RESEARCH ARTICLE Asian Journal of Dairy and Food Research, Volume 44 Special Issue: 142-150 (November 2025)

Applications of Machine Learning Models in Near-infrared Spectroscopy for Small-grain Quality Control

Ardiana Topi[1], Majlinda Sana[2], Marjeta Daja[3], Dritan Topi[3]

10.18805/ajdfr.DRF-565

ABSTRACT

Background: Wheat quality and safety are crucial for global food security and regional agricultural economies. Conventional analytical methods for assessing grain composition and quality-while accurate-are often time-consuming, destructive and resource-intensive. In contrast, near-infrared spectroscopy (NIR) has emerged as a rapid, non-destructive alternative that provides detailed chemical and physical insights. Recent advancements in machine learning have enhanced NIR's analytical capabilities by enabling the development of predictive models that accurately categorize, quantify and monitor grain quality parameters. These models-spanning from linear regression methods to intricate neural networks-enable researchers to discern significant patterns from spectral data, even amidst noise or overlapping signals.

Methods: This study used NIR spectroscopy and machine learning models to evaluate wheat quality across multiple Albanian regions during the 2023 harvest. NIR spectra were preprocessed using standard normalization and smoothing techniques. Machine learning models-including PLSR, SVM and random forest-were trained to predict key quality traits such as protein content and moisture. Model performance was validated using cross-validation metrics (RMSE, R^2, accuracy), demonstrating the potential of NIR-ML integration for rapid, non-destructive grain assessment.

Result: Protein content ranged from 9.6% to 15.0%, gluten from 19.4% to 37.5% and starch from 66.3% to 71.5%, indicating consistent nutritional profiles across samples. Sedimentation index values (21.1-57.8 cm^3) and moisture levels (9.8-12.5%) confirmed compliance with EU food safety standards. The integration of NIR spectroscopy with machine learning enabled rapid, non-destructive prediction of wheat quality traits, supporting scalable decision-making in agri-food systems.

Key words: Albania agriculture, Machine learning, Near-infrared spectroscopy, Non-destructive analysis, Small-grain cereals, Wheat quality.

INTRODUCTION

Cereals are the primary type of grain-based food used by humans for thousands of years. Wheat one of the world's most important cereal crops, belongs to the Poaceae (grass) family (Goesaert *et al.*, 2005; Zhang *et al.,* 2022). While consumption may vary across regions, wheat grain is the most important commodity for Europeans and the global population (Shewry *et al.,* 2009; Williams, 2006). *Triticum durum* and *T. aestivum* are the two most distinguished wheat species, with the latter commonly referred to as soft wheat, while *T. durum* is sometimes known as durum wheat. Soft wheat is cultivated across all temperate regions (Shewry *et al.,* 2009). The seed consists of bran, endosperm and germ, each with a distinct chemical composition. The endosperm contains high amounts of proteins and carbohydrates, while the germ provides vitamins, trace minerals and triglycerides. Although bran is rich in dietary fiber, it is commonly separated from the grain prior to consumption (Javid Iqbal *et al.*, 2022; Khalid *et al.*, 2023). Typically, grains are milled to produce flour, which serves as a primary source of nutrients in many diets. However, alternative consumption practices that retain or reincorporate bran are increasingly explored for their health benefits (Giraldo *et al.*, 2019). Wheat flour primarily contains starch and gluten-forming proteins. Wheat stands out among cereals for its ability to produce bread and other dough products due to its gluten protein. The gluten proteins provide dough elasticity and gas retention (Shewry *et al.,* 2009; Javid Iqbal *et al.,* 2022).

Agriculture accounts for 58% of employment and 21% of the GDP. Arable land covers 24% of the country. Approximately 49% of agricultural land is used to cultivate

[1]Department of Informatics and Technology, Faculty of Engineering, Informatics and Architecture, European University of Tirana, Street Xhanfize Keko, Kompleksi Xhura, Tirana 1000, Albania.
[2]Department of Natural Applied Sciences, Faculty of Professional Studies, University "Aleksandër Moisiu" Durrës, Durrës, Albania.
[3]Department of Chemistry, Faculty of Natural Sciences, University of Tirana, Blvd Zog 1, 25/1, Tirana 1001, Albania.

Corresponding Author: Dritan Topi, Department of Chemistry, Faculty of Natural Sciences, University of Tirana, Blvd Zog 1, 25/1, Tirana 1001, Albania. Email: dritan.topi@unitir.edu.al
ORCIDs: 0000-0002-3655-6516; 0009-0004-4878-6580; 0000-0001-7852-5374; 0009-0000-2657-3446

How to cite this article: Topi, A., Sana, M., Daja, M. and Topi, D. (2025). Applications of Machine Learning Models in Near-infrared Spectroscopy for Small-grain Quality Control. *Asian Journal of Dairy and Food Research*. **44 (Special Issue):** 142-150. doi: 10.18805/ajdfr.DRF-565.

