Journal of Food Composition and Analysis 136 (2024) 106824

Contents lists available at ScienceDirect

Journal of Food Composition and Analysis

journal homepage: www.elsevier.com/locate/jfca

Comparative analysis of deep learning and machine learning-based models for simultaneous prediction of minerals in perilla (*Perilla frutescens* L.) seeds using near-infrared reflectance spectroscopy

Naseeb Singh [a,1], Simardeep Kaur [a,1], Antil Jain [b], Amit Kumar [a], Rakesh Bhardwaj [b,*], Renu Pandey [c,*], Amritbir Riar [d,*]

[a] *ICAR Research Complex for North Eastern Hill Region, Umiam, Meghalaya 793 103, India*
[b] *ICAR, National Bureau of Plant Genetic Resources, New Delhi 110 012, India*
[c] *ICAR, Indian Agricultural Research Institute, New Delhi 110 012, India*
[d] *Department of International Cooperation Research Institute of Organic Agriculture, FiBL, Frick 11, Switzerland*

ARTICLE INFO

Keywords:
Perilla seeds
Deep learning
Minerals prediction
NIRS
Germplasm screening
Machine learning

ABSTRACT

Perilla seeds contain a rich array of essential minerals, thus having the potential to address multiple micronutrient deficiencies at a time. However, traditional methods of mineral estimation are complex, time-consuming, expensive, and require technical expertise. This study includes the development of Near-Infrared Reflectance Spectroscopy (NIRS)-based prediction models for predicting five important minerals (Calcium, Copper, Magnesium, Manganese, and Phosphorus) using machine learning and deep learning techniques. Four models, including 1D Convolutional Neural Networks (1D CNNs), Artificial Neural Networks (ANNs), Random Forests (RFs), and Support Vector Regression (SVR), were developed and evaluated. The developed 1D CNN model outperformed other considered models in predicting calcium, magnesium, and phosphorus content with RPD (Residual Prediction Deviation) values of 1.75, 1.83, and 2.96, respectively. Whereas, SVR performed best in predicting copper and manganese with an RPD of 1.82 and 2.2, respectively. The 1D CNN model demonstrated R^2 (Coefficient of determination) values above 0.65 for all minerals, with a maximum of 0.88 for phosphorus. In addition, the developed models performed superior as compared to the Partial Least Square Regression method (R^2= 0.32). The developed models provide efficient tools for rapidly screening perilla germplasm available in global repositories, thus aiding in the selection of mineral-rich genotypes to mitigate micronutrient deficiencies.

1. Introduction

Orphan or underutilized crops are increasingly recognized as important components of global food and nutritional security strategies due to their inherent nutritional richness and adaptability to diverse ecological conditions (Mabhaudhi et al., 2019; Talabi et al., 2022). The North Eastern Hill (NEH) region of India is a treasure trove of biodiversity and a unique ecosystem in the world, encompassing a diverse array of germplasm and an extensive collection of potentially underutilized crops. *Perilla frutescence* L. is one such potential herbaceous plant that belongs to the mint family Lamiaceae and is native to East Asia, including China, Japan, Korea, and India. Perilla seeds are highly nutritious, containing 35.0–45.0 % polyunsaturated fatty acids (PUFAs) and protein levels ranging from 15.7 % to 23.9 %, making them superior to conventional oil sources in terms of both fatty acid composition and protein content (Longvah and Deosthale, 1998; Longvah et al., 2000). Beyond their nutritional value, perilla seeds and oil have significant industrial and therapeutic potential. Industrially, perilla oil is used in the production of varnishes, paints, and inks due to its quick-drying properties. Therapeutically, perilla oil is rich in omega-3 fatty acids, which have been linked to anti-inflammatory, antioxidant, and cardiovascular health benefits. The plant is also used in traditional medicine to treat respiratory ailments, allergies, and other inflammatory conditions, further highlighting its wide-ranging applications ((Dhyani et al., 2019; Kaur et al., 2024; Wu et al., 2023).

However, the nutritional significance of perilla encompasses not only

* Corresponding authors.
E-mail addresses: rb_biochem@yahoo.com (R. Bhardwaj), renu.pandey.iari@gmail.com (R. Pandey), Amritbir.riar@fibl.org (A. Riar).
[1] Both the authors share first authorship.

https://doi.org/10.1016/j.jfca.2024.106824
Received 22 June 2024; Received in revised form 26 September 2024; Accepted 2 October 2024
Available online 3 October 2024

its fatty acid and protein content but also includes a diverse array of essential minerals crucial for human health, including calcium, copper, magnesium, manganese, phosphorus, and zinc. These minerals are important for various physiological and metabolic processes in both humans and animals (An et al., 2022; González-Montaña et al., 2020; Long and Romani, 2014; Tang et al., 2023). Despite its mineral-rich potential, utilization of perilla remains underexplored, primarily due to the predominant focus on its oil content, leading to the discarding of the defatted seed meal (Kim et al., 2019; Kim and Yoon, 2020). However, several studies have explained the potential of perilla seeds as natural antioxidants and as valuable additions to animal diets, contributing to health improvement and enhanced production (Arjin et al., 2020; Kim et al., 2019; Ruamrungsri et al., 2016). The significant mineral content of perilla seeds and their potential to enhance food and feed quality necessitate the accurate estimation of their mineral content. This requires screening large germplasm to assess mineral levels precisely. Traditional methods such as Atomic Absorption Spectroscopy (AAS), Inductively Coupled Plasma Optical Emission Spectroscopy (ICP OES), and spectrophotometry have been used for this purpose (Khan et al., 2022; Yeung et al., 2017). However, generally, these methods involve complex pretreatment procedures, time-consuming, significant financial investments, and specialized technical expertise (Baianu and Guo, 2011).

In contrast, Near-Infrared Reflectance Spectroscopy (NIRS) offers a promising alternative, providing a non-destructive, cost-effective, and sustainable solution for rapid nutritional profiling of diverse crop germplasm (Cozzolino, 2015; Fassio and Cozzolino, 2004). NIRS facilitates the development of calibration models by correlating spectral attributes with desired parameters, enabling real-time, non-destructive analysis, and streamlining crop nutrition assessment and breeding programs. Previous studies have successfully utilized NIRS-based prediction modeling for the biochemical assessment of various food items; for example, Fassio and Cozzolino (2004) effectively used NIRS to predict the biochemical composition of sunflower seeds through MPLS, while Wang et al. (2014) analyzed the nutritional composition of *faba bean* seeds using NIRS through PLS regression. Additionally, Gohain et al. (2021) developed an NIRS model using Modified Partial Least Square (MPLS) regression for Brassica oilseed. However, recent advances in machine learning (ML) algorithms have further enhanced the predictive accuracy of NIRS models, resulting in better performance compared to traditional methods. For instance, Posom and Maraphum (2023) predicted starch content in cassava using NIRS and ML algorithms, while Ye et al. (2022) reported that ML algorithms can be effectively utilized for the quality inspection of grapes. Bai et al. (2022) used ML learning to predict quality-related parameters of tea. Additionally, Folli et al. (2022) found that SVM provided better results than PLS in analyzing food composition.

Despite the extensive application of NIRS-based prediction modeling in estimating various biochemical constituents across different crops, to the best of our knowledge, mineral content prediction models for Perilla have not been reported so far. One of the challenges in developing such models lies in the poor absorption of energy by minerals in the NIR region, as minerals are often bound (rather than in a free form) to organic molecules. Despite this, accurately predicting mineral content is important for selecting genotypes with high nutritional value. Such predictions can facilitate the development of industrial applications and food products, helping to address micronutrient deficiencies and improve human health through nutritionally superior perilla-based foods., To address this gap, the current study focuses on developing predictive models of the mineral content of perilla seeds by utilizing 88 perilla germplasm samples collected from the NEH region of India. This study used ICP-OES for reference data generation and NIR spectroscopy for spectral acquisition. Utilizing 1D Convolutional Neural Networks (1D CNNs), Artificial Neural Networks (ANNs), Random Forests Regression (RFR), and Support Vector Regression (SVR) algorithms, regression models were developed for profiling the mineral content including calcium, copper, magnesium, manganese, and phosphorus of perilla seeds. The assessment of the developed models was conducted using independent test data. This study is the first report of the integration of NIR spectroscopy with ML and DL models to determine essential minerals in perilla seeds. By developing mineral-specific models, our research aims to provide a novel approach for accurately assessing mineral content, thus, advancing analytical methodologies in agricultural science and nutritional research.

2. Materials and methods

2.1. Overall framework of the present study

The workflow followed in the present study for predicting mineral composition in perilla seeds involved several key steps, as depicted in Fig. 1. Initially, perilla seed samples were collected and loaded into a Near Infrared Spectrometer (NIRS) for spectral analysis. The internal components of the NIRS instrument facilitated the detection of NIR signals emitted from the dried seed samples. These signals were then used to generate a typical NIRS spectrum, indicating peaks corresponding to different sample characteristics. Mineral content estimation for calcium, copper, magnesium, manganese, and phosphorus was conducted for the 88 samples using Inductively Coupled Plasma - Optical Emission Spectroscopy (ICP-OES). Subsequently, the data obtained from ICP-OES analysis was input into the NIRS calibration file. NIRS-based models were then developed using machine learning algorithms, including ANN, RFR, and SVR. Additionally, the 1D CNN technique was used to develop NIRS-based models. Finally, the developed models were validated and their performance was assessed using test data.

2.2. Plant materials and sample preparation

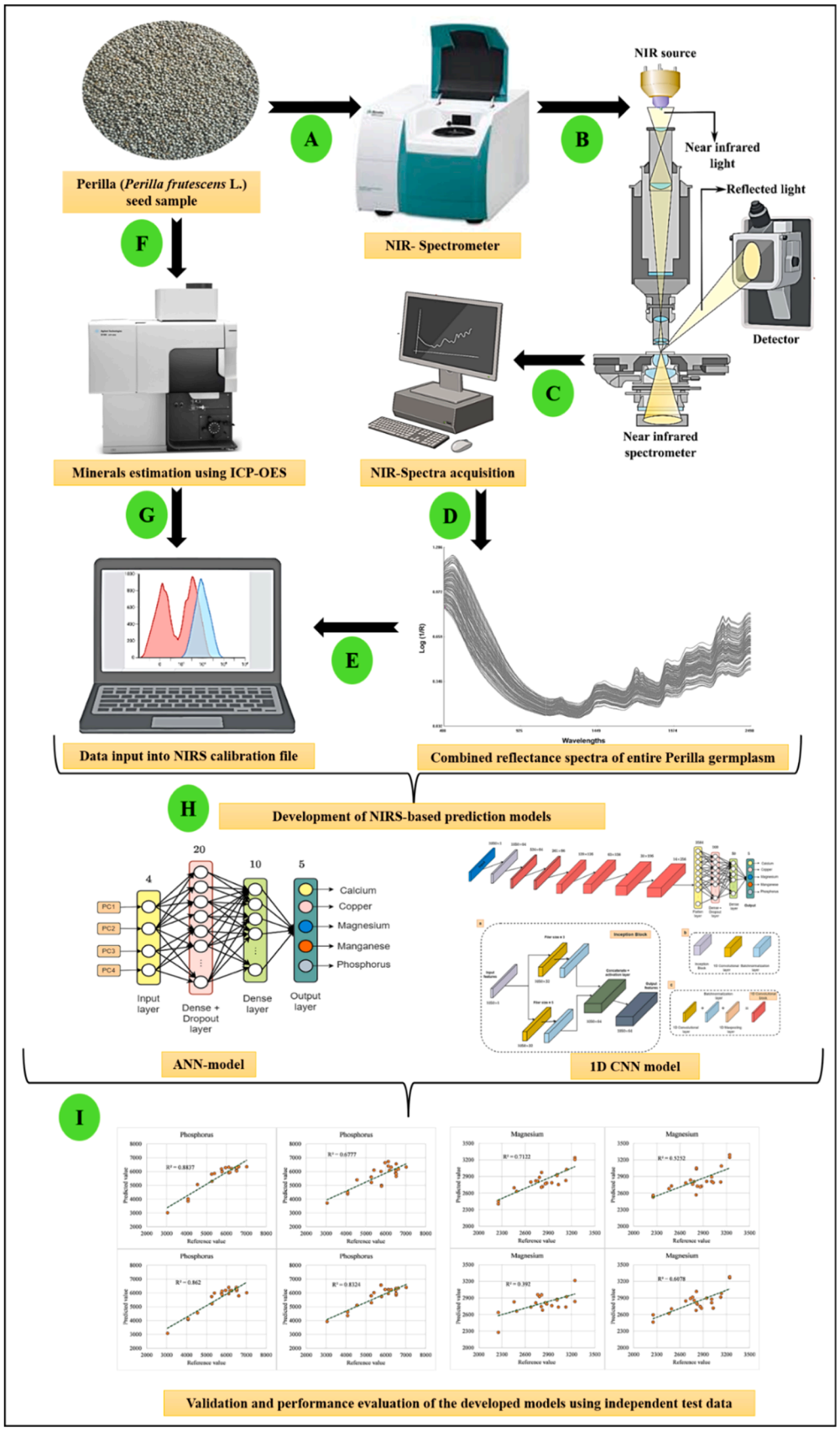
A total of 88 diverse perilla germplasm, collected from five different locations of North Eastern Hill (NEH) region states including Arunachal Pradesh, Manipur, Meghalaya, Nagaland, and Sikkim of India, were considered in the present study. These genotypes were cultivated using an augmented design with seven blocks and three checks, thereby ensuring robust error degrees of freedom. The perilla seeds were grown at the experimental farm of the Division of Crop Science, ICAR-Research Complex for North Eastern Hill Region, Umiam, Meghalaya, India, (Latitude: 25.66° N; Longitude: 91.83° E) during Kharif season (June to October) of the year 2023. The plants were grown under upland conditions, with each plot having an area of 3.5 × 3.0 m and a defined plant spacing (50.0 × 35.0 cm). Standard agronomic practices were followed throughout the crop season to maintain the crop. The seeds were harvested at the physiological maturity stage and dried in a hot air oven at approx. 60ºC for 12 hours. The dried seeds were cleaned and then stored in an air-tight plastic container for further laboratory analyses.

