Citrus Peels by NIR Spectra

erence values, while external validation verified the model's robustness. Overall, the integrated NIR–SR–NC–MLR approach provided a fast, reliable, and non-destructive method for predicting vitamin C content in citrus peels, offering strong potential for industrial applications in citrus waste valorization.

In [18] the study began with the extraction of starch from water chestnuts, followed by size reduction using ball milling to prepare the starch for encapsulation as we can see in fig. 4. The vitamins (D, E, B1, B2) were then dissolved and added to the starch solution, which was subjected to sonication and freeze-
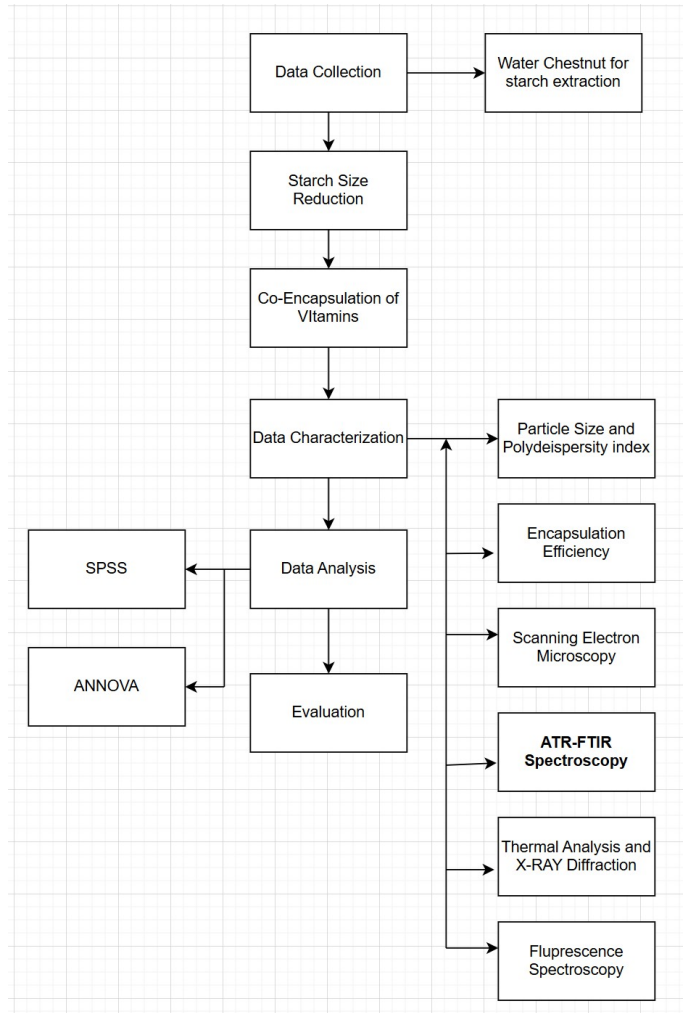
Figure 4: Methodology for Co-encapsulation of multivitamins in micro and nano-sized starch

drying for co-encapsulation. This process ensured the vitamins were incorporated into the micro and nano-sized starch particles. Capsule characterization included measuring particle size, polydispersity, and zeta potential using a ZetaSizer, assessing the physical properties of the encapsulated particles [37]. Encapsulation efficiency was determined by washing, centrifuging, and measuring the vitamin content released through spectrophotometry, providing insight into how effectively each vitamin was encapsulated. Structural analysis using ATR-FTIR, Differential Scanning Calorimetry (DSC), and X-ray Diffraction (XRD) helped study the molecular interactions and thermal properties of the capsules.

Microscopic analysis via Scanning Electron Microscopy (SEM) provided a visual inspection of the capsule morphology, revealing the structural features of the micro capsules. Fluorescence Spectroscopy was employed to measure the fluorescence intensity of the encapsulated vitamins, further investigating the interactions and stability of the encapsulated vitamins. Statistical validation of the results was carried out using ANOVA in SPSS, confirming the experimental findings. The study concluded that co-encapsulation significantly improves the stability and bioavailability of the vitamins, supporting its feasibility for use in food fortification applications.

After synthesizing all this four methodologies we can combine the methodologies for quantification of Vitamin C Concentration Using NIR Spectroscopy and Chemometric Modeling as per the following strategies:

3.1   Sample Collection

Several researchers works with different types of data and analyzed vitamin C content using a diverse range of sample matrices. A powdered mixtures of vitamin C and starch were prepared in varying concentrations (0–35%) to model quantitative analysis in [15]. In [16] they focused on 18 varieties of intact red chilies cultivated under controlled greenhouse conditions in Aceh Province, Indonesia, selected for their natural vitamin C and antioxidant content. The study of [17] utilized citrus peels from 50 different citrus varieties to explore spectral features related to ascorbic acid content. Lastly, the study of [18] investigated water chestnut (Trapa natans) starch as an encapsulation matrix for multivitamins, including vitamins B1, B2, D, and E, to evaluate retention and controlled release behavior through micro and nano-encapsulation techniques.

3.2   Sample Preparation

Across the four studies, distinct sample preparation protocols were employed to accommodate different sample types and analytical objectives. In [15] vitamin C was either blended with starch in powdered form or dissolved in aqueous solutions, producing mixtures with concentrations ranging from 0 to 21% for spectroscopic quantification. Sometimes data collected from fusion methods that could make models to interpret data more accurately by combining related outputs from different instruments [38]. In [16] study it utilized intact red chili samples, where no grinding or extraction was performed; instead, Vis–NIR spectral data were directly collected from the whole fruits to preserve their natural optical characteristics. In [17] study, citrus peels were carefully washed, cleaned of impurities, and cut into standardized 20 mm × 20 mm segments to ensure uniformity during spectral scanning and data acquisition. The [18] study involved a more complex procedure and water chestnut (*Trapa natans*) starch was first extracted and processed, followed by the co-encapsulation of vitamins B_1, B_2, D, and E within the starch matrix to form micro- and nano-capsules for subsequent structural and spectral analysis.

3.3   Spectral Preprocessing and Data Enhancement

Although accurate data collection is essential, the most important step before analyzing chemical data is spectrum preprocessing or cleaning [39]. Data preparation is a necessary step in chemometrics that is done to improve the quality of the data and the signal-to-noise ratio. This is done to make it easier to get useful information from the data, which makes it useful for making and training calibration or classification models. Sometimes even if the full spectrum captured it shows peak related to the samples but that don't give the useful information about how much to predict accurately [40]. The following preprocessing techniques are mostly used to preprocess and enhance the collected data:

Standard Normal Variate (SNV): is a common way to prepare spectral data for chemometrics, especially for near-infrared (NIR) and other types of spectroscopic data. The main goal is to fix scatter effects and baseline shifts, which makes chemometric models easier to understand and more useful for making predictions. But its effect can change based on the application and the data's features. SNV transforms each spectrum by centering and scaling, removing additive and multiplicative effects caused by sample heterogeneity, particle size, or instrument variation [41] [42]. Initially it takes the mean of the each spectrum and then divide each mean centered spectrum by standard deviation which is mathematically presented as follows:

$$\frac{\mathbf{X}_{ij} - \bar{X}_i}{s_i} \ \text{ i is the ith spectrum} \tag{1}$$

In some cases, especially where scattering contains relevant sample information, SNV may inadvertently remove useful signals, reducing model performance. Careful evaluation is needed before applying SNV in such contexts [43] [44].