Submitted: 18-06-2025 **Accepted:** 24-10-2025 **Online:** 05-11-2025

feed crops rather than for food consumption (Diku, 2011). Public concern over wheat flour safety is widespread (Zhang *et al.,* 2022).

Machine learning and NIR spectroscopy

Machine learning models, combined with big data technologies and high-performance computing, create new opportunities for data-intensive science in precision farming and sustainable agriculture. Machine learning (ML) main categories are 1) Supervised learning (SL), 2) Unsupervised learning (UL) and 3) Reinforcement learning (RL). SP algorithms use a training dataset of labeled data to infer a function that predicts new data. UL algorithms directly examine the data and learn patterns from it without requiring human supervision. ML is widely used to solve complex problems, often involving multiple factors. ML models have found applications in the multidisciplinary agri-technologies domain for crop management, furthermore have significantly enhanced the application of near-infrared (NIR) spectroscopy across various fields, including agriculture, medical diagnostics, pharmaceutical products, environmental monitoring and the assessment of food quality and origin.

Advanced NIR food spectral analysis techniques and traditional machine learning

Traditional ML methods are vital in near-infrared (NIR) spectral analysis, addressing multi-collinearity and enhancing generalization. NIR analysis involves pre-processing, feature selection and modeling. ML algorithms like principal component analysis (PCA), partial least squares (PLS), extreme learning machines (ELM), support vector machines (SVM), support vector regression (SVR), decision trees (DT) and random forests (RF). These methods extract significant features, reduce redundancy and create predictive models for NIR applications. Recent advances focus on improving pre-processing, wavelength selection and feature extraction (Sanjeevannavar *et al.,* 2023). Besides traditional ML methods, NIRS is increasingly applied to (1) detecting food contamination and chemical residuals in agricultural products; (2) fostering sustainable agriculture by monitoring crops and soil nutrients; (3) optimizing harvest times for maximum yield and (4) assessing product quality for high-value items. However, ML in near-infrared reflectance spectroscopy (NIRS) is still in its early stages of development compared to other fields due to the difficulty in acquiring specialized data needed for analysis.

Traditional methods for assessing wheat quality-such as wet chemistry, Kjeldahl protein analysis, or sedimentation testing-are often labor-intensive, time-consuming and destructive, limiting their scalability in field or industrial settings. In contrast, near-infrared (NIR) spectroscopy offers a rapid, non-destructive alternative capable of capturing complex chemical and physical traits through spectral signatures. When paired with machine learning algorithms, NIR data can be transformed into predictive models that identify subtle patterns and correlations, even in noisy or overlapping spectral regions. This integration enhances analytical speed, reduces operational costs and enables real-time decision-making, making it particularly suitable for modern agri-food systems. Recent studies (Du *et al*., 2022; Wang *et al*., 2025) have demonstrated the effectiveness of NIR-ML pipelines in predicting wheat processing traits with high accuracy, further validating their applicability in both laboratory and field contexts.

Machine learning (ML) application in agricultural raw data

NIRS devices, chemometric techniques and computer technology have significantly enhanced methodologies (Phuong *et al.,* 2025). Machine learning (ML) models employ a scientific approach in which a machine is trained to learn without being explicitly programmed (Samuel, 2000). It includes three main types: (1) supervised learning (SL), (2) unsupervised learning (UL) and (3) reinforcement learning (RL). In agricultural systems, applications of ML models can be classified into (a) crop management, (b) livestock management, (c) water management and (d) soil management. Within the broader agri-tech sector, machine learning (ML) models have been employed in crop management (61%), yield forecasting (20%) and disease detection (22%) (Liakos *et al.,* 2018). NIR is acknowledged as one of the most promising analytical techniques available today, valued for its non-invasive, rapid and non-destructive qualities, high throughput, simple sample preparation, chemical-free operation, portability and user-friendliness for non-specialists (Abbaspour-Gilandeh *et al.,* 2024). The potential for further advancement is substantial, driven by improvements in optics and recent progress in data science, artificial intelligence, machine learning and deep learning. This advancement enables simultaneous measurement of multiple constituents from a single spectrum.

Recent studies have shown that machine learning-assisted NIR spectroscopy enables rapid, non-destructive prediction of wheat processing traits, outperforming conventional wet chemistry in both efficiency and field applicability (Wang *et al.,* 2025). Compared to conventional analytical techniques, NIR spectroscopy offers faster, non-destructive assessment of grain quality traits, particularly when paired with machine learning models that enhance predictive accuracy and scalability (Du *et al.,* 2022).