2.3. Minerals analysis using ICP-OES

Mineral content analysis, including calcium, copper, magnesium, manganese, and phosphorus was conducted using 0.5 g of dried perilla seed samples. Using a microwave digestor, the samples were subjected to digestion with HNO_3 (10 mL) and 2 mL of H_2O_2. Following digestion, the resulting extract was adjusted to a final volume of 50 mL and filtered through the Whatman No. 42 filter paper (pore size of ~2.5 microns) (Tomar et al., 2021a). The samples were then analyzed using an inductively coupled plasma optical emission spectrometer (ICP-OES) (Agilent Technologies, Model 5110 ICP-OES, Santa Clara, CA, USA) calibrated with standard solutions. The concentration of all the minerals was recorded in parts per million (ppm).

2.4. NIR-spectra acquisition

The perilla seed samples were kept at room temperature (25 °C) for



(caption on next page)

Fig. 1. Illustration of workflow employed in the present study for assessing mineral composition in perilla seeds using Near-Infrared Spectroscopy (NIRS) combined with machine learning (ML) and deep learning (DL) techniques. (A) Collection of perilla seed samples; (B) Loading of samples into a Near Infrared Spectrometer (NIRS); (C) Internal components of a typical NIRS, with spectra obtained from dried seed samples loaded in a circular ring cup with a quartz window (3.8 cm in diameter and 1 cm in thickness), and NIR signals detected by the detector; (D) Illustration of a typical NIRS spectrum indicating peaks for one sample; (E) Presentation of the average combined reflectance spectrum of 88 perilla samples; (F) Estimation of mineral content (Calcium, Copper, Magnesium, Manganese, and Phosphorus) for 88 samples using Inductively Coupled Plasma - Optical Emission Spectroscopy (ICP-OES); (G) Input of data into the NIRS calibration file; (H) Development of NIRS-based models using ML and DL algorithms; here we shown Artificial Neural Networks (ANN) and 1D Convolutional Neural Network (CNN)-models; (I) Validation and inter-performance assessment of the developed models using test data, depicted by accuracy through scatter plots between the reference and predicted values.

6 hours to standardize temperature and moisture levels, as these factors can impact the absorbance and reflectance of NIR waves. Before scanning and every 30.0 minutes thereafter, the NIR spectrometer was calibrated by scanning a check sample P/N 60053128, S/N 83924. Approximately 5.0 g of the seed samples were scanned using the FOSS NIRS DS3 spectrometer (FOSS Nils Foss Alle 1, DK-3400, Hilleroed, Denmark). To obtain the spectra, samples were loaded into a circular ring cup with a quartz window (3.8 cm in diameter and 1.0 cm in thickness). A circular cardboard backing was gently pressed onto the samples to ensure uniform packing without air pockets. Each spectrum represented an average of 32 scans across the range of 400–2500 nm and was recorded as log (1/R) (where R is relative reflectance) at 0.5 nm intervals. A comprehensive evaluation of the entire perilla germplasm, comprising both landraces and accessions from various regions of the NEH region of India, was conducted to estimate the mineral content using ICP-OES. The collective near-infrared (NIR) spectra spanning the wavelength range of 400–2500 nm for seeds sourced from a diverse perilla germplasm collection in the NEH region of India are presented in Fig. 2.

2.5. Development of NIRS-based predictive models

For the development of NIRS-based predictive models for mineral prediction in perilla seeds, two ML-based algorithms (Random Forest Regressor (RFR) and Support Vector Regressor (SVR)) and two DL-based algorithms (Artificial Neural Network (ANN) and 1D Convolutional Neural Networks (1D CNN) were considered in the present study.

2.5.1. Random forest regressor

The random forest algorithm (Breiman, 2001), a popular ML technique based on decision trees, is widely applied for its robustness and accuracy (Fawagreh et al., 2014). It works by constructing a multitude of decision trees, each trained on a bootstrapped subset of the data and a random selection of features. These decision trees act like individual models, predicting the target variable based on the input features (Biau and Scornet, 2016). In the present study, to optimize the performance of the random forest model, hyperparameter tuning via grid search was conducted. This technique involved testing a range of hyperparameter values to identify the optimal configuration. In the present study, grid search used various values for key hyperparameters, including the number of trees (*n_estimators*: 50, 100, or 200), maximum number of features (*max _features*: 'auto' or 'sqrt'), maximum depth of each tree (*max _depth*: 2, 5, 7), minimum number of samples required to split a node (*min _samples_split*: 2, 5, or 10), and minimum number of samples required at each leaf node (*min _samples_leaf*: 1, 2, or 4).

Upon completion of the grid search, a combination of 100 estimators, 'auto' for *max _features*, 2 for *max _depth*, 2 for *min _samples_split*, and 1 for *min _samples_leaf* was found to be best-performer as hyperparameters. Thus, the RFR model was trained with these hyperparameters to maximize the predictive performance of the model.

2.5.2. Support vector regressor

Support Vector Regressor (SVR) (Cortes and Vapnik, 1995) was applied in the present study for the prediction of minerals in perilla seeds. It is a widely used ML algorithm and is known for its effectiveness in finding optimal hyperplanes for class separation (Gualtieri and Chettri, 2000). Using the scikit-learn library (Pedregosa et al., 2011) in Python, an SVR model was constructed. To maximize the performance of the SVR model, hyperparameter tuning through grid search was conducted (Duan et al., 2003). This process involved evaluating various combinations of hyperparameters to identify the optimal configuration. In the present study, the optimized value of the regularization parameter (C), kernel type, and kernel coefficient (gamma) were determined.

The grid search spanned a range of C (regularization parameter) values, including 0.1, 1.0, 10.0, and 100.0, to explore different levels of regularization. Additionally, multiple kernel types, including linear, radial basis function (RBF), and polynomial, were considered to capture different decision boundary shapes. Furthermore, the gamma values were varied, including 'scale' and 'auto', to adjust the influence of individual training samples on the decision boundary. Model performance

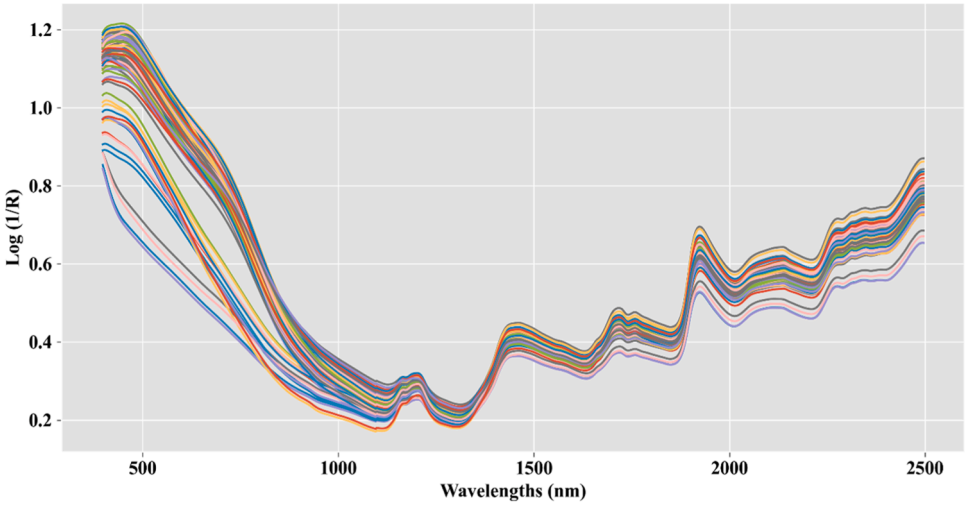


Fig. 2. A combined plot of the reflectance spectrum of the entire perilla germplasm (88 samples).

was evaluated using cross-validation techniques to ensure robustness and reliability (Sweet et al., 2023). Following the grid search process, the best-performing hyperparameters for the SVR model were determined. The optimal configuration of the regressor included a regularization parameter of C=1.0, employing a 'rbf' kernel, and setting gamma to 'scale'.

2.5.3. Artificial neural network

Artificial neural network (ANN) models can effectively capture the complex non-linear relationships present in the input and target data (Hornik et al., 1989); thus, the ANN model was used for predicting the mineral content in perilla. The architecture of the ANN model for the prediction of mineral composition comprised two hidden layers, with the first layer consisting of 20 neurons and the second layer comprising 10 neurons. The rectified linear unit (ReLU) activation function (Nair and Hinton, 2010) can effectively handle nonlinearities in the data, the ReLU activation function was used for both hidden layers. To mitigate the risk of overfitting during the training of the model, a dropout layer was incorporated after the first hidden layer. Dropout is a regularization technique that randomly drops a fraction of neurons during training, preventing the network from becoming dependent on specific neurons (Srivastava et al., 2014). In this ANN model, a dropout rate of 20.0 % was applied after the first hidden layer, ensuring that the network remains robust and generalizes well to unseen data. For the output layer of the ANN model, a linear activation function was used (Rumelhart et al., 1986). This activation function is suitable for regression tasks, as it allows the model to output continuous values without imposing any constraints on the predicted mineral content. Fig. 3 shows the architecture of the ANN model designed for predicting mineral content in perilla seeds.

2.5.4. 1D convolutional neural networks

In the present study, in addition to the SVR, RFR, and ANN model, 1D Convolutional Neural Networks (1D CNNs) were also utilized for mineral predictions. 1D CNNs can effectively capture complex patterns within spectral data, thus, enabling accurate predictions (Sang et al., 2022; Shen and Viscarra Rossel, 2021). Fig. 4 shows the architecture of the 1D CNN model designed for predicting mineral content in perilla seeds. As shown in Fig. 4, the architecture of the 1D CNN model is composed of six convolutional blocks followed by a prediction head. Each convolutional block comprises a 1D convolutional layer, a batch normalization layer (Ioffe and Szegedy, 2015), and a 1D max-pooling layer. This sequential arrangement facilitates feature extraction from spectral data while reducing feature size for computational efficiency as model depth increases (Gholamalinezhad and Khosravi, 2020).

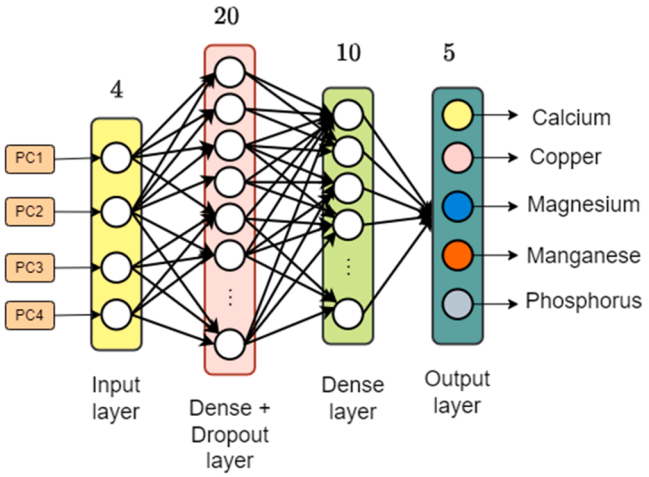


Fig. 3. The architecture of the constructed ANN model designed for predicting mineral content in perilla seeds.

To enhance the ability of the model to capture local patterns and mitigate potential noises in the spectral data, an inception module into the architecture was integrated into the architecture (Mishra et al., 2022; Zhang et al., 2021). This inception module (Fig. 4(a)) was constructed with two parallel 1D convolutional layers with different filter sizes (1×3 and 1×5) to capture both local and global features. The outputs of these parallel layers were concatenated to form a comprehensive feature representation and the non-linearity function (Nair and Hinton, 2010) was subsequently introduced using the ReLU activation function. This module was placed at the input of the model, prior to the first convolutional block, ensuring early integration of local and global information into the extracted features. At the end of the model, the prediction head was placed to predict the concentrations of various minerals. This head transforms the features extracted by the final convolutional block into a 1D tensor using a flattened layer, followed by two dense layers with 100 and 50 neurons, respectively. To prevent overfitting, a dropout layer with a 20.0 % dropout rate was inserted between these two dense layers. Throughout the model, the ReLU activation function was used to introduce the non-linearity. However, for the output layer, containing five neurons to predict mineral content, a linear activation function was utilized, as the regression task does not require additional non-linearity in mapping model predictions to continuous mineral content values (Rumelhart et al., 1986).

2.6. Pre-processing of data

In the present study, 1D CNN, ANN, RFR, and SVR models were used to predict mineral content in perilla using NIRS data. While 1D CNN models can directly process the raw spectral data, other models (ANN, SVR, and RFR models) require dimensionality reduction of the input data. To address this, principal component analysis (PCA) was utilized to reduce the dimensionality of the raw spectral data. PCA effectively captures the underlying structure and patterns within the data while reducing its dimensionality (Beattie and Esmonde-White, 2021). By transforming the raw spectral data into a lower-dimensional space represented by principal components, a more compact and informative representation of the spectral information was achieved in previous studies (Beattie and Esmonde-White, 2021; He et al., 2007, 2006; Howley et al., 2006). Principal Component Analysis (PCA) was applied to reduce the dimensionality of the spectral data, retaining four principal components. These components explained over 99.0 % of the cumulative variance, ensuring that the important spectral information necessary for accurate mineral prediction was preserved while minimizing overfitting risk in machine learning-based models. The high cumulative variance (99.0 % >) indicates that the principal components successfully captured the majority of the variability present in the raw spectral data.

Fig. 5 illustrates the cumulative variance explained by the derived principal components. This plot shows the relative importance of each principal component in summarizing the spectral information. These four principal components were utilized as input features for the random forest, support vector, and ANN models to predict mineral content in Perilla seeds.

The output variables have inherent variability across different scales. Such disparities in scale among the output features can potentially impact the training and predictive capabilities of models (Singh and Singh, 2020). To mitigate this issue, pre-processing of output variables was performed. To standardize the scale of the output features, the *MinMax* scaler from scikit-learn was used. This scaler transforms each feature to a common scale (ranging between 0 and 1 in the present study). This step effectively normalizes the output features, bringing them onto the same scale and eliminating any potential biases arising from differences in scale during the training of models.