Multiplicative Scatter Correction (MSC): MSC is the most commonly used pre-processing technique which is closely folloed by SNV and it is uded for correcting baseline and scattering effects [39]. Using linear regression, MSC adjusts each spectrum by fitting it to a reference (usually the mean spectrum). This fixes both additive (offset) and multiplicative (scaling) effects. The corrected spectrum is then used to do more chemometric modeling [45] [46]. To implement MSC the first step is to find the slope "b" and intercept "a" of the equation given here. The parameters are the spectrum corrected and final coefficients.

$$X_{\text{spe}} = a + bX_{\text{ref}} \tag{2}$$

where X_{spe} and X_{ref} are specific and reference spectra respectively.

$$\mathbf{X}_{\text{final}} = \frac{\mathbf{X}_{\text{spe}} - \mathbf{a}}{\mathbf{b}} \tag{3}$$

Savitzky–Golay Correction (SGC): Savitzky–Golay filtering, also known as SGC, is a common pre-processing method in chemometrics that is used to smooth and separate spectroscopic data. SGC uses a moving window polynomial regression on the data to smooth out the signal or find its derivatives. This method works especially well for keeping sharp peaks and subtle features in spectra intact, while simple averaging filters can change these details [47] [48]. SGC can make noise louder when calculating higher-order derivatives and may add artifacts near data boundaries [48]. However, new research suggests that different filters or better SGC variants can fix these problems.

Mean-Centering Transformation (MCT): Mean-Centering Transformation (MCT) is a basic preprocessing step in chemometrics, used especially for multivariate data analysis like spectroscopy. In recent time researchers use MCT before using multivariate calibration methods like Principal Component Analysis (PCA), Partial Least Squares (PLS), and Principal Component Regression (PCR) [17] [49]. In some cases, for certain types of spectral data (e.g., ATR-FTIR), mean-centering alone may not significantly improves model performance in comparison to others pre-processing techniques.

Normalization Correction (NC): Standard Normal Variate (SNV), Variable Sorting for Normalization (VSN), Probabilistic Quotient Normalization (PQN), and normalization to total intensity are widely used. These methods correct for sample-to-sample differences in path length, concentration, or instrument response, and are often combined with other pre-processing steps like derivatives or smoothing [43] [50]. Normalization is crucial in NIR, Raman, and NMR spectroscopy as well.

1st/2nd Derivative Approach: In chemometrics, first (D1) and second (D2) derivative preprocessing are common methods for improving spectral data for both quantitative and qualitative analysis. D1 focuses on the rate of change in spectral intensity, which fixes baseline drift and makes peak resolution betterand it is especially good at telling apart overlapping bands and making classification better in samples that are hard to classify [41] [51]. D2 shows local curvature and makes it easier to see closely spaced or hidden peaks, but it can also make noise louder. Researchers often use D2 with smoothing (like Savitzky–Golay) to cut down on noise and show small details [43].

Across the four studies, various spectral preprocessing techniques were employed to improve data quality, minimize noise, and enhance the interpretability of spectral information prior to chemometric modeling. In [17] Savitzky–Golay smoothing (SGS) was applied to reduce random noise while preserving peak sharpness and spectral resolution. Both the [16] and [17] utilized the Standard Normal Variate (SNV) method to correct scattering effects and normalize the spectral data, thereby improving consistency among samples. Multiplicative Scatter Correction (MSC) was a common technique in those studies, effectively eliminating baseline drifts and scatter-induced distortions.

The following table summarize the use of some preprocessing techniques what actually used in quantification in vitamnin C for the synthesis:

Table 2: Spectral Preprocessing Techniques Used in Different Methodologies

References	Preprocessing Approach	Purpose
[15]	Multiplicative Scatter Correction (MSC)	Minimize scattering and instrument drift before PLS modeling
[16]	Standard Normal Variate (SNV)	Correct reflectance variations on curved chili surfaces
[17]	SGS, NC, MSC, 1st/2nd derivative, BC, SNV, MCT	Compare eight preprocessing schemes; NC gave best linear response
[18]	Minimal preprocessing; baseline subtraction	FTIR/FT-Raman inherently stable, focused on interpreting bond shifts

Additionally, the third study incorporated Normalization Correction (NC) to achieve more precise spectral calibration and scaling across datasets. Furthermore, the first and third studies applied first- and second-derivative transformations to enhance the resolution of overlapping absorption bands and reveal subtle spectral features, thereby improving the robustness of subsequent chemometric analyses.

3.4   Chemometric Modeling and Prediction

Across the four studies, chemometric modeling served as the central analytical approach to establish quantitative relationships between spectral data and vitamin-related parameters. Partial Least Squares Regression (PLSR) was the primary method employed in all studies to correlate spectral reflectance with vitamin concentration or antioxidant activity, providing robust calibration and prediction models. In some cases Simplified PLS models (sometimes even 9–12 key wavelengths) performed nearly as well as full-range models [52]. PLS coefficients were used to generate chemical maps also. In [16], Support Vector Machine Regression (SVMR) was introduced as an alternative to PLSR to capture nonlinear relationships within the Vis–NIR spectral data, demonstrating improved accuracy in complex datasets. In some cases The SVMR method is much better than Partial Least Squares (PLS) to evaluates the efficacy of Quantitative Structure-Activity Relationships (QSAR) analysis utilizing authentic chemical data [53]. SVMR is frequently combined with preprocessing (e.g., SNV, MSC) and variable selection to maximize predictive power [54].

It has been experimented with the k-Nearest Neighbors (kNN) algorithm as a simple, non-parametric predictive model to classify and estimate vitamin C levels based on spectral similarity [16]. kNN classifies or predicts a sample by finding the k closest data points (neighbors) in the feature space and us-

ing their majority class (for classification) or average value (for regression) [55]. Weighted kNN gives more weight to neighbors that are closer, while ensemble and kernel-based kNN methods make the model more stable and accurate, especially with datasets that are noisy or have a lot of dimensions [55]. Meanwhile, in [17], incorporated Multiple Linear Regression (MLR) following wavelength selection to simplify model structure while retaining predictive efficiency. MLR is used for making quantitative predictions (like finding out the concentrations of analytes from sensor or spectral data), modeling processes, and designing experiments. In chemometrics, it helps connect spectral or sensor data to chemical properties. It works best when there are fewer predictors than samples and multicollinearity is low [56]. Collectively, these chemometric approaches highlight the versatility of both linear and nonlinear regression strategies for transforming high-dimensional spectral data into reliable quantitative predictions of vitamin C content and related biochemical attributes.

3.5   Wavelength Selection and Optimization

In [17] study, various wavelength selection and optimization techniques were applied to refine the spectral data and enhance model performance by focusing on the most informative regions of the spectrum. Regression Coefficients (RC) were first utilized to identify wavelengths with the strongest correlation to vitamin C concentration, ensuring that the selected variables contributed significantly to model accuracy. It represents the change in the outcome variable for a one-unit change in the predictor, in the original units of measurement [57]. Stepwise Regression (SR) was then employed as an iterative statistical approach to progressively include or exclude wavelengths based on their predictive significance. Additionally, advanced algorithms such as the Successive Projections Algorithm (SPA) and Competitive Adaptive Reweighted Sampling (CARS) were implemented for dimensionality reduction, minimizing redundant spectral information and improving computational efficiency. By choosing the right variables, we can make a reliable model that is less expensive to run and can be used to get accurate results [40]. Some of the techniques for choosing variables that are used on spectral data are;

Interval-based PLS (iPLS): Interval-based PLS (iPLS) is a feature selection approach that builds Partial Least Squares (PLS) models on evenly spaced intervals across the spectrum's wavelengths. The prediction performance of interval PLS models for the subintervals and the global (full-wavelength) model is compared using the same latent variable. These PLS models are created on spectral subintervals of comparable width. The main basis for the comparison is the cross-validation metric RMSECV (root mean squared error of cross-validation). To obtain a complete picture of the model, additional metrics such as slope, offset, and r^2 (squared correlation coefficient) should also be examined [58]. Prior to applying iPLS, outliers detected in samples or measurements, typically discovered through PLS inner relation plots should be removed. The primary purpose of iPLS is to provide a thorough overview of model performance across different wavelength divisions, enabling the identification of informative spec-

tral regions and the elimination of those that are not. The resulting iPLS plots clearly illustrate how sensitive the PLS algorithm is to noise.