NIR versatility in food and feed analysis

Near-infrared (NIR) spectroscopy is a versatile analytical technique that operates within the wavelength range of 780 nm to 2500 nm. It measures the absorption, emission, reflection and diffuse reflection of light, providing valuable insights into the molecular structure and composition of various substances (Ozaki *et al.,* 2017). NIR spectroscopy measures absorption bands resulting from overtones and combination excitations of molecular vibrations. These bands are typically smooth and broad, which necessitates high

signal-to-noise ratios and stable instrumentation for accurate quantitative analysis (Gao *et al.,* 2021). NIR spectroscopy determines the nutritional content of agricultural products, including cereals, fruits, vegetables and animal feed, by quantifying protein, starch, oil and micronutrients (Johnson *et al.,* 2020). As a rapid, non-destructive technique, it monitors product quality, detects stored-grain insects and manages food logistics (Abbaspour-Gilandeh *et al.,* 2024). This method evaluates seed quality, including variety discrimination, germination rate, moisture content and vigor (Qiu *et al.,* 2005). It aids in food safety by enabling online analysis of proteins, polysaccharides and polyphenols, as well as the detection of adulteration (Xu *et al.,* 2025). NIR evaluates the quality of fruits and vegetables, including soluble solids, acidity, moisture, texture, ripeness and overall quality (Sirisomboon, 2018). Challenges such as calibration and accuracy issues stem from environmental variability, while robust calibration models can enhance reliability (Xu *et al.,* 2019). Technological advancements, including cloud computing, the internet of things (IoT) and machine learning, are expected to enhance real-time monitoring and predictive modeling, thereby expanding the agricultural applications of near-infrared reflectance (NIR) spectroscopy (Phuong *et al.,* 2025). NIR offers advantages in crop quality assessment, including non-destructive analysis, minimal sample preparation, rapidity and high accuracy across various crop types, integrating with information and communication technology (Sirisomboon, 2018). NIR enables the rapid evaluation of crop viability, moisture content and other indicators in real-time (Qiu *et al.,* 2005; Barbin *et al*., 2013; Johnson *et al.,* 2020). NIRS effectively determines the nutritional composition of cereal grains, including protein, carbohydrate and lipid content (Kays 2015). Recent studies have also explored the use of NIR spectroscopy for predicting mineral composition in fortified wheat flours, demonstrating its versatility beyond traditional macronutrient analysis (Martínez-Martín *et al.,* 2023).

Near-infrared spectroscopy applications in agriculture

Near-infrared (NIR) spectroscopy is a technique used to assess various physical characteristics and chemical compounds related to the quantity and quality of agricultural products (Shewry *et al.,* 2009). The accuracy of this methodology relies on the NIR model. Moreover, macroconstituents like starch and fat, along with essential micronutrients such as amino acids, dietary fiber and amylose, are commonly analyzed using near-infrared reflectance (NIR) spectroscopy. It is also used in quality control, primarily by measuring flour moisture during milling or dough processing (Chadalavada *et al.,* 2022).

Additionally, NIRS aids in identifying food contaminants, *e.g.*, mycotoxins and ergot bodies (Delwiche, 2021). NIR offers several advantages, including cost savings and faster assessment than traditional methods. It operates without prior sample preparation or chemical agents, providing significant benefits for quality control and process monitoring at an industrial scale (Shewry *et al.,* 2009). NIR spectroscopy's spectral range provides reliable, consistent insights into food quality (Ozaki *et al.,* 2017). Its temporal range extends from the visible to the mid-infrared (Schuster *et al.,* 2023).

NIR analysis involves data acquisition, noise reduction, calibration and evaluation (Cen and He, 2007). Analyzing data establishes the relationship between the unique features of the investigated sample and its transmittance or absorption values (Zhang *et al.,* 2022). The scope of NIR applications continues to expand, including mineral profiling in composite flours such as wheat-lentil blends (Martínez-Martín *et al.,* 2023). Recent studies in Indian agricultural contexts have explored the use of NIR spectroscopy for evaluating grain quality traits, highlighting its potential for rapid, non-destructive analysis in post-harvest systems (Venkatesan *et al*., 2020).

MATERIALS AND METHODS

Wheat samples collection

A total of 75 wheat samples were collected during the harvesting season from three agricultural regions in Albania-Fieri, Elbasan and Korça-representing diverse climatic conditions. These samples were collected by the established sampling protocol and stored in bags. Eventually, they were delivered to the laboratory and kept in a refrigerated chamber at 4°C in the dark until undergoing physicochemical analysis. The experimental procedures and data analyses were conducted at the Kosovo Agriculture Institute, Pejë, Kosovo, and Department of Chemistry, Faculty of Natural Sciences, Tirana, Albania. The research was conducted during the 2023 harvesting season, with sample collection, spectral analysis, and model development carried out between June and September 2023.