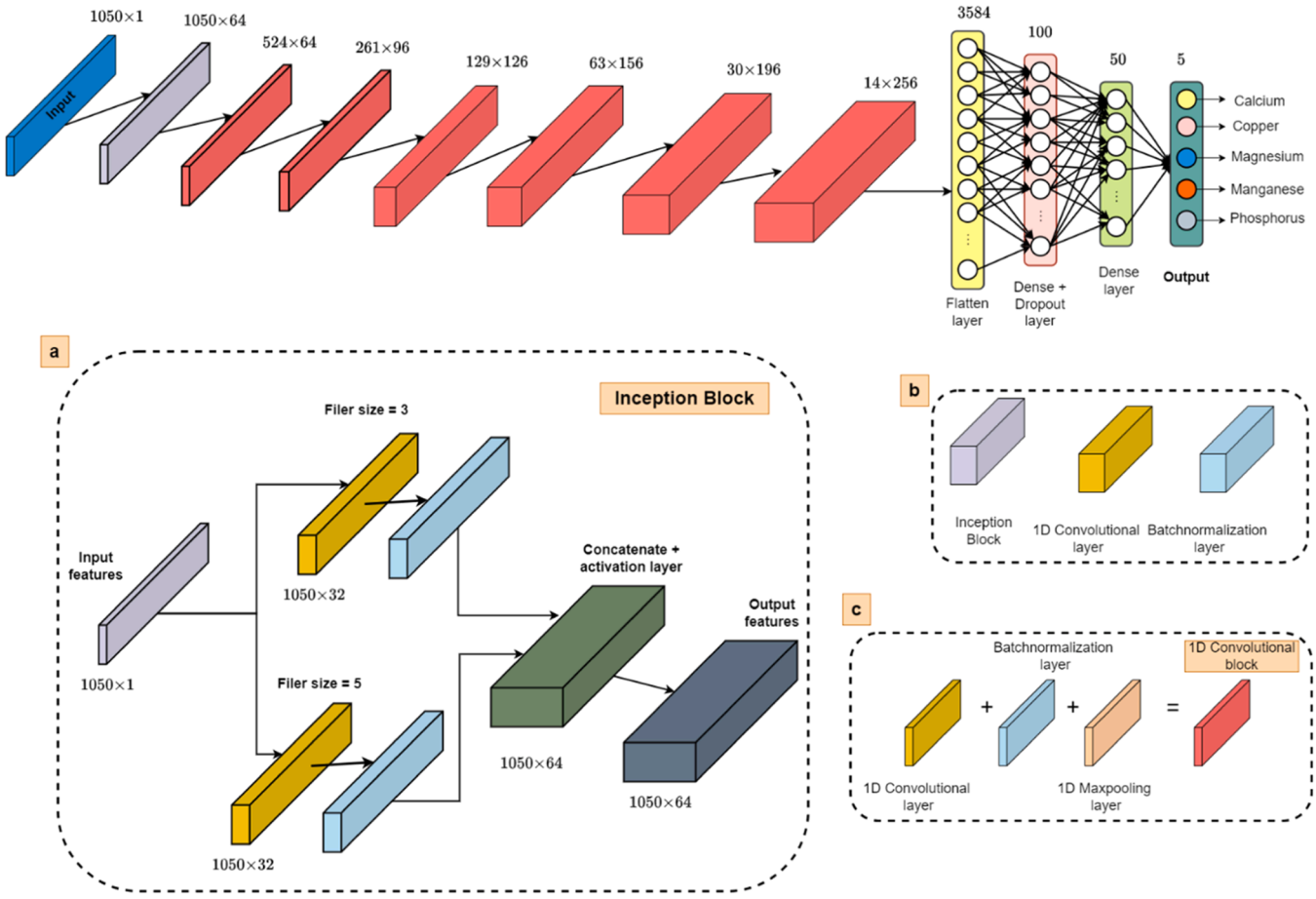


Fig. 4. Architecture of the 1D Convolutional Neural Network (1D CNN) for predicting mineral content in Perilla seeds. (a) The inception module featuring two parallel 1D convolutional layers with varying filter sizes (1×3 and 1×5) captures both local and global spectral features. (b) 1D convolutional layer, batch normalization layer, and max-pooling layer used in architecture of 1d CNN model. (c) The architecture includes multiple convolutional blocks, each consisting of a 1D convolutional layer, batch normalization layer, and max-pooling layer, which progressively extract important features from the NIR spectra while reducing the data dimensionality. The final prediction head, composed of dense layers with ReLU activation and a dropout layer, transforms the extracted features into predicted mineral concentrations.

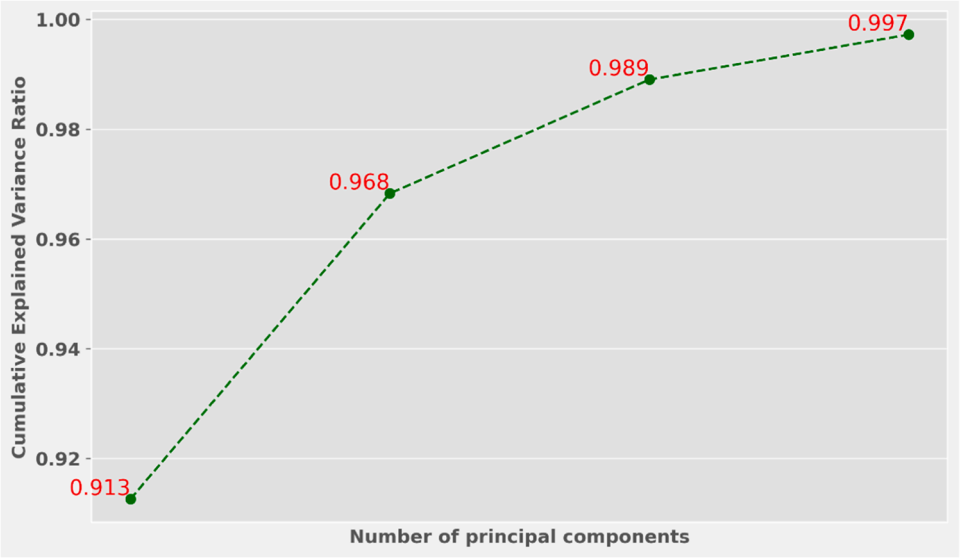


Fig. 5. Plot illustrating the cumulative explained variance by principal components (PCs) alongside the variance explained by each PC.

2.7. Training of deep-learning models

To train the constructed deep learning-based models (1D CNN and ANN models), the dataset was divided into training and testing sets, comprising 75.0 % and 25.0 % of the data, respectively. This allocation ensured that the testing data remained unseen during the training phase. To facilitate effective model training, a robust 5-fold cross-validation technique was adopted. This approach involved dividing the training dataset into five subsets, with each subset serving as a validation set while the model underwent training on the remaining four subsets (Sweet et al., 2023). This process was repeated five times, thus allowing the comprehensively assess the performance of the model across different data subsets and ensuring its generalizability. During the training phase, the root mean squared error (RMSE) metric was used as the loss function to quantify the disparity between predicted and reference values. Additionally, the mean absolute error (MAE) metric was utilized to measure the accuracy of the model in predicting mineral contents. The Adam optimizer (Kingma and Ba, 2017) with a learning rate of 0.0001 was used to update the model weights during training over 250 epochs with a batch size of seven instances. K-Fold Cross-Validation identified the best-performing model from the five folds. This model was then further refined through fine-tuning using the entire training dataset. This fine-tuning allowed the model to capture more complex patterns present across the entire data, thus, enhancing its ability to predict mineral content values with better accuracy.

2.8. Partial Least Squares Regression (PLSR) model

Partial Least Squares Regression (PLSR) provides efficient handling of multicollinearity and high-dimensional data while minimizing the risk of overfitting, making it suitable for small sample sizes (Farahani et al., 2010; Khatri et al., 2021), thus used by numerous researchers for predicting quantity through spectral data (Wold et al., 2001). In the present study, in addition to ML and DL, the PLSR model was also investigated for comprehensive evaluation. The PLSR model was implemented using the Scikit-learn library (Pedregosa et al., 2011), and a range of numbers of components was tested to determine the optimal number of components to be used in PSLR compilation. As depicted in Fig. 6 and Fig. 7, the Mean Squared Error (MSE) was observed to be lowest and the Residual Prediction Deviation (RPD) highest when using five numbers of components. Therefore, the PLSR model was compiled with five components while keeping the other hyperparameters at their default values.

2.9. Performance evaluation of the developed models

The developed models (1D CNNs, ANN, RFR, and SVR) were compared to evaluate their effectiveness in predicting mineral content in perilla. A comprehensive set of performance metrics was used to assess their predictive accuracy. The coefficient of Determination (R^2) metric (Eq. (1)) was used to indicate the proportion of variance in predicted mineral values explained by the model. A higher R^2 value (closer to 1.0) signifies a better fit between the actual and predicted data, where 1.0 represents a perfect fit (Bucchianico, 2007).

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

Where n is the number of samples, y_i represents actual mineral values, $\bar{y}$ is the mean of actual mineral values, and $\hat{y}_i$ is the predicted mineral values.

Additionally, Root Mean Squared Error (RMSE) (Eq. (2)) and Mean Absolute Error (MAE) (Eq. (3)) metrics were utilized to assess the performance (Hodson, 2022). RMSE measures the square root of the average of the squared differences between the actual and predicted values, while MAE quantifies the average absolute difference. Lower MSE and MAE values signify superior model performance, indicating smaller deviations between predicted and observed mineral values.

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

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \ldots \quad (3)$$

Where n is the number of samples, y_i represents actual mineral values, and $\hat{y}_i$ is the predicted mineral values.

The Residual Prediction Deviation (RPD) metric (Eq. (4)) provides a robust evaluation of the model by considering both the variability in

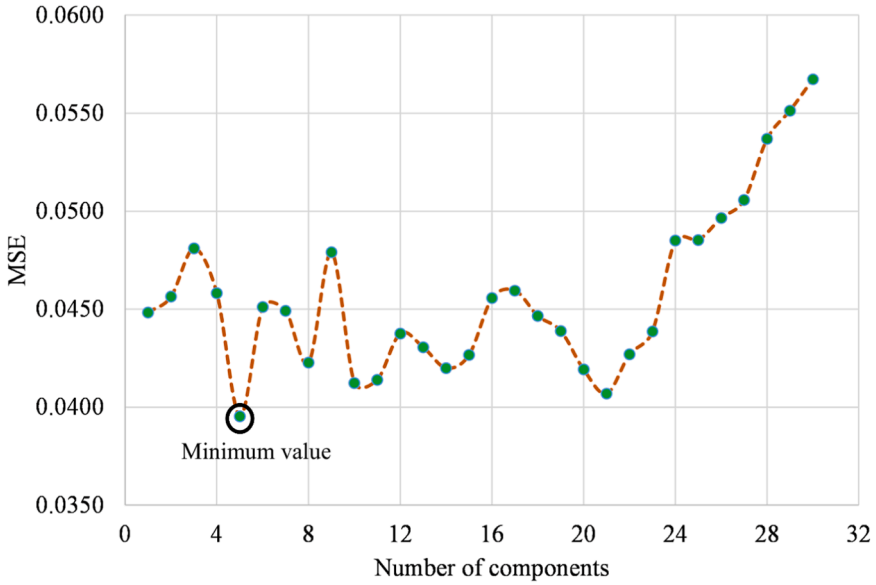


Fig. 6. Line plot illustrating the Mean Squared Error (MSE) values achieved with varying numbers of components in Partial Least Squares Regression (PLSR) model. The black circle highlights the minimum MSE value obtained with five components.

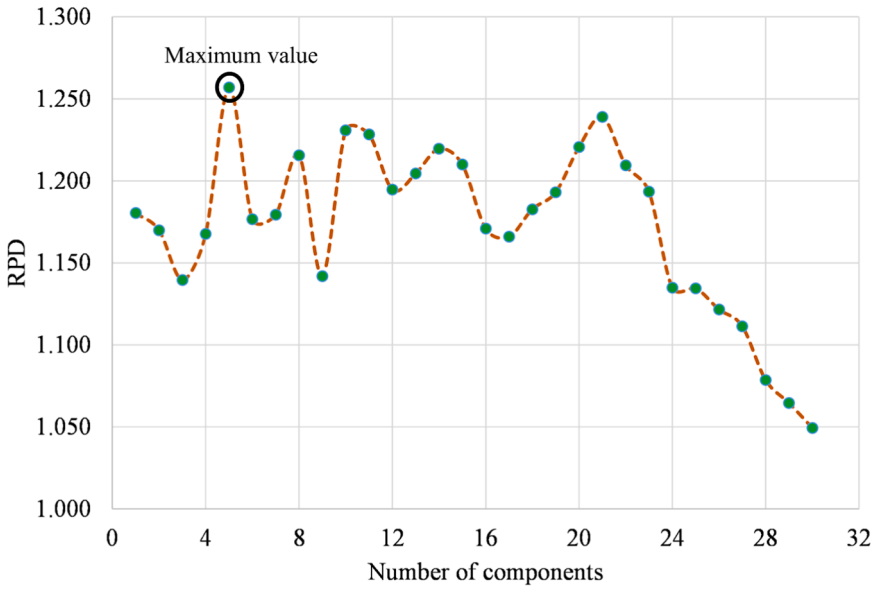


Fig. 7. Line plot illustrating the Residual Prediction Deviation (RPD) values achieved with varying numbers of components in the Partial Least Squares Regression (PLSR) model. The black circle highlights the maximum RPD value obtained with five components.

observed data and the prediction error (Viscarra Rossel et al., 2006). Thus, developed models were evaluated using the RPD metric as well. Higher RPD values indicate better predictive performance by the model.

$$RPD = \frac{Standard Deviation (SD) of\ Observed\ Values}{Root\ mean\ squared\ error} \ldots \quad (4)$$

2.10. Statistical analysis

The prediction accuracy of the developed models was assessed through a paired t-test conducted at a 95.0 % confidence interval using IBM SPSS Statistics 21.0, and the results were presented in the form of the standard error of the mean, standard deviation (SD), t-value, and p-values.