Backward Interval Partial Least Squares (biPLS): Backward interval partial least squares (biPLS) is a sophisticated variable selection method employed in chemometrics to improve the efficacy of PLS models, particularly for spectroscopic data. By systematically getting rid of spectral intervals that don't give much information, biPLS makes models more accurate, stable, and easy to understand. biPLS iteratively removes the least informative spectral intervals based on RMSECV until the optimal subset yielding the lowest prediction error is identified [59]. An alternative to biPLS is FiPLS which is sequentially adds the most informative spectral intervals based on RMSECV improvement until no further performance gain is achieved [58].

Moving Window Partial Least Squares (mwPLS): It is a method for choosing variables that is commonly used in chemometrics, especially when looking at spectroscopic data. The spectrum is split into windows that can be either fixed or variable in size. A PLS model is created for each window, and its ability to make predictions (like the error or residual sum of squares) is tested [60]. Windows with the least amount of prediction errors are said to have informative variables [61] [62]. The process helps to eliminate noise and areas that aren't important, so the model can focus on spectral intervals that meaningful. Selected intervals can be combined to build a final PLS model, or separate models from each interval can be aggregated for improved performance.

In NIR chemometrics, wavelength selection is used to find the spectral regions that are most closely related to the target property and get rid of extra or noisy variables. People often use iPLS, CARS, SPA, GA-PLS, and VIP to find a good balance between model accuracy, interpretability, and speed of computation. In my opinion, although not explicitly stated across all studies, Principal Component Analysis (PCA) could also be integrated into the workflow to capture major spectral variance and reduce data complexity before regression modeling. Collectively, it can be stated that, these wavelength optimization techniques allowed the chemometric models to achieve greater precision, robustness, and interpretability in vitamin C prediction from complex spectral datasets.

3.6   Model Optimization and Validation

All four studies emphasized rigorous model validation to ensure the reliability and generalizability of the chemometric predictions. Cross-validation was consistently applied, with datasets typically divided into calibration (approximately 75%) and validation (approximately 25%) subsets to assess model performance on unseen data and prevent overfitting. In the second and third studies, statistical significance tests such as the F-test and t-test were employed to confirm the reliability of the regression coefficients and to evaluate whether the predictive

results were statistically meaningful. Several key evaluation metrics were used across the studies: the coefficient of determination (R^2) quantified the strength of correlation between predicted and reference vitamin values; the Root Mean Squared Error (RMSE) for both calibration and validation phases measured the magnitude of prediction errors; and the Residual Predictive Deviation (RPD) assessed model robustness by comparing predictive accuracy relative to the variability in the reference dataset. Together, these validation strategies ensured that the chemometric models were not only accurate but also stable and transferable for predicting vitamin C content and related nutritional parameters from spectral data.

3.7 Statistical Analysis and Interpretation

Statistical evaluation played a vital role in interpreting and validating the findings across the four studies. In [18], SPSS software was utilized to perform statistical analyses, including Analysis of Variance (ANOVA), to assess the significance of differences observed in encapsulation efficiency, particle characteristics, and vitamin retention across formulations. Following model validation, the predictive outcomes from the chemometric analyses were interpreted in terms of their accuracy and reliability for quantifying individual vitamins or assessing antioxidant activity within the tested samples. The statistical interpretation ensured that variations in predicted values were scientifically justified and that the developed models exhibited strong predictive power and reproducibility, thereby confirming the robustness of the overall analytical approach.

3.8 Evaluation and Applications

The collective findings from the four studies demonstrate the broad applicability of spectroscopic and chemometric techniques for vitamin analysis and formulation development. The study [15] showcased the potential of NIR and FTIR spectroscopy for rapid and accurate quantification of vitamin C in both pharmaceutical and food matrices, offering a non-destructive alternative to traditional wet-chemical assays. The study [16] expanded this approach by integrating Vis–NIR spectroscopy with machine learning algorithms, enabling efficient, non-invasive assessment of vitamin C and antioxidant levels in fresh produce such as red chilies. The study [17] emphasized wavelength optimization strategies to enhance predictive precision, establishing an efficient model for rapid vitamin C estimation in diverse citrus varieties. The study [18] extended the application of spectroscopic and statistical tools beyond quantification which demonstrating successful co-encapsulation of multiple vitamins (B_1, B_2, D, and E) within water chestnut starch matrices, resulting in thermally stable and bioavailable multivitamin formulations suitable for food fortification and nutritional enhancement. Collectively, these studies underline the power of integrated spectroscopy in chemometrics frameworks in advancing rapid, reliable, and sustainable vitamin analysis and delivery systems.

Overall we can compare each methodology to others in the following way in table 3:

Table 3: Comparative and Synthetic Methodological Insights

Dimension	Reference [15]	Reference [16]	Reference [17]	Reference [18]
Spectroscopy	NIR, FT-NIR, FTIR-ATR, FTIR DRIFT, FT-Raman	Vis–NIR	NIR	ATR-FTIR, FT-Raman
Sample Type	Powder & Solution	Red chili	Citrus Peel	water chestnut followed by vitamin co-encapsulation
Preprocessing	MSC	SNV	SGS, NC, MSC, 1st/2nd derivative, BC, SNV, MCT	baseline subtraction
Modeling	PLSR	PLSR, kNN, SVMR	PLSR, MLR	Correlation Validation
Evaluation Metrices	R^2, Standard Error of Cross-Validation (SECV)	R^2, RMSE	R^2, RMSE	Structural Validation
Innovation	Multi-modal spectroscopy	Machine learning integration	Variable selection optimization	Structural chemometric fusion

4 Result Analysis and Comparative Discussions

The results section across the four studies demonstrates the successful application of Near-Infrared (NIR) spectroscopy coupled with chemometric models for the rapid and non-destructive determination of vitamin C content in various sample matrices such as citrus peels and other biological materials. The studies explored different spectroscopic techniques, pre-processing methods, and model optimization techniques to improve prediction accuracy and validation performance.

All three studies [15] [16] [17] focused on vitamin C quantification demonstrated that spectroscopic techniques, coupled with appropriate modeling, can achieve high accuracy and rapid results, often rivalling or exceeding traditional methods in practicality. In [15] multi-technique comparison, each spectroscopic method yielded strong predictive performance for vitamin C in the prepared samples. Using PLS regression on spectral data, they reported correlation coefficients (R^2) for predicted vs. actual vitamin C concentration near or above

0.97 for most methods: specifically, $R^2 \approx 0.999$ for FTIR-ATR (the best), 0.992 for FT-NIR, 0.988 for dispersive NIR, 0.976 for DRIFTS, 0.966 for FTIR-PAS, and 0.950 for FT-Raman. The corresponding prediction errors were low on the order of 0.2% to 3.0% depending on the method. Notably, FTIR-ATR (a mid-IR technique measuring liquid samples) achieved the lowest error (0.2%) and essentially a perfect fit ($R^2 \approx 0.999$), indicating superb accuracy for vitamin C in solution [15]. NIR-based methods also performed very well ($R^2 \sim 0.98$–0.99) despite the overlapping bands in the NIR region, thanks to the robustness of PLS in extracting the vitamin signal. FT-Raman showed a slightly lower R^2 (0.95) still high, but indicating a bit less sensitivity in this setup, possibly due to lower signal levels or fluorescence interference.