3. Results and discussion

3.1. Minerals profiling and NIR-spectra analysis of perilla germplasm

In the present investigation, ICP-OES was used to estimate the concentration of the five minerals in diverse perilla germplasm of 88 samples. As shown in Fig. 8, the mineral composition of perilla seeds showed significant variability, suggesting its potential for diverse nutritional and agricultural applications. Calcium is vital for bone and teeth formation, muscle contraction, nerve transmission, and blood clotting (González-Montaña et al., 2020). Significant variability was observed for calcium levels ranging from a minimum of 4509 ppm to a maximum of 6399 ppm, with a mean concentration of 5416 ppm. Copper acts as a cofactor in enzymes such as cytochrome c oxidase, vital for cellular respiration, and superoxide dismutase, essential for antioxidant defense against free radicals. It also facilitates reactions in enzymes like tyrosinase, involved in melanin production, and lysyl oxidase, vital for connective tissue formation (An et al., 2022; Tang et al., 2023). Copper content exhibited variability as well, with concentrations ranging from 11 ppm to 19 ppm, and a mean of 15 ppm. Magnesium (Mg) participates in hundreds of enzymatic reactions, muscle and nerve function, bone health, and cardiovascular regulation pressure (Long and Romani, 2014). Magnesium content showed a wide range from 2252 ppm to 3243 ppm, with a mean concentration of 2792 ppm. Manganese is crucial for metabolic processes as it serves as a cofactor for enzymes involved in carbohydrate, amino acid, and cholesterol metabolism, supporting energy production and overall cellular function. Manganese levels displayed considerable variability, ranging from 38 ppm to 67 ppm, and a mean concentration of 49 ppm. Phosphorus (P) is integral to DNA and RNA formation, energy production, bone health, and cell signaling (Serna and Bergwitz, 2020). Phosphorus content exhibited significant variability, with minimum, maximum, and mean concentrations of 3047 ppm, 7166 ppm, and 6006 ppm, respectively. These findings highlight the significant variability in mineral composition in perilla germplasm, indicating its potential for various nutritional and agricultural applications.

The collective near-infrared (NIR) spectra spanning the wavelength range of 400–2500 nm for seeds sourced from a diverse perilla germplasm collection in the NEH region of India are presented in Fig. 2. During NIR-spectra acquisition, a completely randomized design was used to ensure that the spectral acquisition and evaluation of minerals content was randomly assigned to each sample unit, with every sample having an equal probability of receiving any treatment. We observed that distinguishing NIR regions visually becomes difficult due to the prevalence of highly overlapping and broad combination bands arising from fundamental vibrations (Cozzolino, 2015). Additionally, understanding matrix effects, particularly in biological materials like perilla seeds, is crucial for ensuring model applicability across diverse matrices. Furthermore, it has been observed that variations in sample composition, moisture content, and particle size can significantly impact prediction accuracy, requiring a uniform seed sample size. Therefore, each sample was scanned twice to ensure the check of the spectral error, if any, and the subsequent data analysis utilized the average spectrum of each sample.

The rationale behind selecting specific five minerals including Calcium, Copper, Magnesium, Manganese, and Phosphorus, for our study, lies in their key roles in human health, their known abundance in perilla seeds (Dhyani et al., 2019), and their suitability for accurate and robust modeling. As mentioned in the previous paragraph, calcium is vital for bone health and various cellular processes, while Magnesium is essential for muscle and nerve function, blood sugar control, and bone development. Phosphorus is crucial for the formation of bones and teeth and plays a role in energy production and storage (Palacios et al., 2021; Serna and Bergwitz, 2020). Copper and Manganese, although required in smaller amounts, are indispensable trace minerals. Copper is necessary for red blood cell formation, nerve health, and immune function, (An et al., 2022), while Manganese is involved in bone formation,

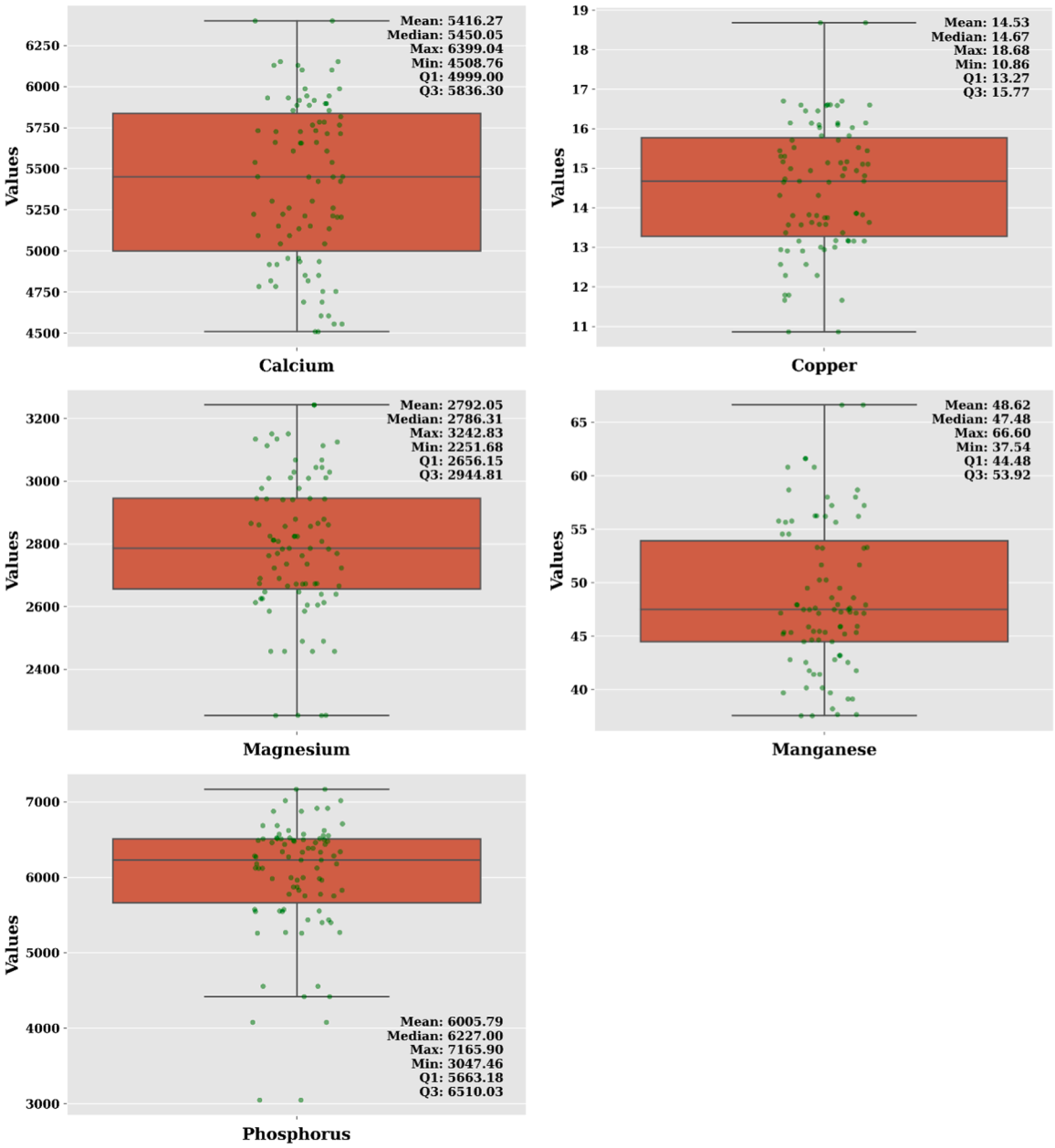


Fig. 8. Box plot depicting the statistical description of the minerals content (in ppm) in perilla germplasm.

metabolism, and antioxidant defense. These minerals are present in higher abundance compared to other trace elements in perilla seeds, which enhances the accuracy and robustness of our models. Understanding the relationships between the calibrated trait and its associated wavelength (even indirectly with bound compounds) can become complex due to an overlap between the NIR band vibrations associated with different traits. The presence of a trait in extremely low concentration can be a limiting factor, as these traits usually display low wavelength regression coefficients and NIR absorption band numbers (Tomar et al., 2021b). Our results indicate that the prediction accuracy of a model is low when a trait is present in lesser amounts. By focusing on these five abundant minerals, we developed NIRS-based prediction models that are both reliable and efficient, facilitating the rapid screening of perilla germplasm to select mineral-rich genotypes and mitigate micronutrient deficiencies.

3.2. *Performance of the developed models*

An independent test dataset, comprising 25.0 % of total data was used for the evaluation of developed models. The estimation of the five minerals for this set was also conducted using the same method (i.e., ICP- OES) as performed in the calibration set. Statistical metrics: Coefficient of Determination (R^2), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Residual Prediction deviation (RPD) were used to quantify the performance of the developed models.

The performance of different models in predicting mineral content varied across all the minerals. For calcium, as shown in Fig. 9, the 1D CNN model achieved an R^2 value of 0.69, indicating good predictive accuracy, while SVR and RFR models had lower R^2 values of 0.51 and 0.50, respectively. The performance of the ANN model lies between 1D CNN and SVR models in terms of predictive accuracy, with an R^2 value of 0.57. In the case of copper (Fig. 10), the SVR model showed the best performance with an R^2 value of 0.78, followed by 1D CNN and RFR models with R^2 values of 0.64 each. The ANN model exhibited the lowest

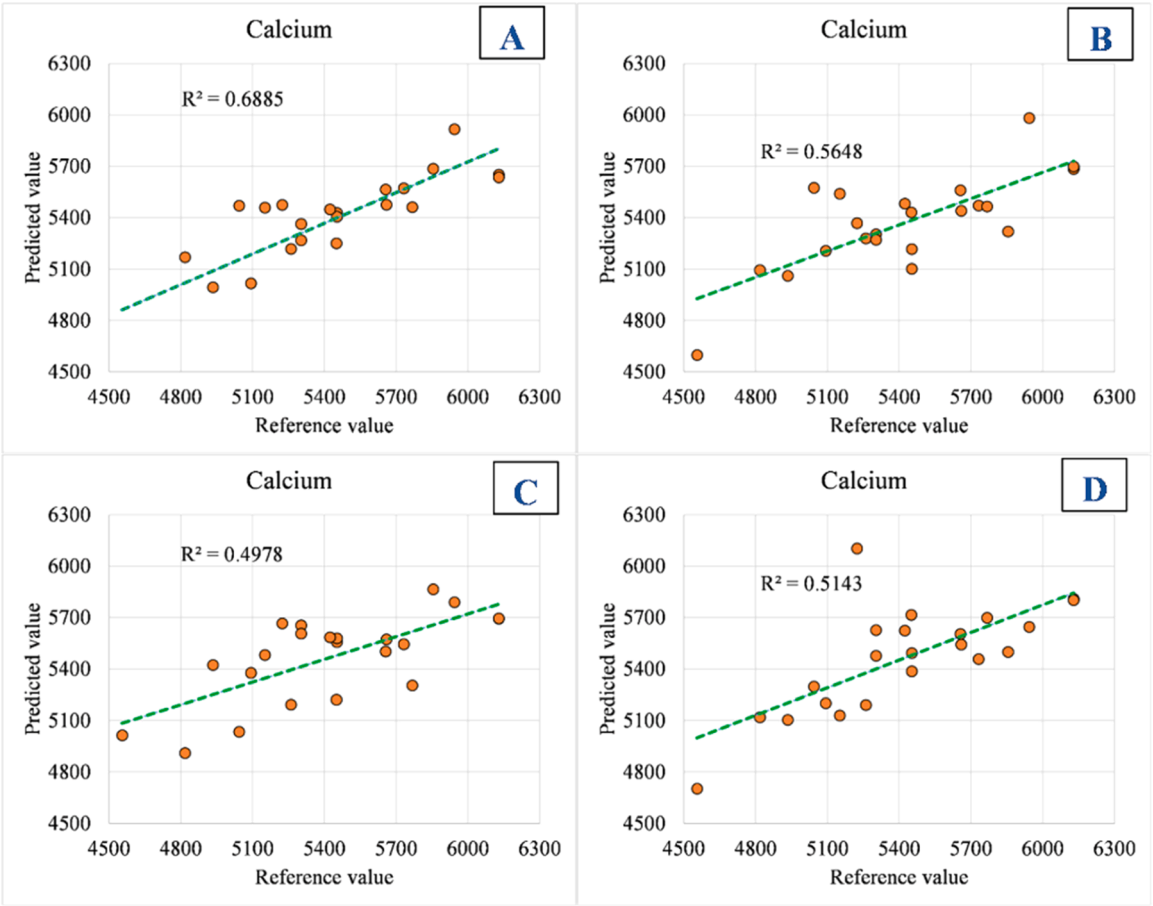


Fig. 9. Scatter plots of reference and predicted values for calcium content (in ppm) in perilla seeds. Each graph represents a linear equation with the coefficient of determination (R^2) using NIRS-based (A) 1D CNN (Convolutional neural network); (B) ANN (Artificial Neural Network); (C) RFR: Random Forest Regressor; (D) SVR: Support Vector Regressor models.

predictive accuracy for copper, with an R^2 value of 0.36. For magnesium, as shown in Fig. 11, the 1D CNN model achieved the highest R^2 value of 0.71, followed by SVR (R^2 =0.61), RFR (R^2 = 0.39), and ANN (R^2 = 0.53) models. For manganese (Fig. 12), SVR had the highest predictive accuracy with an R^2 value of 0.85, followed by RFR (R^2 = 0.72), 1D CNN (R^2 = 0.67), and ANN (R^2 = 0.56) models. As shown in Fig. 13, for phosphorus, the 1D CNN model demonstrated the highest R^2 value of 0.88, followed closely by the RFR (R^2 = 0.86) and SVR (R^2 = 0.83) models. The ANN model had a lower R^2 value of 0.68 for phosphorus prediction. Fig. 14 illustrates a bar plot comparing the Coefficient of determination (R^2) values for the models.

The performance of each developed model was evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) metrics, as summarized in Table 1. For calcium prediction, the 1D CNN model achieved the lowest MAE of 177.09, outperforming SVR (MAE = 219.64) and ANN (MAE = 211.93) models, while RFR had the highest MAE (244.27). For copper prediction, SVR exhibited the lowest MAE of 0.78, closely followed by 1D CNN (MAE = 0.86), while RFR had an MAE of 0.92 and ANN showed the highest MAE (1.34). For magnesium, 1D CNN had the lowest MAE of 130.21, followed by SVR (MAE = 147.59) and ANN (MAE = 159.30), with RFR displaying the highest MAE (175.29). In predicting manganese content, SVR showed the lowest MAE of 2.94, followed by RFR (MAE = 3.66), 1D CNN (MAE = 3.96), and ANN with the highest MAE (4.45). For phosphorus prediction, RFR demonstrated the lowest MAE of 259.14, followed by 1D CNN (MAE = 285.60), while SVR had an MAE of 366.20 and ANN exhibited the highest MAE (484.96).

In terms of RMSE performance, as shown in Table 1, the 1D CNN model shows the lowest RMSE for calcium prediction, with a value of 232.66, followed by ANN (RMSE = 272.29). RFR exhibited an RMSE of 290.00, while SVR had the highest RMSE at 283.50. For copper prediction, SVR achieved the lowest RMSE of 0.96, closely followed by 1D CNN (RMSE = 1.15). RFR had an RMSE of 1.20, while ANN displayed the highest RMSE at 1.60. In terms of magnesium prediction, 1D CNN showed the lowest RMSE of 148.74, followed by RFR with an RMSE of 211.50. SVR had an RMSE of 170.98, while ANN exhibited the highest RMSE at 186.26. SVR also demonstrated the lowest RMSE for manganese prediction, with a value of 3.52, followed by RFR at 4.37. The RMSE for 1D CNN was 5.01, while ANN had the highest RMSE at 5.77. For phosphorus prediction, 1D CNN showed the lowest RMSE of 333.86, followed by RFR with an RMSE of 365.46. SVR had an RMSE of 453.42, while ANN exhibited the highest RMSE at 554.16.

The evaluation of different models reveals variations in performance across the five minerals. The 1D CNN model consistently demonstrates superior performance in terms of both R^2 values and error metrics such as MAE and RMSE for three minerals (Calcium, Magnesium, and Phoshphorus). This indicates that the 1D CNN model is more effective in capturing the underlying patterns in the spectral data, resulting in more accurate predictions compared to other models. However, the SVR model outperforms the 1D CNN model for Copper and Manganese prediction. Additionally, the RFR model shows competitive performance for phosphorus prediction in terms of both MAE and RMSE.

As shown in Fig. 15, a comparison of model performance using Residual Prediction Deviation (RPD) revealed variations across different minerals. For calcium, the 1D CNN and ANN models showed moderate predictive performance with RPD values of 1.75 and 1.50, respectively,

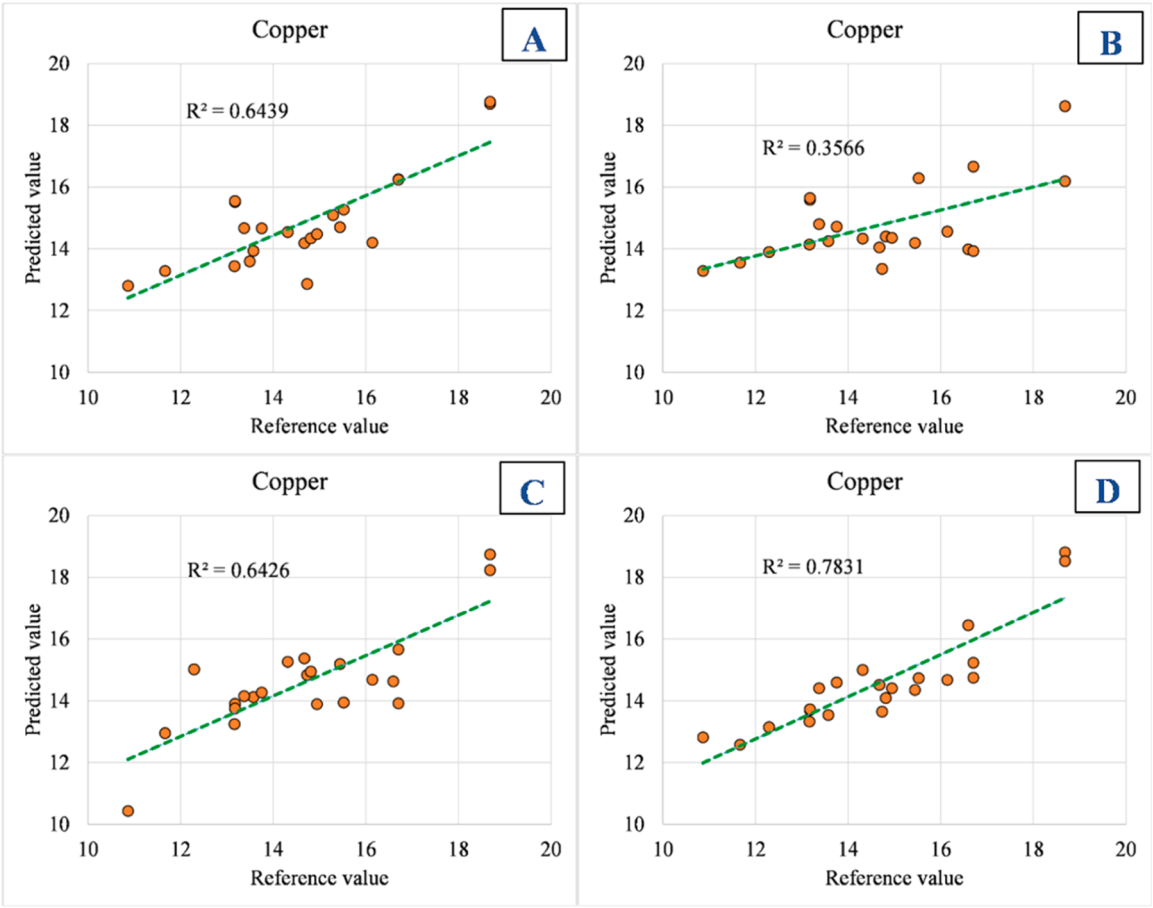


Fig. 10. Scatter plots of reference and predicted values for copper content (in ppm) in perilla seeds. Each graph represents a linear equation with the coefficient of determination (R^2) using NIRS-based (A) 1D CNN (Convolutional neural network); (B) ANN (Artificial Neural Network); (C) RFR: Random Forest Regressor; (D) SVR: Support Vector Regressor models.

while the SVR and RFR models demonstrated slightly lower RPD values of 1.44 and 1.41, respectively. In predicting copper content, the SVR model performed well with an RPD of 1.82, followed closely by the 1D CNN model with an RPD of 1.49. For magnesium, the 1D CNN model exhibited the highest RPD value of 1.83, indicating relatively high predictive accuracy, while the SVR and ANN models also performed well with RPD values of 1.60 and 1.47, respectively. In the case of manganese, the SVR model yielded the highest RPD of 2.20, suggesting relatively high predictive accuracy, followed by the RFR model with an RPD of 1.79 and the 1D CNN model with an RPD of 1.58. Finally, for phosphorus, the 1D CNN model demonstrated excellent predictive accuracy with the highest RPD value of 2.96, followed by the RFR model with an RPD of 2.71, and the SVR model with an RPD of 2.18.

These results highlight the superior predictive potential of the 1D CNN and SVR models over ANN and RFR for the estimation of the mineral composition of perilla seeds. Extracting the complex relationships between mineral composition and their associated wavelengths can be challenging, primarily due to the overlap in NIR band vibrations. An additional complexity arises when a trait is present in extremely low concentrations. These particular traits typically exhibit regression coefficients and NIR absorption band numbers at lower wavelengths. Our findings suggest that 1D CNN yields superior predictions, demonstrating lower MAE and RMSE and higher RPD in predicting the minerals that are present in large amounts for instance, Calcium (average = 5416 ppm), Magnesium (average = 2792 ppm), and Phosphorus (average = 6006 ppm). While SVR performed better in predicting the minerals that are present in lesser amounts including Copper (average = 15 ppm) and Manganese (average = 49 ppm).

3.3. Comparative assessment of developed models

The comparative assessment of 1D CNN, ANN, RFR, and SVR models for predicting mineral contents in perilla seeds shows variations in performances across minerals. Fig. 14 and Fig. 15 provide a comprehensive comparative analysis of the performance of models in terms of both R^2 and RPD metrics for each mineral. For calcium estimation, the 1D CNN model was found to be the best performer, yielding the highest R^2 (0.69), RPD (1.75), and the lowest MAE (177.09) and RMSE (232.66) values, indicative of its superior predictive accuracy. Similarly, in the case of magnesium and phosphorus, the 1D CNN model consistently shows superior performance across all metrics, verifying its robustness in predicting mineral content. However, for copper estimation, the SVR model displayed the highest R^2 (0.78) and RPD (1.82) values, along with the lowest MAE (0.78) and RMSE (0.96), suggesting better prediction by SVR compared to other models. Similarly, in predicting manganese content, the SVR model outperformed other models in terms of R^2 (0.85) and RPD (2.20), as well as having the lowest MAE (2.94) and RMSE (3.52). In summary, while both 1D CNN and SVR models exhibited strong predictive capabilities, the 1D CNN performed better in predicting Calcium, Magnesium, and Phosphorus, whereas, for Copper and Manganese, the SVR model was found to be more effective, thus, highlighting the importance of selecting models based on the specific mineral.

The superior performance of the 1D CNN model compared to SVR, RFR, and ANN can be attributed to several factors. While SVR, RFR, and ANN models were trained on principal components (PCs) extracted from the spectral data, the 1D CNN model directly utilized raw spectral data

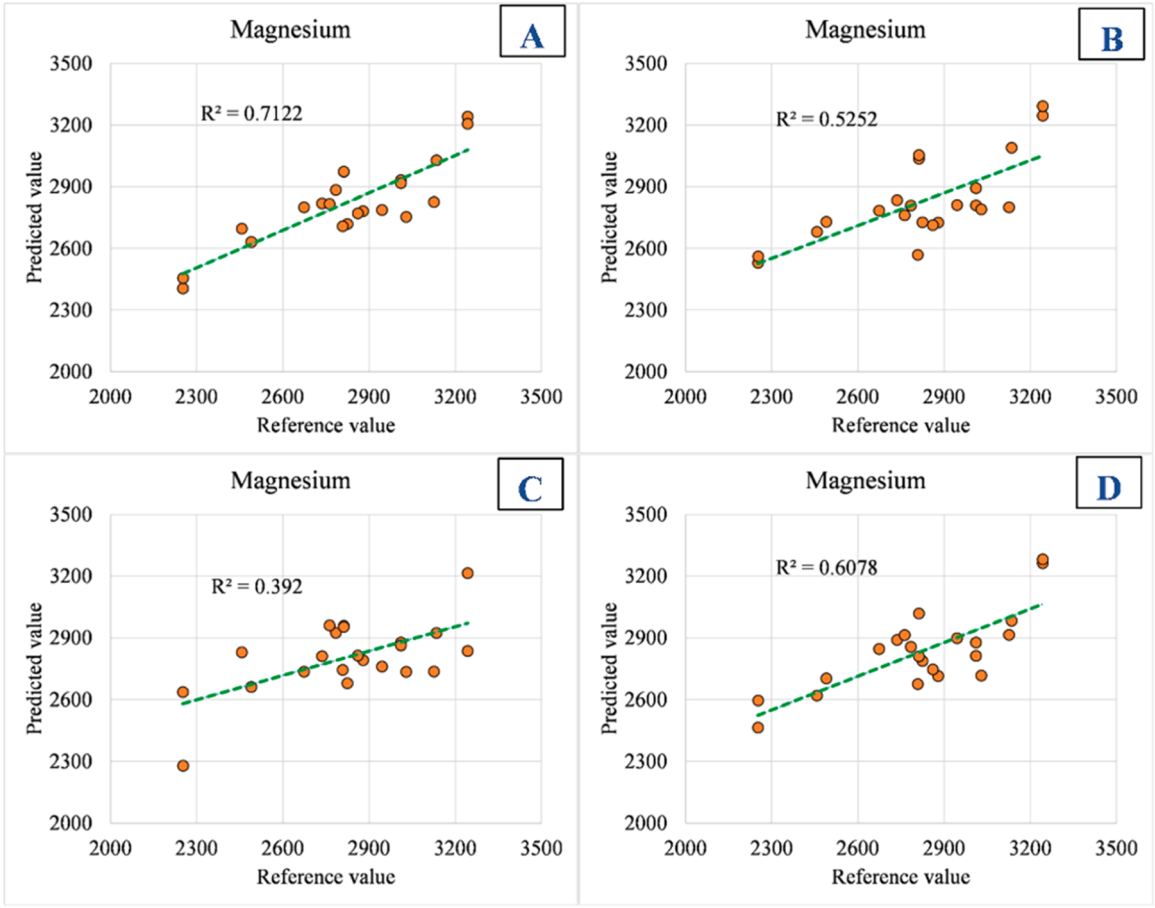


Fig. 11. Scatter plots of reference and predicted values for magnesium content (in ppm) in perilla seeds. Each graph represents a linear equation with the coefficient of determination (R^2) using NIRS-based (A) 1D CNN (Convolutional neural network); (B) ANN (Artificial Neural Network); (C) RFR: Random Forest Regressor; (D) SVR: Support Vector Regressor models.

as input. This difference in input data is important because the use of PCs may lead to information loss (Howley et al., 2006), thereby limiting the ability of a model to capture the complexities of the spectral patterns associated with mineral content. In contrast, the 1D CNN model can effectively extract both local and global information from the raw spectra (Kiranyaz et al., 2021), enabling it to capture a more comprehensive representation of the spectral data. Additionally, the architecture of the 1D CNN model is suitable for sequential data analysis (Malek et al., 2018; Zeng et al., 2021), allowing it to apply convolutional filters across the sequential data (spectral data in the present study), thus, facilitating the identification and extraction of relevant features related to each mineral. This capability of the 1D CNN model is advantageous for spectral data analysis which is sequential. In contrast, SVR, RFR, and ANN models are not ideally suited for sequential data analysis due to their lack of internal memory or inability to process spatial information (Kiranyaz et al., 2021; Sherstinsky, 2020). This limitation may lead to suboptimal performances by SVR, RFR, and ANN models when used with sequential data. Overall, the ability of the 1D CNN model to utilize raw spectral data, extract complex features, and effectively analyze sequential data contributes to its superior performance compared to SVR, RFR, and ANN models in predicting mineral content from spectral data.