Table 4: Calibration and validation model evaluation of the NIR, MIR and FT-Raman methods. [15]

Method	Factor No.	Calibration			Validation	
		PRESS	SECV	R²	SEE	R²
FTIR-ATR	11	0.0002	0.23%	0.999	0.28%	0.998
DRIFTS	9	0.0148	1.82%	0.976	1.93%	0.975
FTIR-PAS	9	0.0183	1.95%	0.966	1.45%	0.985
NIR	5	0.0280	1.35%	0.992	1.64%	0.980
FT-NIR	3	0.0062	0.96%	0.992	1.64%	0.980
FT-Raman	5	0.0219	2.47%	0.950	2.68%	0.941

In the method of [16] achieved remarkable results in predicting vitamin C content in intact red chilies using Vis–NIR spectroscopy combined with machine learning. Among the tested models, the k-Nearest Neighbors (kNN, $k = 1$) model after SNV preprocessing performed best, showing an almost perfect correlation ($r \approx 0.999$) with minimal prediction error (RMSE ≈ 0.001) and an exceptionally high RPD = 31.62, indicating outstanding predictive accuracy. The PLS regression model also demonstrated strong performance ($R^2 \approx 0.989$, RMSEP ≈ 0.993, RPD ≈ 6.74), while SVM regression ($R^2 \approx 0.825$, RPD ≈ 1.77) underperformed, likely due to kernel or parameter limitations.

Table 5: Calibration and validation performances of various machine learning algorithms for prediction of vitamin C. [16]

Method	Hyperparameter	Calibration		Prediction		RPD
		r_{cal}	RMSEC	r_{val}	RMSEP	
PLSR	LV $= 22$	0.995	0.735	0.989	0.993	6.742
SVMR	C $= 100$, $g =$ scale, $d = 2$, $k =$ linear	0.930	2.816	0.825	3.787	1.768
kNN	n-neighbors $= 1$	0.999	0.001	0.999	0.001	31.622

For antioxidant activity prediction, kNN again produced near-perfect results ($r_{\mathrm{val}} \approx 1.0$). The study highlighted that SNV preprocessing was critical to remove scattering effects and enhance spectral signal quality. Spectral features

around 740 nm and 960–1000 nm were associated with pigments and water content, indirectly reflecting vitamin C and antioxidant levels. Overall, the findings confirm that Vis–NIR spectroscopy combined with SNV and kNN provides a rapid, non-destructive, and highly accurate approach for quantifying vitamin C and antioxidant activity in whole produce, though recalibration may be required for other crop types.

Table 6: Calibration and validation performances of various machine learning algorithms for prediction of antioxidants. [16]

Method	Hyperparameter	Calibration		Prediction		RPD
		r_{cal}	RMSEC	r_{val}	RMSEP	
PLSR	LV = 40	0.999	0.864	1.000	0.129	171.953
SVMR	C = 100, g = scale, d = 2, k = linear	0.964	5.917	0.967	5.627	3.914
kNN	n-neighbors = 1	0.999	0.855	1.000	0.000	316.228

Figure 5 shows scatter plots comparing predicted and measured vitamin C levels in red chilies using PLSR, SVMR, and kNN models. PLSR performed well ($r = 0.989$, RMSEP = 0.993, RPD = 6.742), while SVMR showed weaker prediction ($r = 0.843$, RMSEP = 3.787, RPD = 1.356). The kNN model achieved the best performance ($r = 0.9995$, RMSEP = 0.001, RPD = 31.622), indicating excellent predictive accuracy. Overall, kNN provided the most reliable and precise results, confirming its robustness for non-destructive vitamin C prediction in intact red chilies.
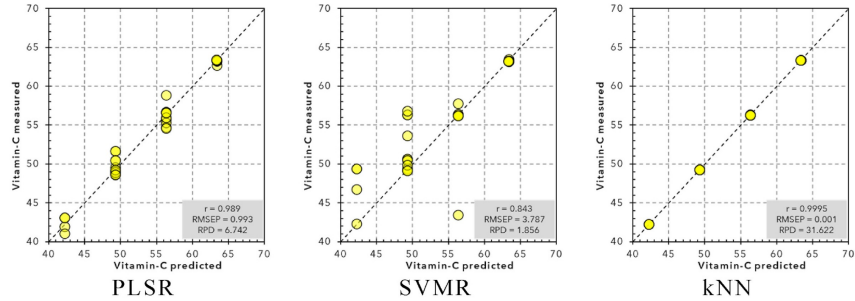

Figure 5: Performance validation to predict vitamin C on intact red chilies. [16]

In [17], a rapid and non-destructive approach was developed to estimate vitamin C content in citrus peels using NIR spectroscopy (912–1667 nm) coupled with linear algorithms. The full-spectrum NC–PLS model achieved strong predictive accuracy ($r_p = 0.956$, RMSEP = 13.8 mg/100 g), confirming that Normalization Correction (NC) was the best preprocessing method. To simplify the model, Stepwise Regression (SR) was used to select 15 optimal wavelengths, which were then applied in a Multiple Linear Regression (SR–NC–MLR) model. This reduced-variable model maintained comparable accuracy

($r_p = 0.949$, RMSEP $= 14.8\,\mathrm{mg}/100\,\mathrm{g}$) to the full-spectrum PLS while being faster and more practical. External validation on 40 independent peel samples yielded $R^2 = 0.9558$, confirming model robustness and generalizability. The study concluded that NIR spectroscopy combined with SR–NC–MLR offers a simple, rapid, and feasible method for estimating vitamin C in citrus peels, supporting industrial applications in citrus waste valorization and nutritional quality assessment.

Figure 6 shows the external validation results of the SR–NC–MLR model for predicting vitamin C content in citrus peels. The scatter plot compares predicted versus measured vitamin C values for 40 independent samples, displaying a strong linear correlation ($R^2 = 0.9558$). The regression line ($y = 0.922x + 17.055$) closely follows the 1:1 reference line, indicating high predictive accuracy and minimal deviation. These results confirm the model's robustness, reliability, and suitability for rapid, non-destructive estimation of vitamin C in citrus peels.
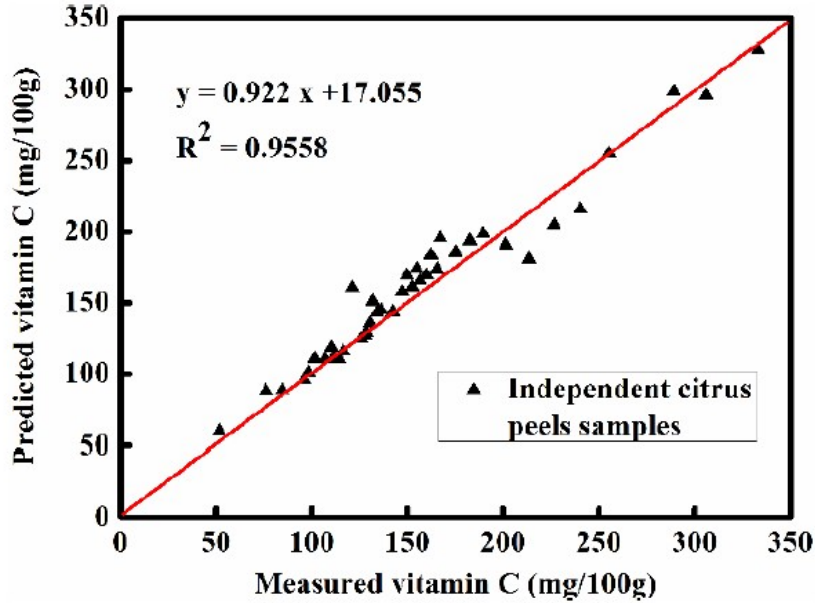

Figure 6: Predicted and measured values of vitamin C contents [17].