However, the SVR model predicts Copper and Manganese contents in perilla accurately. In present study, as can be observed from Fig. 8, Copper and Manganese have significantly lower concentration ranges (11–19 ppm for Copper and 38–67 ppm for Manganese) compared to Calcium (4509–6399 ppm), Magnesium (2252–3243 ppm), and Phosphorus (3047–7166 ppm). Lower concentrations can lead to subtler spectral signatures that may require a different modeling approach. SVR might be more adept at handling these subtle variations. Moreover, the presence of outliers in the copper and manganese data, as illustrated in Fig. 8, could have influenced the performance difference between SVR and 1D CNN. The ability of SVR in handling outliers effectively may have been advantageous in capturing the variability in Copper and Manganese levels (Mohammed Rashid et al., 2022; Nishiguchi et al., 2009; Wang and Li, 2017). Outliers can significantly impact model performance (Uzun Ozsahin et al., 2022), and the ability of SVR to mitigate their influence may have led to more accurate predictions for these minerals.

3.4. Comparison of developed models with PLSR

For comprehensiveness, the performance of developed ML and DL-based models was compared with the PLSR model. From Table 2, a significant performance gap between the PLSR model and other models developed in the present study was observed. For instance, in terms of MAE, the PLSR model shows higher values across all minerals compared to the 1D CNN, SVR, and RFR models. For calcium, the MAE for PLSR is 279.85, whereas for 1D CNN, it is significantly lower at 177.09. Similarly, for copper, magnesium, manganese, and phosphorus, the MAE values for PLSR were found to be higher compared to other models. A similar pattern was observed in RMSE values as well, where PLSR shows larger values compared to other models, thus indicating greater prediction errors. For instance, for calcium, the RMSE for PLSR was 359.24, while for 1D CNN, it was lower at 232.66. Furthermore, the R^2 values for PLSR were found to be lower across minerals compared to other models,

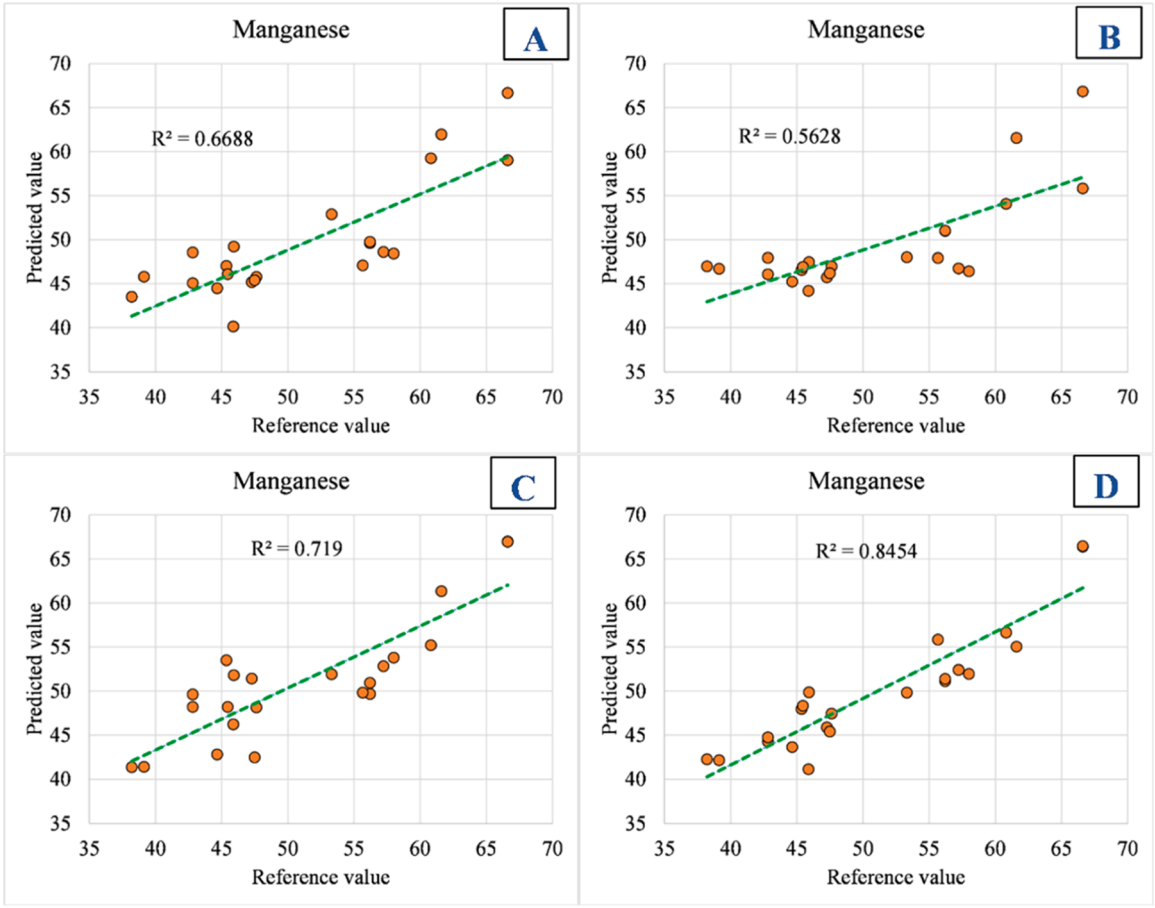


Fig. 12. Scatter plots of reference and predicted values for manganese content (in ppm) in perilla seeds. Each graph represents a linear equation with the coefficient of determination (R^2) using NIRS-based (A) 1D CNN (Convolutional neural network); (B) ANN (Artificial Neural Network); (C) RFR: Random Forest Regressor; (D) SVR: Support Vector Regressor models.

indicating weaker correlations between predicted and actual values. This performance gap between PLSR and other developed models in the present study shows that the PLSR model has limited predictive capability in estimating mineral content in Perilla seeds.

Although NIR spectroscopy primarily detects molecular vibrations in bonds like O-H, N-H, and C-H, minerals indirectly affect the NIR spectra through associations with organic matter, hydration states, and matrix effects. Advanced ML and DL models can learn complex, non-linear relationships in these spectra, enabling accurate mineral content predictions. A recent study on rice bean and adzuki bean validated this approach using MPLS, demonstrating correlations between NIR spectra and the mineral content of iron (R^2= 0.26), copper (R^2= 0.47), and zinc (R^2= 0.45), despite the indirect detection (John et al., 2023). In agreement with this study, our results also indicate that while MPLS provides low accuracy for mineral prediction, using advanced feature selection algorithms like 1D CNN and SVR can significantly improve the accuracy and performance evaluation of the models.

In the present study, high accuracy was achieved for phosphorus and moderate accuracy for four other minerals, although no combined model for these minerals has been reported so far. The MPLS method gave lower R^2 values compared to 1D CNN and SVR models. One reason for the challenge in developing calibrations for microelements is the poor absorption of energy by minerals in the NIR region, as they are often bound to organic molecules rather than in a free form, resulting in poor absorption. In the future, robust models can be developed using vast datasets and advanced techniques like deep learning and hybrid algorithms to enhance the accuracy and reliability of NIR-based mineral predictions.

3.5. Statistical analysis of the developed models

To determine the comparability of mean values of a dependent variable with both reference and predicted values for estimated biochemical parameters, a paired t-test with a 95 % confidence interval was executed. The resulting p-value exceeded the significance threshold of 0.05, suggesting the robust accuracy and reliability of the models, as indicated in Table 2. Specifically, for 1D CNN, the p-values were 0.40, 0.44, 0.99, 0.14, and 0.77 for Calcium, Copper, Magnesium, Manganese, and Phoshphorus, respectively. For SVR, the p-values were 0.51, 0.68, 0.76, 0.15, and 0.52 for Calcium, Copper, Magnesium, Manganese, and Phoshphorus respectively (Table 3). These findings highlight a lack of statistically significant differences in means between the NIRS-based prediction method and the different algorithms used in the present study.

4. Limitations and future scope

The present study presents significant advancements in using NIRS coupled with machine learning and deep learning-based models for mineral prediction in Perilla seeds. While the results are promising, there are opportunities for further enhancement and broader application of the developed models. One potential area for future research is the exploration of hybrid deep learning models, such as 1D CNNs combined with Long Short-Term Memory (LSTM) layers or attention mechanisms, which could improve the ability of deep learning-based models to capture both local and sequential dependencies in the spectral data. Additionally, using Transformer-based models, which have shown success in

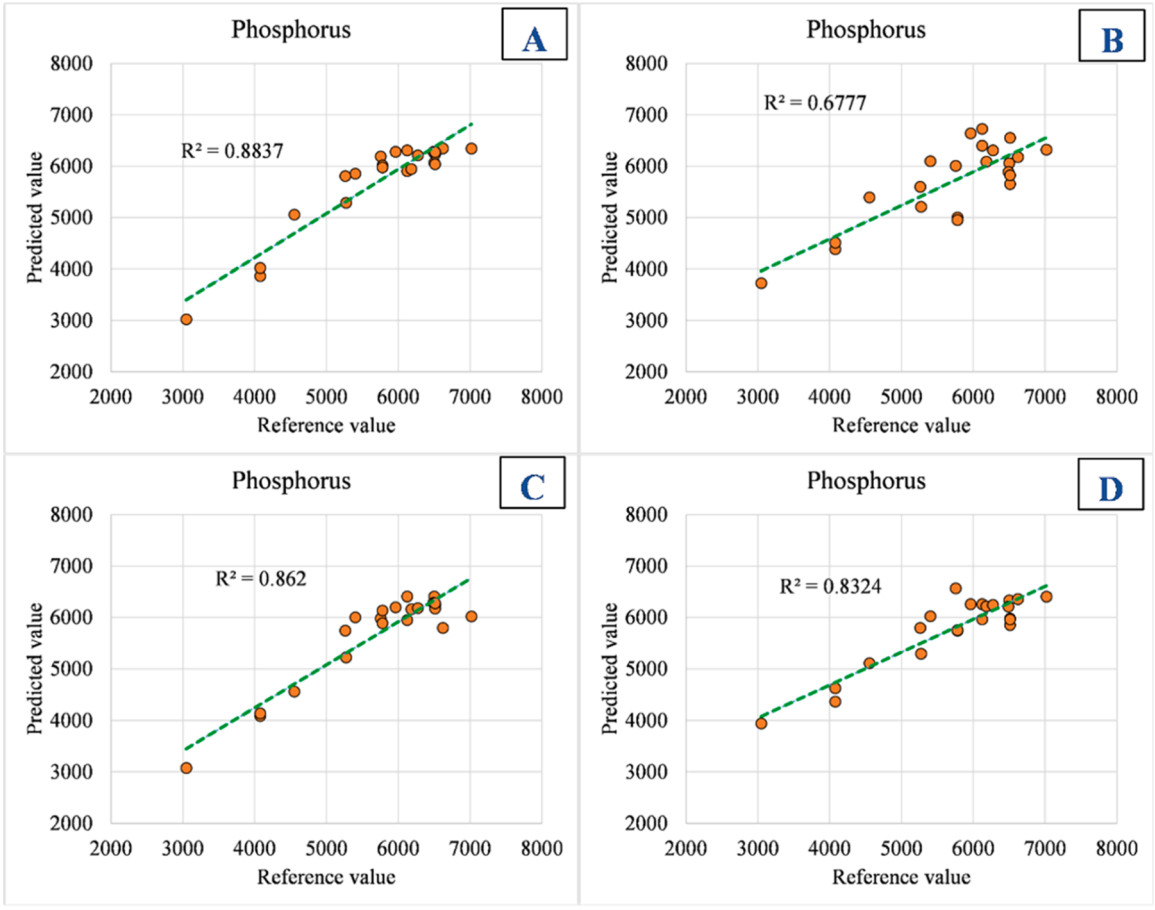


Fig. 13. Scatter plots of reference and predicted values for phosphorus content (in ppm) in perilla seeds. Each graph represents a linear equation with the coefficient of determination (R^2) using NIRS-based (A) 1D CNN (Convolutional neural network); (B) ANN (Artificial Neural Network); (C) RFR: Random Forest Regressor; (D) SVR: Support Vector Regressor models.

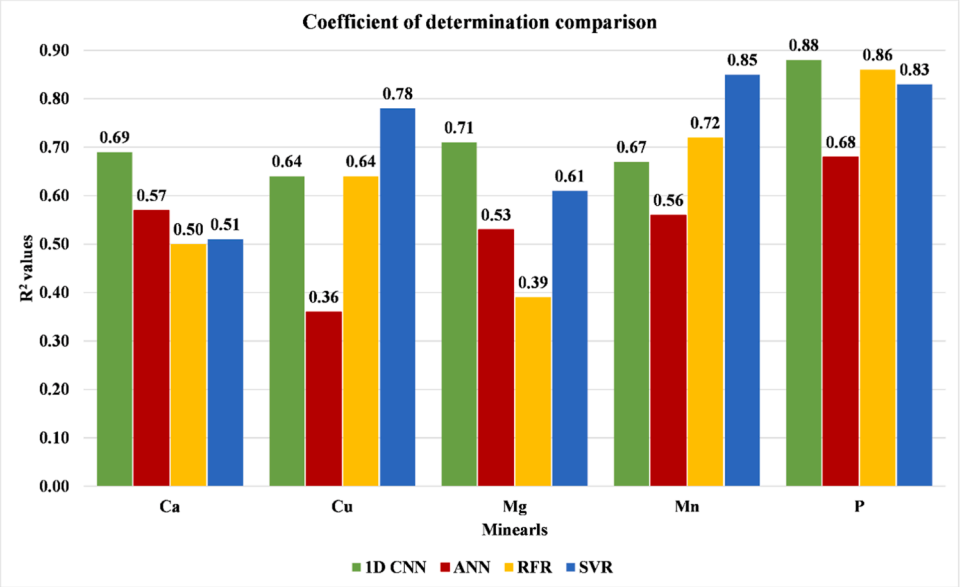


Fig. 14. Bar plot comparing the Coefficient of determination (R^2) values of 5 minerals among 1D CNN (Convolutional neural network), ANN (Artificial Neural Network), RFR: Random Forest Regressor, and SVR: Support Vector Regressor.

other sequential data tasks, could further enhance prediction accuracy by focusing on key spectral features. Another promising direction is the application of wavelength selection techniques to identify key wavelengths that contribute most to the prediction of mineral content. This could reduce computational complexity and improve model interpretability while maintaining or even enhancing predictive

Table 1
MAE and RMSE metrics of 5 minerals utilizing 1D CNN, ANN, RFR, and SVR models.