In [18], the study demonstrated that nano-sized starch particles significantly improved the co-encapsulation efficiency and stability of multivitamins (D, E, B_1, B_2) compared to micro-sized ones. Encapsulation efficiency increased from $35{\rightarrow}46\%$ for vitamin D, $81{\rightarrow}89\%$ for vitamin E, $83{\rightarrow}85\%$ for vitamin B_1, and $76{\rightarrow}78\%$ for vitamin B_2, confirming that nano-starch provides a superior barrier and higher surface interaction. FTIR spectra revealed a strong band at $877\,\mathrm{cm}^{-1}$, characteristic of vitamin ring structures, validating vita-

min incorporation. Fluorescence spectra showed higher emission intensity in nano-capsules, indicating enhanced molecular interactions and reduced quenching due to stronger vitamin–starch binding. Thermal (DSC) and XRD analyses confirmed better thermal stability and reduced crystallinity in nano-capsules, implying amorphous dispersion and improved protection against degradation. SEM images showed uniform, porous nanostructures. Overall, nano-starch carriers provided better vitamin retention, thermal resistance, and molecular integration—demonstrating their potential as stable, bioavailable carriers for food fortification.

The results from all four studies clearly demonstrate the potential of NIR spectroscopy combined with chemometric models (PLS, MLR, kNN) for rapid and accurate estimation of vitamin C content in a variety of food matrices, including citrus peels and chilies. The findings show that spectroscopy is a non-destructive, quick, and reliable alternative to traditional wet-chemistry methods. Furthermore, wavelength selection, model optimization, and preprocessing significantly enhance the predictive accuracy of the models, ensuring they are both practical and accurate for real-world applications such as food quality control and nutrient labeling. The nano-encapsulation study also highlights the potential of using nanotechnology to improve the stability and bioavailability of vitamins for use in food fortification, with the added advantage of rapid prediction models for vitamin content.

4.1 Comparative Analysis

Across all studies, the integrated findings highlight that FTIR-ATR and NIR spectroscopy, particularly using NC–PLS and SR–NC–MLR models, achieved outstanding predictive performance ($R^2 > 0.95$ with low RMSEP), demonstrating their effectiveness for rapid, non-destructive vitamin C quantification. Normalization Correction (NC) and Standard Normal Variate (SNV) preprocessing methods were most effective in eliminating scattering and baseline noise, thereby enhancing model robustness [41] [42]. Among modeling approaches, PLS and MLR provided strong and reliable linear correlations, while kNN outperformed others for spectra with distinct features, and SVM showed limited efficiency without kernel optimization. The external validation in [17] further confirmed the model's generalizability with $R^2 \approx 0.956$, supporting its applicability in real-world or field scenarios. Complementing these analytical advances, in [18] demonstrated that nano-starch encapsulation significantly improves vitamin stability and retention, providing a practical extension from quantification to nutrient preservation. Collectively, these findings establish a comprehensive, low-cost, and eco-friendly framework for vitamin C monitoring, stability assurance, and food fortification using spectroscopic and encapsulation-based innovations. In the table shows the overall comparison of all these methods with the results.

Table 7: Comparison of Spectroscopic and Chemometric Studies on Vitamin C and Multivitamin Analysis

Study	Sample/Matrix	Technique / Model Used	Spectral Range / Process	Best Model / Optimization	Performance Metrics	Key Findings
[15]	Vitamin C in aqueous & powdered samples	FTIR-ATR, FT-NIR, Dispersive NIR, DRIFTS, FTIR-PAS, FT-Raman	400–2500 nm (FTIR & NIR regions)	PLS Regression	• **FTIR-ATR:** $R^2 = 0.999$, SECV = 0.23% • **FT-NIR:** $R^2 = 0.988$ • **FT-Raman:** $R^2 = 0.950$	FTIR-ATR gave the highest accuracy and lowest error; NIR methods offered fast, non-destructive quantification (∼5 s/sample). Raman was less sensitive but moisture-independent.
[16]	Intact red chilies (vitamin C and antioxidants)	Visible–NIR spectroscopy + ML models (kNN, PLSR, SVM)	381–1065 nm	SNV-preprocessed kNN ($k = 1$)	• **kNN:** $r = 0.999$, RMSE = 0.001, RPD = 31.62 • **PLSR:** $R^2 = 0.989$, RMSEP = 0.993, RPD = 6.74 • **SVMR:** $R^2 = 0.825$, RPD = 1.77	SNV + kNN achieved near-perfect prediction; PLSR also strong; SVM ($R^2 \approx 0.82$) underperformed. Demonstrated feasibility of rapid, *in-situ* vitamin C assessment in produce.
[17]	Citrus peels (249 samples / 50 varieties)	NIR (912–1667 nm) + Linear Algorithms	Full (375 wavelengths) and 15 selected wavelengths	SR–NC–MLR (optimized from NC–PLS)	• **NC–PLS:** $r_p = 0.956$, RMSEP = 13.8 mg/100 g • **SR–NC–MLR:** $r_p = 0.949$, RMSEP = 14.8 mg/100 g • **External Validation:** $R^2 = 0.9558$	Normalization Correction (NC) improved spectra; Stepwise Regression (SR) reduced variables from 375 to 15 without accuracy loss. Validated rapid NIR quantification for citrus waste reuse.

[18]	Multivitamins (D, E, B_1, B_2) in micro & nano starch capsules	FTIR, Fluorescence, DSC, XRD Characterization	Encapsulation in water-chestnut starch	Nano-sized starch vs. Micro-starch	• **Encapsulation Efficiency (%):**
					  – Vit D: 35 → 46.3
					  – Vit E: 81.2 → 89.3
					  – Vit B_1: 83.1 → 84.9
					  – Vit B_2: 76.1 → 77.6
						Nano-starch improved encapsulation efficiency & thermal stability. FTIR showed vitamin bands (877 cm^{-1}), fluorescence confirmed strong vitamin–starch interactions. Enhanced vitamin retention and bioavailability.

4.2 Limitations and Challenges

In [15], the study utilized pure vitamin C–starch mixtures and aqueous solutions, without testing real food matrices, which limits its applicability to complex products where spectral interferences are common. The authors noted that moisture strongly absorbs in the MIR region (3700–3100 and 1640 cm^{-1}), overlapping with vitamin C peaks and thereby reducing FTIR reliability for high-moisture samples such as fresh foods or juices. The Partial Least Squares (PLS) models were developed under controlled laboratory conditions using a fixed calibration set; in my opinion, such models may not perform as accurately when applied to new sample types or different instruments without re-calibration. This work did not assess how other food components (fats, proteins, sugars, polyphenols) influence vitamin C prediction, and overlapping NIR absorption bands may cause cross-correlation in complex samples.

The reviewed studies on vitamin C determination using chemometric and spectroscopic methods like spanning FTIR, FT-NIR, Vis–NIR, and NIR-based models which share several common and study-specific limitations that affect their generalizability, accuracy, and field applicability. Most notably, sample diversity was limited, as analyses were conducted primarily on pure vitamin C powders, aqueous solutions, or samples from a single crop type or geographic region (e.g., red chilies from Aceh, citrus peels from China). This narrow scope restricts model transferability to real-world food matrices containing pigments, sugars, fats, or proteins that introduce spectral interference. Moisture interference remains a key challenge, particularly in the MIR region, where strong water absorption overlaps with vitamin C signals, making FTIR less reliable for high-moisture foods.

The spectral range limitations in several studies (e.g., 381–1065 nm and 912–1667 nm) excluded deeper NIR or MIR regions where stronger overtone bands of vitamin C occur, leading to reliance on indirect spectral correlations with water, color, or carbohydrate features rather than distinct vitamin C peaks [16] [17]. Linear chemometric models (PLS, MLR) dominated the analyses, with limited exploration of nonlinear or hybrid methods (e.g., SVM, ANN, kNN ensembles) that might capture more complex relationships. Furthermore, models were often product- or instrument-specific, built under controlled laboratory settings with helium-purged systems or stable lighting, and hence required recalibration for new instruments or sample types.

Datasets were typically small and homogeneous (often fewer than 200 samples within narrow concentration ranges), increasing the risk of overfitting (especially in kNN models with $R^2 \approx 0.999$) and limiting robustness across wider natural variability. The lack of external and independent validation often relying only on hold-out subsets from the same batch further constrains generalization. Additionally, preprocessing dependence (e.g., only NC or SNV applied) and limited feature-selection diversity (e.g., absence of MSC, derivatives, or adaptive scatter corrections) reduce reproducibility and optimization potential.

Most studies compared only between spectroscopic techniques without benchmarking against high-precision reference methods such as HPLC or UPLC; instead, they relied on titrimetric assays that introduce reference error. Instrumental and environmental dependencies such as the need for purged atmospheres, stable temperature, and fixed angles make field or industrial deployment difficult. Physical factors like particle size, scattering, and unstandardized sample preparation may also distort spectral linearity. Finally, while wavelength selection methods (such as SR) successfully identified key spectral regions, few studies provided chemical interpretations of these wavelengths, limiting mechanistic understanding.

In [18], the study successfully demonstrates that nano-sized starch particles improve encapsulation and stability for multivitamins, but it does not quantify or analyze vitamin C, the most oxidation-sensitive vitamin. Analytical techniques (FTIR, XRD, fluorescence) were used qualitatively, not quantitatively; no chemometric or correlation model was developed.

Overall, these limitations highlight the need for future research that incorporates diverse and real-world food matrices, employs larger and more varied datasets, integrates nonlinear and hybrid chemometric approaches, and performs rigorous cross-validation against chromatographic standards under variable environmental conditions, optimization and wavelength selection to achieve robust, interpretable, and transferable vitamin C prediction models.

5 Discussion and Possible Research Directions

Chemometric approaches have demonstrated high accuracy and efficiency in the quantitative determination of vitamin C in various matrices using NIR and FTIR spectroscopy. Based on the synthesis of the reviewed studies and the

current research scope, future investigations will aim to extend and optimize chemometric applications for broader vitamin analysis, model generalization, and stability assessment in complex food systems.

Possible Research Questions

- Can we identify an optimal sampling and preprocessing protocol to minimize scattering and baseline variation in different vitamin-rich matrices (fresh produce, processed foods, and encapsulated forms)?

- Can we determine the most informative wavelength regions for differentiating and quantifying vitamin C under varying physical and chemical conditions?

- Can we develop a robust and transferable chemometric model capable of accurately predicting vitamin C across multiple sample types and instrumental platforms?

- Can hybrid or ensemble models combining PLSR, MLR, and non-linear machine learning algorithms (e.g., kNN, ANN, SVR) further enhance prediction accuracy and model interpretability?

Derived from these research questions, the key objectives are:

- Acquire representative NIR of vitamin C from diverse sources (fresh fruits, powders, and encapsulated systems) to build a comprehensive reference library.

- Perform systematic preprocessing (e.g., NC, SNV, MSC, derivative transformations) and wavelength selection (e.g., iPLS, SR, CARS, SPA) to identify optimal feature sets for robust prediction.

- Develop and validate chemometric calibration models (PLSR, SR–NC–MLR, kNN) using cross-validation and external datasets to ensure generalizability.

- Extend modeling frameworks to other vitamins (B_5, B_9, D, E) to explore multi-vitamin quantification using spectral data fusion.

- Integrate findings with nano-encapsulation research to establish correlations between spectral characteristics and vitamin stability profiles under environmental and processing stresses.

This research pathway aims to advance the non-destructive, rapid, and eco-friendly quantification of vitamins, leading to practical tools for real-time quality control, food fortification monitoring, and nutritional analytics in the agri-food and pharmaceutical sectors.

6   Conclusion

This synthesis consolidates the findings of four key studies that collectively demonstrate the advancement of Near-Infrared (NIR) spectroscopy and chemometric modeling as reliable, rapid, and non-destructive tools for determining vitamin C concentration across various biological and food matrices. The integrated review of [15], [16], [17], and [18] underscores a paradigm shift in analytical chemistry from reagent-based, time-consuming assays to data-driven spectrochemical modeling frameworks.

Across the studies, NIR, FTIR, and Vis–NIR spectroscopy proved capable of achieving high predictive accuracy (R^2 values between 0.95–0.999) for vitamin C quantification using models such as Partial Least Squares Regression (PLSR), Multiple Linear Regression (MLR), and k-Nearest Neighbors (kNN). Proper preprocessing techniques notably Normalization Correction (NC) and Standard Normal Variate (SNV) were essential for minimizing spectral scattering, improving calibration, and enhancing model robustness. Meanwhile, Stepwise Regression (SR) and related wavelength-selection algorithms successfully reduced redundant spectral variables, enabling efficient and interpretable linear models without significant loss of accuracy.

The reviewed evidence highlights that machine learning integration further enhances the predictive capacity of chemometric models. For instance, the kNN model in [16] achieved near-perfect prediction for vitamin C in red chilies ($r \approx 0.999$, RPD ≈ 31.6), proving the potential of nonlinear approaches in handling complex spectral datasets. Similarly, [17] showed that streamlined SR–NC–MLR models using only 15 optimal wavelengths can rival full-spectrum PLSR models in citrus peel analysis, indicating the feasibility of simplified, cost-effective industrial applications.

Complementing these quantification studies, in [18] contributed by exploring nano-starch encapsulation, demonstrating that nano-sized carriers significantly improve vitamin stability, encapsulation efficiency, and molecular retention. Although not centered on vitamin C quantification, this work extends the analytical framework toward vitamin preservation, suggesting a future integration between spectroscopic quantification and nanotechnological stabilization.

However, several limitations remain. The models were developed on limited, homogeneous datasets (≤ 200 samples), often restricted to specific products (e.g., chilies, citrus peels), and validated under controlled laboratory conditions. Consequently, model transferability across instruments, environmental conditions, and food matrices remains untested. Moreover, reference data were derived mainly from titrimetric assays rather than chromatographic gold standards (HPLC/UPLC), potentially introducing calibration bias.

In summary, this synthesis concludes that NIR spectroscopy combined with chemometric and machine learning models provides a powerful, eco-friendly, and high-throughput alternative to conventional analytical techniques for vitamin C quantification. Future research should focus on expanding sample diversity, integrating nonlinear ensemble models (e.g., hybrid PLS–ANN, SVM), implementing cross-platform validation, and merging quantification with stability

analysis through nano-encapsulation and spectral tracking. Such advancements will enable real-time, portable, and universal models for vitamin C monitoring, contributing to sustainable quality control in the food, nutrition, and pharmaceutical sectors.

References

[1] National Institutes of Health, Office of Dietary Supplements, "Vitamin c — health professional fact sheet," https://ods.od.nih.gov/factsheets/VitaminC-HealthProfessional/, 2024, accessed: 2025-10-08.

[2] S. Arya, M. Mahajan, and P. Jain, "Non-spectrophotometric methods for the determination of vitamin c," *Analytica Chimica Acta*, vol. 417, no. 1, pp. 1–14, 2000.

[3] R. R. Eitenmiller, W. Landen Jr, and L. Ye, *Vitamin analysis for the health and food sciences*. CRC press, 2016.

[4] M. Kurek, N. Benaida-Debbache, I. Elez Garofulić, K. Galić, S. Avallone, A. Voilley, and Y. Waché, "Antioxidants and bioactive compounds in food: critical review of issues and prospects," *Antioxidants*, vol. 11, no. 4, p. 742, 2022.

[5] J. B. Johnson, K. B. Walsh, M. Naiker, and K. Ameer, "The use of infrared spectroscopy for the quantification of bioactive compounds in food: A review," *Molecules*, vol. 28, no. 7, p. 3215, 2023.

[6] C. Hong, M. Shi, S.-X. Wang, Y. Yang, and Z. Pu, "Novel analysis based on raman spectroscopy in nutrition science," *Analytical Methods*, 2025.

[7] A. Madurapperumage, N. Johnson, P. Thavarajah, L. Tang, and D. Thavarajah, "Fourier-transform infrared spectroscopy (ftir) as a high-throughput phenotyping tool for quantifying protein quality in pulse crops," *The Plant Phenome Journal*, vol. 5, no. 1, p. e20047, 2022.