Models	MAE					RMSE				
	Calcium	Copper	Magnesium	Manganese	Phoshphorus	Calcium	Copper	Magnesium	Manganese	Phoshphorus
1D CNN	177.09	0.86	130.21	3.96	285.60	232.66	1.15	148.74	5.01	333.86
ANN	211.93	1.34	159.30	4.45	484.96	272.29	1.60	186.26	5.77	554.16
RFR	244.27	0.92	175.29	3.66	259.14	290.00	1.20	211.50	4.37	365.46
SVR	219.64	0.78	147.59	2.94	366.20	283.50	0.96	170.98	3.52	453.42

*MAE: Mean Absolute Error; RMSE: Root Mean Squared Error; 1D CNN: Convolutional neural network; ANN: Artificial Neural Network; RFR: Random Forest Regressor; SVR: Support Vector Regressor.

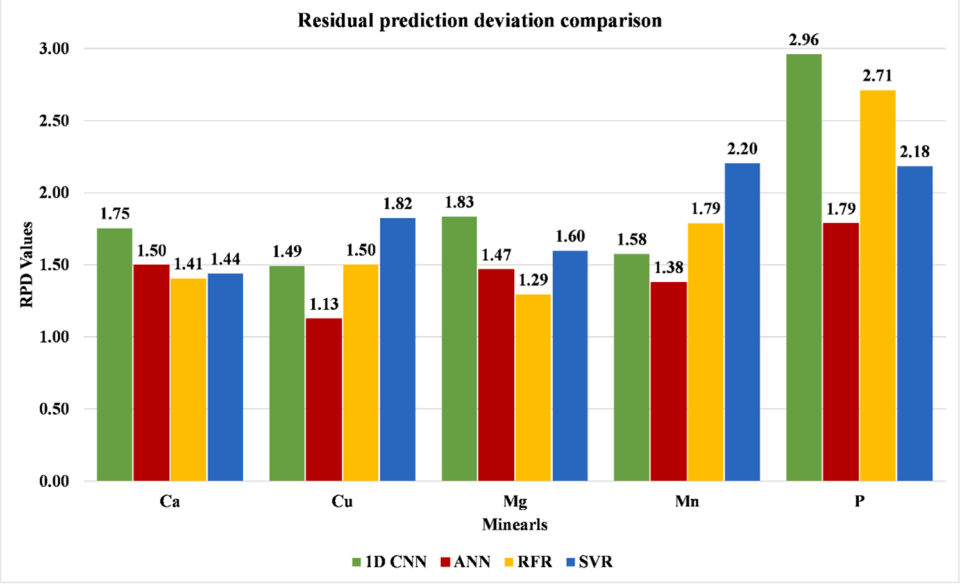


Fig. 15. Bar plot comparing the Residual Prediction Deviation (RPD) values of 5 minerals among 1D CNN (Convolutional neural network), ANN (Artificial Neural Network), RFR: Random Forest Regressor, and SVR: Support Vector Regressor.

Table 2
Evaluation metrics attained for minerals prediction in perilla using Partial Least Squares Regression model.

	MAE	RMSE	R^2
Calcium	279.85	359.24	0.20
Copper	1.47	1.75	0.23
Magnesium	203.81	251.84	0.13
Manganese	5.77	6.92	0.29
Phosphorus	593.02	804.07	0.32

performance. Future work could also explore the use of transfer learning, where models trained on one crop can be fine-tuned for application to different crops. This would significantly broaden the utility of these models and enable rapid screening across diverse agricultural contexts.

5. Conclusions

This study presents a novel application of NIRS-based prediction modeling combined with machine learning (ML) and deep learning (DL) techniques, offering a fast, reliable, and efficient approach for screening large germplasm of perilla to accurately estimate essential minerals. By employing 1D CNNs, ANNs, RFR, and SVR, models were developed and validated to profile key five minerals (Calcium, Copper, Magnesium, Manganese, and Phosphorus) in perilla seeds. The adaptability of ML and DL techniques in predicting mineral content from NIR-spectral data was demonstrated, with the 1D CNN model outperforming others for Calcium, Magnesium, and Phosphorus prediction, achieving RPD values of 1.75, 1.83, and 2.96, respectively. Similarly, the SVR model performed better for Copper and Manganese, attaining RPD values of 1.82 and 2.2. The 1D CNN model also showed high R^2 values (>0.65) for all minerals, with a maximum of 0.88 for Phosphorus, compared to the lower performance of the PLSR model (R^2 = 0.32). The superior performance of the 1D CNN model can be attributed to its ability to capture complex non-linear relationships within the spectral data, making it a powerful tool for rapid, non-destructive screening of perilla germplasm. These models streamline the identification of mineral-rich genotypes, which are crucial for addressing micronutrient deficiencies in both human health and the food industry. Moreover, the models assist in eliminating fewer desirable genotypes, enhancing efficiency in germplasm screening. This study highlight the broader applicability of ML and DL-based approaches in the food industry, presenting rapid techniques to predict essential mineral compositions, ultimately contributing to the development of nutrient-rich food products.

Funding

This work has been supported by the in-house project of first two authors. This work was also supported by the International collaborative project 'Consumption of Resilient Orphan Crops & Products for Healthier Diets' (CROPS4HD), which is co-funded by the Swiss Agency for Development and Cooperation, Global Programme Food Security (SDC GPFS), and executed in India through FiBL (Research Institute of Organic Agriculture), and Alliance of Bioversity and CIAT with ICAR-National Bureau of Plant Genetic Resources, New Delhi – 110 012 (India).

Table 3
Paired sample t-test at 95 % confidence interval.

Method	Paired differences						t-value	p-value
	Pair	Mean	SD	SEM	95 % confidence interval of the difference			
					Lower	Upper		
1D CNN	Calcium (Ref. vs Pred.)	42.60273	234.10580	49.91153	−61.19397	146.39943	.854	.403
	Copper (Ref. vs Pred.)	−.19607	1.16457	.24829	−.71241	.32027	−.790	.439
	Magnesium (Ref. vs Pred.)	.44455	152.23497	32.45660	−67.05266	67.94175	.014	.989
	Manganese (Ref. vs Pred.)	1.58735	4.86885	1.03804	−.57138	3.74608	1.529	.141
	Phosphorus (Ref. vs Pred.)	21.32591	341.01628	72.70492	−129.87224	172.52406	.293	.772
ANN	Calcium (Ref. vs Pred.)	54.03227	273.15485	58.23681	−67.07780	175.14235	.928	.364
	Copper (Ref. vs Pred.)	−.08662	1.63868	.34937	−.81317	.63993	−.248	.807
	Magnesium (Ref. vs Pred.)	−4.96500	190.57507	40.63074	−89.46125	79.53125	−.122	.904
	Manganese (Ref. vs Pred.)	1.73203	5.62842	1.19998	−.76347	4.22753	1.443	.164
	Phosphorus (Ref. vs Pred.)	12.47591	567.05813	120.89720	−238.94358	263.89540	.103	.919
RFR	Calcium (Ref. vs Pred.)	−42.61455	293.60216	62.59619	−172.79045	87.56136	−.681	.503
	Copper (Ref. vs Pred.)	.08043	1.22102	.26032	−.46094	.62180	.309	.760
	Magnesium (Ref. vs Pred.)	18.59318	215.63506	45.97355	−77.01405	114.20041	.404	.690
	Manganese (Ref. vs Pred.)	−.01359	4.47670	.95444	−1.99845	1.97127	−.014	.989
	Phosphorus (Ref. vs Pred.)	41.40182	371.65594	79.23731	−123.38119	206.18483	.523	.607
SVR	Calcium (Ref. vs Pred.)	−40.70955	287.15941	61.22259	−168.02890	86.60981	−.665	.513
	Copper (Ref. vs Pred.)	.08696	.97838	.20859	−.34682	.52075	.417	.681
	Magnesium (Ref. vs Pred.)	−11.41818	174.60952	37.22688	−88.83571	65.99934	−.307	.762
	Manganese (Ref. vs Pred.)	1.09816	3.42209	.72959	−.41911	2.61543	1.505	.147
	Phosphorus (Ref. vs Pred.)	−64.86864	459.31778	97.92688	−268.51873	138.78146	−.662	.515

*SD: Standard Deviation; SEM: Standard Error of Mean; Ref.- Reference values, Pred.- Predicted values; 1D CNN: Convolutional neural network; ANN: Artificial Neural Network; RFR: Random Forest Regressor; SVR: Support Vector Regressor.

Author statement

The authors affirm that the manuscript is entirely original and has not been previously published. They collectively grant consent for its submission to the '*Journal of Food Composition and Analysis*'.

CRediT authorship contribution statement

Naseeb Singh: Writing – original draft, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Simardeep Kaur:** Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Antil Jain:** Formal analysis. **Amit Kumar:** Writing – review & editing, Resources. **Rakesh Bhardwaj:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Renu Pandey:** Writing – review & editing, Supervision, Resources. **Amritbir Riar:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

All the authors declare no conflict of interest and state that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors express gratitude to the Directors of the ICAR- Research Complex for NEH Region, Umiam, Meghalaya – 793103 (India), ICAR-National Bureau of Plant Genetic Resources, and ICAR- Indian Agricultural Research Institute, New Delhi – 110012 (India), for providing the necessary facilities for conducting this study.

References

An, Y., Li, S., Huang, X., Chen, X., Shan, H., Zhang, M., 2022. The role of copper homeostasis in brain disease. IJMS 23, 13850. https://doi.org/10.3390/ijms232213850.

Arjin, C., Pringproa, K., Hongsibsong, S., Ruksiriwanich, W., Seel-audom, M., Mekchay, S., Sringarm, K., 2020. In vitro screening antiviral activity of Thai medicinal plants against porcine reproductive and respiratory syndrome virus. BMC Vet. Res 16, 102. https://doi.org/10.1186/s12917-020-02320-8.

Bai, X., Zhang, L., Kang, C., Quan, B., Zheng, Y., Zhang, X., Song, J., Xia, T., Wang, M., 2022. Near-infrared spectroscopy and machine learning-based technique to predict quality-related parameters in instant tea. Sci. Rep. 12, 3833. https://doi.org/10.1038/s41598-022-07652-z.

Baianu, I., Guo, J., 2011. NIR calibrations for soybean seeds and soy food composition analysis: total carbohydrates, oil, proteins and water contents. Nat. Prec. https://doi.org/10.1038/npre.2011.6611.1.

Beattie, J.R., Esmonde-White, F.W.L., 2021. Exploration of principal component analysis: deriving principal component analysis visually using spectra. Appl. Spectrosc. 75, 361–375. https://doi.org/10.1177/0003702820987847.

Biau, G., Scornet, E., 2016. A random forest guided tour. TEST 25, 197–227. https://doi.org/10.1007/s11749-016-0481-7.

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Bucchianico, A.D., 2007. Coefficient of Determination (R2. In: Ruggeri, F., Kenett, R.S., Faltin, F.W. (Eds.), Encyclopedia of Statistics in Quality and Reliability. Wiley. https://doi.org/10.1002/9780470061572.eqr173.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn 20, 273–297. https://doi.org/10.1007/BF00994018.

Cozzolino, D., 2015. Foodomics and infrared spectroscopy: from compounds to functionality. Curr. Opin. Food Sci. 4, 39–43. https://doi.org/10.1016/j.cofs.2015.05.003.

Dhyani, A., Chopra, R., Garg, M., 2019. A Review on Nutritional Value, Functional Properties and Pharmacological Application of Perilla (Perilla frutescens L.). Biomed. Pharmacol. J. 12, 649–660. https://doi.org/10.13005/bpj/1685.

Duan, K., Keerthi, S.S., Poo, A.N., 2003. Evaluation of simple performance measures for tuning SVM hyperparameters. Neurocomputing 51, 41–59. https://doi.org/10.1016/S0925-2312(02)00601-X.

Farahani, H.A., Rahiminezhad, A., Same, L., Immannezhad, K., 2010. A Comparison of Partial Least Squares (PLS) and Ordinary Least Squares (OLS) regressions in predicting of couples mental health based on their communicational patterns. Procedia - Soc. Behav. Sci. 5, 1459–1463. https://doi.org/10.1016/j.sbspro.2010.07.308.

Fassio, A., Cozzolino, D., 2004. Non-destructive prediction of chemical composition in sunflower seeds by near infrared spectroscopy. Ind. Crops Prod. 20, 321–329. https://doi.org/10.1016/j.indcrop.2003.11.004.

Fawagreh, K., Gaber, M.M., Elyan, E., 2014. Random forests: from early developments to recent advancements. Syst. Sci. Control Eng. 2, 602–609. https://doi.org/10.1080/21642583.2014.956265.

Folli, G.S., Santos, L.P., Santos, F.D., Cunha, P.H.P., Schaffel, I.F., Borghi, F.T., Barros, I.H.A.S., Pires, A.A., Ribeiro, A.V.F.N., Romão, W., Filgueiras, P.R., 2022. Food analysis by portable NIR spectrometer. Food Chem. Adv. 1, 100074. https://doi.org/10.1016/j.focha.2022.100074.