[8] H. Z. Amanah, S. S. Tunny, R. E. Masithoh, M.-G. Choung, K.-H. Kim, M. S. Kim, I. Baek, W.-H. Lee, and B.-K. Cho, "Nondestructive prediction of isoflavones and oligosaccharides in intact soybean seed using fourier transform near-infrared (ft-nir) and fourier transform infrared (ft-ir) spectroscopic techniques," *Foods*, vol. 11, no. 2, p. 232, 2022.

[9] N. Quijano-Ortega, C. A. Fuenmayor, C. Zuluaga-Dominguez, C. Diaz-Moreno, S. Ortiz-Grisales, M. García-Mahecha, and S. Grassi, "Ftir-atr spectroscopy combined with multivariate regression modeling as a preliminary approach for carotenoids determination in cucurbita spp." *Applied Sciences*, vol. 10, no. 11, p. 3722, 2020.

[10] L. F. Leopold, N. Leopold, H.-A. Diehl, and C. Socaciu, "Quantification of carbohydrates in fruit juices using ftir spectroscopy and multivariate analysis," *Journal of Spectroscopy*, vol. 26, no. 2, pp. 93–104, 2011.

[11] M. Lesnoff, D. Andueza, C. Barotin, P. Barre, L. Bonnal, J. A. Fernandez Pierna, F. Picard, P. Vermeulen, and J.-M. Roger, "Averaging and stacking partial least squares regression models to predict the chemical compositions and the nutritive values of forages from spectral near infrared data," *Applied Sciences*, vol. 12, no. 15, p. 7850, 2022.

[12] W. Q. Khan, M. Farrar, M. Awrangjeb, S. H. Bai, S. J. Trueman, H. Wallace, T. E. Richards, and W. Arshid, "Assessment of macadamia nutrients using hyperspectral data and machine learning," in *2024 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2024, pp. 530–537.

[13] L. P. Osco, A. P. M. Ramos, M. M. Faita Pinheiro, É. A. S. Moriya, N. N. Imai, N. Estrabis, F. Ianczyk, F. F. d. Araújo, V. Liesenberg, L. A. d. C. Jorge *et al.*, "A machine learning framework to predict nutrient content in valencia-orange leaf hyperspectral measurements," *Remote Sensing*, vol. 12, no. 6, p. 906, 2020.

[14] D. Wang, M. Zhang, B. Adhikari, and L. Zhang, "Determination of polysaccharide content in shiitake mushroom beverage by nir spectroscopy combined with machine learning: A comparative analysis," *Journal of Food Composition and Analysis*, vol. 122, p. 105460, 2023.

[15] H. Yang and J. Irudayaraj, "Rapid determination of vitamin c by nir, mir and ft-raman techniques," *Journal of Pharmacy and Pharmacology*, vol. 54, no. 9, pp. 1247–1255, 2002.

[16] S. Hafsah, F. Ahmad, N. D. Arianti, E. Saputra, S. Hartuti *et al.*, "Rapid and non-destructive determination of vitamin c and antioxidant activity of intact red chilies using visible near-infrared spectroscopy and machine learning tools," *Case Studies in Chemical and Environmental Engineering*, vol. 8, p. 100435, 2023.

[17] W. Zhang, M. Lin, H. He, Y. Wang, J. Wang, and H. Liu, "Toward achieving rapid estimation of vitamin c in citrus peels by nir spectra coupled with a linear algorithm," *Molecules*, vol. 28, no. 4, p. 1681, 2023.

[18] M. Ahmad, I. Hassan, M. Shah, A. Gani, and K. Muthukumarappan, "Co-encapsulation of multivitamins in micro & nano-sized starch, target release, capsule characterization and interaction studies," *International Journal of Biological Macromolecules*, vol. 240, p. 124367, 2023.

[19] K. R. Borba, P. C. Spricigo, D. P. Aykas, M. C. Mitsuyuki, L. A. Colnago, and M. D. Ferreira, "Non-invasive quantification of vitamin c, citric acid, and sugar in 'valência'oranges using infrared spectroscopies," *Journal of Food Science and Technology*, vol. 58, no. 2, pp. 731–738, 2021.

[20] V. A. Tirado-Kulieva, E. Hernández-Martínez, and J.-P. Suomela, "Non-destructive assessment of vitamin c in foods: a review of the main findings and limitations of vibrational spectroscopic techniques," *European Food Research and Technology*, vol. 248, no. 8, pp. 2185–2195, 2022.

[21] N. Khuriyati, A. C. Sukartiko, M. A. F. Falah, and R. N. Alfiani, "Determination of total carotene and vitamin c in chili powder (capsicum annuum l.) non-destructively using near-infrared spectroscopy," *Int. J. Adv. Sci. Eng. Inf. Techno.*, vol. 13, pp. 69–76, 2023.

[22] F. P. d. Moraes, R. C. Costa, C. d. L. M. d. Morais, F. G. M. d. Medeiros, T. R. N. Fernandes, R. T. Hoskin, and K. M. G. d. Lima, "Estimation of ascorbic acid in intact acerola (malpighia emarginata dc) fruit by nirs and chemometric analysis," *Horticulturae*, vol. 5, no. 1, p. 12, 2019.

[23] C. Malegori, E. J. N. Marques, S. T. de Freitas, M. F. Pimentel, C. Pasquini, and E. Casiraghi, "Comparing the analytical performances of micro-nir and ft-nir spectrometers in the evaluation of acerola fruit quality, using pls and svm regression algorithms," *Talanta*, vol. 165, pp. 112–116, 2017.

[24] K. Kusumiyati, A. A. Munawar, and D. Suhandy, "Fast, simultaneous and contactless assessment of intact mango fruit by means of near infrared spectroscopy," *AIMS Agriculture and Food*, vol. 6, no. 1, pp. 172–184, 2021.

[25] R. Alhajjeh and H. Al-Ali, "Application of ftir spectroscopy method for the quantification of ascorbic acid in bulk materials and pharmaceutical formulation," *Iraqi Journal of Pharmaceutical Sciences*, vol. 32, no. 3, pp. 186–194, 2023.

[26] A. Todorovski, M. Petreska, K. Shapkovska, N. Geskovski, and J. Acevska, "Screening the capability of atr-ftir for simultaneous quantification of vitamin b1, b6 and b12 in powder blend," *Maced. Pharm. Bull.*, vol. 68, pp. 43–44, 2022.

[27] J. Zhang, G. Liu, Y. Li, M. Guo, F. Pu, and H. Wang, "Rapid identification of lamb freshness grades using visible and near-infrared spectroscopy (vis-nir)," *Journal of Food Composition and Analysis*, vol. 111, p. 104590, 2022.

[28] K. Najjar and N. Abu-Khalaf, "Non-destructive quality measurement for three varieties of tomato using vis/nir spectroscopy," *Sustainability*, vol. 13, no. 19, p. 10747, 2021.

[29] M. del Río Celestino and R. Font, "Using vis-nir spectroscopy for predicting quality compounds in foods," p. 4845, 2022.

[30] H. Kaur, B. Rana, D. Tomar, S. Kaur, and K. C. Jena, "Fundamentals of atr-ftir spectroscopy and its role for probing in-situ molecular-level interactions," in *Modern Techniques of Spectroscopy: Basics, Instrumentation, and Applications*. Springer, 2021, pp. 3–37.

[31] A. Whitley, E. Leroy, and F. Adar, "Ft-ir–raman combination: The perfect analytical solution for vibrational spectroscopists," 2009.

[32] D. W. Shipp, F. Sinjab, and I. Notingher, "Raman spectroscopy: techniques and applications in the life sciences," *Advances in Optics and Photonics*, vol. 9, no. 2, pp. 315–428, 2017.

[33] F. Bergin and H. Shurvell, "Applications of fourier transform raman spectroscopy in an industrial laboratory," *Applied spectroscopy*, vol. 43, no. 3, pp. 516–522, 1989.

[34] E. Smith and G. Dent, *Modern Raman spectroscopy: a practical approach*. John Wiley & Sons, 2019.

[35] R. Pandiselvam, V. Prithviraj, M. Manikantan, A. Kothakota, A. V. Rusu, M. Trif, and A. Mousavi Khaneghah, "Recent advancements in nir spectroscopy for assessing the quality and safety of horticultural products: A comprehensive review," *Frontiers in Nutrition*, vol. 9, p. 973457, 2022.

[36] H. E. Tahir, Z. Xiaobo, X. Jianbo, G. K. Mahunu, S. Jiyong, J.-L. Xu, and D.-W. Sun, "Recent progress in rapid analyses of vitamins, phenolic, and volatile compounds in foods using vibrational spectroscopy combined with chemometrics: A review," *Food Analytical Methods*, vol. 12, no. 10, pp. 2361–2382, 2019.

[37] M. Ahmad and A. Gani, "Ultrasonicated resveratrol loaded starch nanocapsules: Characterization, bioactivity and release behaviour under in-vitro digestion," *Carbohydrate Polymers*, vol. 251, p. 117111, 2021.

[38] M. Asachi and M. A. Camargo-Valero, "Multi-sensors data fusion for monitoring of powdered and granule products: Current status and future perspectives," *Advanced Powder Technology*, vol. 34, no. 7, p. 104055, 2023.

[39] Å. Rinnan, F. Van Den Berg, and S. B. Engelsen, "Review of the most common pre-processing techniques for near-infrared spectra," *TrAC Trends in Analytical Chemistry*, vol. 28, no. 10, pp. 1201–1222, 2009.

[40] Z. Xiaobo, Z. Jiewen, M. J. Povey, M. Holmes, and M. Hanpin, "Variables selection methods in near-infrared spectroscopy," *Analytica chimica acta*, vol. 667, no. 1-2, pp. 14–32, 2010.

[41] Y. Dai, B. Yan, F. Xiong, R. Bai, S. Wang, L. Guo, and J. Yang, "Tanshinone content prediction and geographical origin classification of salvia miltiorrhiza by combining hyperspectral imaging with chemometrics," *Foods*, vol. 13, no. 22, p. 3673, 2024.

[42] G. Li and S. Deng, "Quantitative analysis of near-infrared spectroscopy using the best-1dconvnet model," *Processes*, vol. 12, no. 2, p. 272, 2024.

[43] P. Mishra, D. N. Rutledge, J.-M. Roger, K. Wali, and H. A. Khan, "Chemometric pre-processing can negatively affect the performance of near-infrared spectroscopy models for fruit quality prediction," *Talanta*, vol. 229, p. 122303, 2021.

[44] J. Luypaert, S. Heuerding, Y. Vander Heyden, and D. Massart, "The effect of preprocessing methods in reducing interfering variability from near-infrared measurements of creams," *Journal of pharmaceutical and biomedical analysis*, vol. 36, no. 3, pp. 495–503, 2004.

[45] E. Teye, C. L. Amuah, T. McGrath, and C. Elliott, "Innovative and rapid analysis for rice authenticity using hand-held nir spectrometry and chemometrics," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 217, pp. 147–154, 2019.

[46] I. S. Helland, T. Næs, and T. Isaksson, "Related versions of the multiplicative scatter correction method for preprocessing spectroscopic data," *Chemometrics and Intelligent Laboratory Systems*, vol. 29, no. 2, pp. 233–241, 1995.

[47] H. Khadem, H. Nemat, J. Elliott, and M. Benaissa, "In vitro glucose measurement from nir and mir spectroscopy: Comprehensive benchmark of machine learning and filtering chemometrics," *Heliyon*, vol. 10, no. 10, 2024.

[48] B. Zimmermann and A. Kohler, "Optimizing savitzky–golay parameters for improving spectral resolution and quantification in infrared spectroscopy," *Applied spectroscopy*, vol. 67, no. 8, pp. 892–902, 2013.

[49] C. D. Kappatou, J. Odgers, S. Garcia-Munoz, and R. Misener, "An optimization approach coupling preprocessing with model regression for enhanced chemometrics," *Industrial & Engineering Chemistry Research*, vol. 62, no. 15, pp. 6196–6213, 2023.

[50] J. A. Cayuela-Sánchez, "'specific normalization'method for nirs calibrations," *NIR news*, vol. 34, no. 3-4, pp. 7–10, 2023.

[51] S.-Y. Kim and J.-H. Ha, "Rapid determination of the geographical origin of kimchi by fourier transform near-infrared spectroscopy coupled with chemometric techniques," *Scientific Reports*, vol. 14, no. 1, p. 24581, 2024.

[52] H.-J. He, C. Zhang, X. Bian, J. An, Y. Wang, X. Ou, and M. Kamruzzaman, "Improved prediction of vitamin c and reducing sugar content in sweetpotatoes using hyperspectral imaging and lars-enhanced lasso variable selection," *Journal of Food Composition and Analysis*, vol. 132, p. 106350, 2024.

[53] A. Demiriz, K. Bennett, C. Breneman, and M. Embrechts, "Support vector machine regression in chemometrics, computing science and statistics," in

Proceedings of the 33rd Symposium on the Interface, American Statistical Association for the Inter face Foundation of North America. Washington, DC, 2001.

[54] D. Malavi, A. Nikkhah, P. Alighaleh, S. Einafshar, K. Raes, and S. Van Haute, "Detection of saffron adulteration with crocus sativus style using nir-hyperspectral imaging and chemometrics," *Food Control*, vol. 157, p. 110189, 2024.

[55] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of k-nearest neighbour (knn) algorithm and its different variants for disease prediction," *Scientific Reports*, vol. 12, no. 1, p. 6256, 2022.

[56] L. M. Cozmuta, "The application of multiple linear regression methods to ftir spectra of fingernails for predicting gender and age of human subjects," *Heliyon*, vol. 11, no. 4, 2025.

[57] N. Roustaei, "Application and interpretation of linear-regression analysis," *Medical Hypothesis, Discovery and Innovation in Ophthalmology*, vol. 13, no. 3, p. 151, 2024.

[58] L. Norgaard, A. Saudland, J. Wagner, J. P. Nielsen, L. Munck, and S. B. Engelsen, "Interval partial least-squares regression (ipls): A comparative chemometric study with an example from near-infrared spectroscopy," *Applied spectroscopy*, vol. 54, no. 3, pp. 413–419, 2000.

[59] F. Liu, T. Shen, W. Kong, J. Peng, C. Zhang, K. Song, W. Wang, C. Zhang, and Y. He, "Quantitative analysis of cadmium in tobacco roots using laser-induced breakdown spectroscopy with variable index and chemometrics," *Frontiers in plant science*, vol. 9, p. 1316, 2018.

[60] J.-H. Jiang, R. J. Berry, H. W. Siesler, and Y. Ozaki, "Wavelength interval selection in multicomponent spectral analysis by moving window partial least-squares regression with applications to mid-infrared and near-infrared spectroscopic data," *Analytical chemistry*, vol. 74, no. 14, pp. 3555–3565, 2002.

[61] W. Yang, W. Wang, R. Zhang, F. Zhang, Y. Xiong, T. Wu, W. Chen, and Y. Du, "A modified moving-window partial least-squares method by coupling with sampling error profile analysis for variable selection in near-infrared spectral analysis," *Analytical Sciences*, vol. 36, no. 3, pp. 303–309, 2020.

[62] J. Xie, D. Sun, J. Cai, and F. Cai, "Waveband selection with equivalent prediction performance for ftir/atr spectroscopic analysis of cod in sugar refinery waste water." *Computers, Materials & Continua*, vol. 59, no. 2, 2019.