Gholamalinezhad, H., Khosravi, H., 2020. Pooling Methods in Deep Neural Networks. a Rev. https://doi.org/10.48550/ARXIV.2009.07485.
Gohain, B., Kumar, P., Malhotra, B., Augustine, R., Pradhan, A.K., Bisht, N.C., 2021. A comprehensive Vis-NIRS equation for rapid quantification of seed glucosinolate content and composition across diverse Brassica oilseed chemotypes. Food Chem. 354, 129527. https://doi.org/10.1016/j.foodchem.2021.129527.
González-Montaña, J.-R., Escalera-Valente, F., Alonso, A.J., Lomillos, J.M., Robles, R., Alonso, M.E., 2020. Relationship between Vitamin B12 and Cobalt Metabolism in Domestic Ruminant: An Update. Animals 10, 1855. https://doi.org/10.3390/ani10101855.
Gualtieri, J.A., Chettri, S., 2000. Support vector machines for classification of hyperspectral data. IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium. Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment. Proceedings (Cat. No.00CH37120). Presented at the IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium. Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment. IEEE, Honolulu, HI, USA, pp. 813–815. https://doi.org/10.1109/IGARSS.2000.861712.
He, Y., Feng, S., Deng, X., Li, X., 2006. Study on lossless discrimination of varieties of yogurt using the Visible/NIR-spectroscopy. Food Res. Int. 39, 645–650. https://doi.org/10.1016/j.foodres.2005.12.008.
He, Y., Li, X., Deng, X., 2007. Discrimination of varieties of tea using near infrared spectroscopy by principal component analysis and BP model. J. Food Eng. 79, 1238–1242. https://doi.org/10.1016/j.jfoodeng.2006.04.042.
Hodson, T.O., 2022. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. Geosci. Model Dev. 15, 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022.
Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Netw. 2, 359–366. https://doi.org/10.1016/0893-6080(89)90020-8.
Howley, T., Madden, M.G., O'Connell, M.-L., Ryder, A.G., 2006. The Effect of Principal Component Analysis on Machine Learning Accuracy with High Dimensional Spectral Data. In: Macintosh, A., Ellis, R., Allen, T. (Eds.), Applications and Innovations in Intelligent Systems XIII. Springer London, London, pp. 209–222. https://doi.org/10.1007/1-84628-224-1_16.
Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. arXiv:1502.03167 [cs]. https://doi.org/⟨10.48550/arXiv.1502.03167⟩.
John, R., Bartwal, A., Jeyaseelan, C., Sharma, P., Ananthan, R., Singh, A.K., Singh, M., Gayacharan, Rana, J.C., Bhardwaj, R., 2023. Rice bean-adzuki bean multitrait near infrared reflectance spectroscopy prediction model: a rapid mining tool for trait-specific germplasm. Front. Nutr. 10, 1224955. https://doi.org/10.3389/fnut.2023.1224955.
Kaur, S., Godara, S., Singh, N., Kumar, A., Pandey, R., Adhikari, S., Jaiswal, S., Singh, S. K., Rana, J.C., Bhardwaj, R., Singh, B.K., Riar, A., 2024. Multivariate Data Analysis Assisted Mining of Nutri-rich Genotypes from North Eastern Himalayan Germplasm Collection of Perilla (Perilla frutescens L.). Plant Foods Hum. Nutr. https://doi.org/10.1007/s11130-024-01220-8.
Khan, S.R., Sharma, B., Chawla, P.A., Bhatia, R., 2022. Inductively Coupled Plasma Optical Emission Spectrometry (ICP-OES): a Powerful Analytical Technique for Elemental Analysis. Food Anal. Methods 15, 666–688. https://doi.org/10.1007/s12161-021-02148-4.
Khatri, P., Gupta, K.K., Gupta, R.K., 2021. A review of partial least squares modeling (PLSM) for water quality analysis. Model. Earth Syst. Environ. 7, 703–714. https://doi.org/10.1007/s40808-020-00995-4.
Kim, J.M., Liceaga, A.M., Yoon, K.Y., 2019. Purification and identification of an antioxidant peptide from perilla seed (*Perilla frutescens*) meal protein hydrolysate. Food Sci. Nutr. 7, 1645–1655. https://doi.org/10.1002/fsn3.998.
Kim, J.M., Yoon, K.Y., 2020. Functional properties and biological activities of perilla seed meal protein hydrolysates obtained by using different proteolytic enzymes. Food Sci. Biotechnol. 29, 1553–1562. https://doi.org/10.1007/s10068-020-00810-x.
Kingma, D.P., Ba, J., 2017. Adam: A Method for Stochastic Optimization. arXiv: 1412.6980 [cs].
Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., Inman, D.J., 2021. 1D convolutional neural networks and applications: A survey. Mech. Syst. Signal Process. 151, 107398. https://doi.org/10.1016/j.ymssp.2020.107398.
Long, S., Romani, A.M., 2014. Role of cellular magnesium in human diseases. Austin J. Nutr. Food Sci. 2, 1–19. PMID: 25839058; PMCID: PMC4379450.
Mabhaudhi, T., Chimonyo, V.G.P., Hlahla, S., Massawe, F., Mayes, S., Nhamo, L., Modi, A.T., 2019. Prospects of orphan crops in climate change. Planta 250, 695–708. https://doi.org/10.1007/s00425-019-03129-y.
Malek, S., Melgani, F., Bazi, Y., 2018. One-dimensional convolutional neural networks for spectroscopic signal regression. J. Chemom. 32, e2977. https://doi.org/10.1002/cem.2977.
Mishra, P., Passos, D., Marini, F., Xu, J., Amigo, J.M., Gowen, A.A., Jansen, J.J., Biancolillo, A., Roger, J.M., Rutledge, D.N., Nordon, A., 2022. Deep learning for near-infrared spectral data modelling: Hypes and benefits. TrAC Trends Anal. Chem. 157, 116804. https://doi.org/10.1016/j.trac.2022.116804.
Mohammed Rashid, A., Midi, H., Dhhan, W., Arasan, J., 2022. Detection of outliers in high-dimensional data using nu support vector regression. J. Appl. Stat. 49, 2550–2569. https://doi.org/10.1080/02664763.2021.1911965.
Nair, V., Hinton, G.E., 2010. Rectified Linear Units Improve Restricted Boltzmann Machines. International Conference on International Conference on Machine Learning. Omnipress, 2600. Anderson, St Madison, WI, United States, Haifa, Israel.
Nishiguchi, J., Kaseda, C., Nakayama, H., Arakawa, M., Yun, Y., 2009. Practical Approach to Outlier Detection Using Support Vector Regression. In: Köppen, M., Kasabov, N., Coghill, G. (Eds.), Advances in Neuro-Information Processing, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 995–1001. https://doi.org/10.1007/978-3-642-02490-0_121.
Palacios, C., Hofmeyr, G.J., Cormick, G., Garcia-Casal, M.N., Pena-Rosas, J.P., Betran, A. P., 2021. Current calcium fortification experiences: A review. Ann. N. Y. Acad. Sci. 1484, 55–73. https://doi.org/10.1111/nyas.14481.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830.
Posom, J., Maraphum, K., 2023. Achieving prediction of starch in cassava (Manihot esculenta Crantz) by data fusion of Vis-NIR and Mid-NIR spectroscopy via machine learning. J. Food Compos. Anal. 122, 105415. https://doi.org/10.1016/j.jfca.2023.105415.
Ruamrungsri, N., Siengdee, P., Sringarm, K., Chomdej, S., Ongchai, S., Nganvongpanit, K., 2016. In vitro cytotoxic screening of 31 crude extracts of Thai herbs on a chondrosarcoma cell line and primary chondrocytes and apoptotic effects of selected extracts. Vitr. Cell. Dev. Biol. -Anim. 52, 434–444. https://doi.org/10.1007/s11626-016-0006-4.
Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533–536. https://doi.org/10.1038/323533a0.
Sang, X., Zhou, R., Li, Y., Xiong, S., 2022. One-Dimensional Deep Convolutional Neural Network for Mineral Classification from Raman Spectroscopy. Neural Process Lett. 54, 677–690. https://doi.org/10.1007/s11063-021-10652-1.
Serna, J., Bergwitz, C., 2020. Importance of Dietary Phosphorus for Bone Metabolism and Healthy Aging. Nutrients 12, 3001. https://doi.org/10.3390/nu12103001.
Shen, Z., Viscarra Rossel, R.A., 2021. Automated spectroscopic modelling with optimised convolutional neural networks. Sci. Rep. 11, 208. https://doi.org/10.1038/s41598-020-80486-9.
Sherstinsky, A., 2020. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. Phys. D: Nonlinear Phenom. 404, 132306. https://doi.org/10.1016/j.physd.2019.132306.
Singh, D., Singh, B., 2020. Investigating the impact of data normalization on classification performance. Appl. Soft Comput. 97, 105524. https://doi.org/10.1016/j.asoc.2019.105524.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958. ⟨http://jmlr.org/papers/v15/srivastava14a.html⟩.
Sweet, L., Müller, C., Anand, M., Zscheischler, J., 2023. Cross-Validation Strategy Impacts the Performance and Interpretation of Machine Learning Models. Artif. Intell. Earth Syst. 2, e230026. https://doi.org/10.1175/AIES-D-23-0026.1.
Talabi, A.O., Vikram, P., Thushar, S., Rahman, H., Ahmadzai, H., Nhamo, N., Shahid, M., Singh, R.K., 2022. Orphan Crops: A Best Fit for Dietary Enrichment and Diversification in Highly Deteriorated Marginal Environments. Front. Plant Sci. 13, 839704. https://doi.org/10.3389/fpls.2022.839704.
Tang, X., Yan, Z., Miao, Y., Ha, W., Li, Z., Yang, L., Mi, D., 2023. Copper in cancer: from limiting nutrient to therapeutic target. Front. Oncol. 13, 1209156. https://doi.org/10.3389/fonc.2023.1209156.
Tomar, M., Bhardwaj, R., Kumar, M., Pal Singh, S., Krishnan, V., Kansal, R., Verma, R., Yadav, V.K., Dahuja, A., Ahlawat, S.P., Rana, J.C., Bollinedi, H., Ranjan Kumar, R., Goswami, S., T, V., Satyavathi, C.T., Praveen, S., Sachdev, A., 2021a. Nutritional composition patterns and application of multivariate analysis to evaluate indigenous Pearl millet ((Pennisetum glaucum (L.) R. Br.) germplasm. J. Food Compos. Anal. 103, 104086. https://doi.org/10.1016/j.jfca.2021.104086.
Tomar, M., Bhardwaj, R., Kumar, M., Singh, S.P., Krishnan, V., Kansal, R., Verma, R., Yadav, V.K., Dahuja, A., Ahlawat, S.P., Chand Rana, J., Satyavathi, C.T., Praveen, S., Sachdev, A., 2021b. Development of NIR spectroscopy based prediction models for nutritional profiling of pearl millet (Pennisetum glaucum (L.)) R.Br: A chemometrics approach. LWT 149, 111813. https://doi.org/10.1016/j.lwt.2021.111813.
Uzun Ozsahin, D., Taiwo Mustapha, M., Saleh Mubarak, A., Said Ameen, Z., Uzun, B., 2022. Impact of Outliers and Dimensionality Reduction on the Performance of Predictive Models for Medical Disease Diagnosis. 2022 International Conference on Artificial Intelligence in Everything (AIE). Presented at the 2022 International Conference on Artificial Intelligence in Everything (AIE). IEEE, Lefkosa, Cyprus, pp. 79–86. https://doi.org/10.1109/AIE57029.2022.00023.
Viscarra Rossel, R.A., McGlynn, R.N., McBratney, A.B., 2006. Determining the composition of mineral-organic mixes using UV–vis–NIR diffuse reflectance spectroscopy. Geoderma 137, 70–82. https://doi.org/10.1016/j.geoderma.2006.07.004.
Wang, J., Liu, H., Ren, G., 2014. Near-infrared spectroscopy (NIRS) evaluation and regional analysis of Chinese faba bean (Vicia faba L.). Crop J. 2, 28–37. https://doi.org/10.1016/j.cj.2013.10.001.
Wang, T., Li, Z., 2017. Outlier detection in high-dimensional regression model. Commun. Stat. - Theory Methods 46, 6947–6958. https://doi.org/10.1080/03610926.2016.1140783.
Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58, 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1.
Wu, X., Dong, S., Chen, H., Guo, M., Sun, Z., Luo, H., 2023. Perilla frutescens: A traditional medicine and food homologous plant. Chin. Herb. Med. 15, 369–375. https://doi.org/10.1016/j.chmed.2023.03.002.
Ye, W., Xu, W., Yan, T., Yan, J., Gao, P., Zhang, C., 2022. Application of Near-Infrared Spectroscopy and Hyperspectral Imaging Combined with Machine Learning Algorithms for Quality Inspection of Grape: A Review. Foods 12, 132. https://doi.org/10.3390/foods12010132.

Yeung, V., Miller, D.D., Rutzke, M.A., 2017. Atomic Absorption Spectroscopy, Atomic Emission Spectroscopy, and Inductively Coupled Plasma-Mass Spectrometry. In: Nielsen, S.S. (Ed.), Food Analysis, Food Science Text Series. Springer International Publishing, Cham, pp. 129–150. https://doi.org/10.1007/978-3-319-45776-5_9.

Zeng, F., Peng, W., Kang, G., Feng, Z., Yue, X., 2021. Spectral Data Classification By One-Dimensional Convolutional Neural Networks. 2021 IEEE International Performance, Computing, and Communications Conference (IPCCC). Presented at the 2021 IEEE International Performance, Computing, and Communications Conference (IPCCC). IEEE, Austin, TX, USA, pp. 1–6. https://doi.org/10.1109/IPCCC51483.2021.9679444.

Zhang, X., Yang, J., Lin, T., Ying, Y., 2021. Food and agro-product quality evaluation based on spectroscopy and deep learning: A review. Trends Food Sci. Technol. 112, 431–441. https://doi.org/10.1016/j.tifs.2021.04.008.