Physicochemical analysis

The standard approach for assessing grain protein levels typically involves near-infrared (NIR) spectroscopy, applicable to whole grains, milled grains or flour. This technique was based on the relationship between NIR spectroscopy and total nitrogen analysis. A total of 75 representative samples from the whole batch underwent laboratory testing. These samples were utilized to calibrate a partial least squares regression (PLSR) model, which was then used to estimate the amounts of water, protein, starch, β-glucan, fat and moisture.

According to Standard EN 15948:2015 and Commission Regulation (EC) 687/2008, the FOSS device (InfratecTM 1241) utilizes the near-infrared wavelength region to analyze kernels, enabling the assessment of key parameters such as fat, hydration, protein, gluten, sediment and starch in whole kernels (Fig 1) (EN, 2015; EC, 2008). This device consists of a sample chamber, a light source and monochromator, a wavelength selection processor and a detector. A minimum bulk sample of 500 grams is introduced into a hopper. The essential data is produced through

reflectance mode within the 1100-2500 nm wavelength range. The greater light energy in the lower range facilitates deeper penetration into the seed kernel (EN, 2015). Utilizing Beer's law, various methods are employed to assess optical data from absorbed samples, aiming to establish a connection between analyte concentration and sample absorbance (A) at specific wavelengths.

The multiple regression equation is used for calibration as follows:

$$Y = Bo + Bi (- \log Ri) N + E \quad ....(1)$$

Y = Concentration percentage of the absorber.
Bo = Regression intercept.
Bi = Regression coefficient.
i = Index of the specific wavelength.
Ri = Reflectance at that wavelength.
N = Total number of wavelengths used in the regression.
E = Random error.

Spectral preprocessing

Applying machine learning (ML) to near-infrared (NIR) spectroscopy poses challenges that researchers must address to improve model accuracy and reliability. These challenges fall into data-related, model-related and practical implementation issues, which may be irrelevant or noisy, obscuring the relationship between the spectra and target variables. Effective pre-processing methods, such as Savitzky-Golay filters, are crucial for improving data quality. Traditional machine learning (ML) models still lead in near-infrared (NIR) applications. In this context, data pre-processing and feature selection are key elements that enhance data quality, reduce dimensionality, extract meaningful information from NIR spectra and boost model performance.



Fig 1: FOSS near-infrared reflectance (InfratecTM 1241).

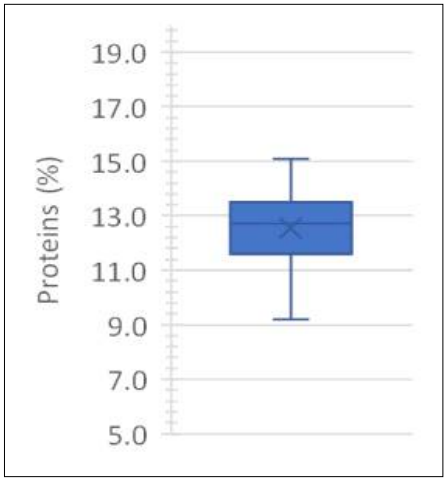


Fig 2: Protein content in wheat samples (%).

Pre-processing NIR data is essential for enhancing model performance and involves techniques like noise reduction, baseline correction, resolution enhancement, centering, smoothing, derivative calculation, detrending and scaling. Most studies employ only one or two methods, which are often chosen based on experience rather than a systematic evaluation. Since some machine learning algorithms struggle with noisy data, pre-processing aims to reduce additive and multiplicative noise bias. Common NIR data pre-processing functions include (1) mean centering; (2) standard normal variate (SNV); (3) multiplicative scatter correction (MSC); (4) extended multiplicative scatter correction (EMSC); (5) inverse scatter correction (ISC) and (6) Savitzky-Golay smoothing (Zhang *et al.,* 2022). Effective pre-processing is essential to enhance model accuracy and reliability. Techniques such as Savitzky-Golay filtering, Standard Normal Variate (SNV) and Multiplicative Scatter Correction (MSC) are commonly used to reduce baseline variation, correct scatter effects and improve signal clarity. In this study, raw NIR spectra were pre-processed using SNV, MSC and Savitzky-Golay smoothing. These methods were selected to minimize additive and multiplicative noise, enhance resolution, and ensure consistent input quality for downstream machine learning analysis.

Machine learning models

Three machine learning models-partial least squares regression (PLSR), support vector machines (SVM), and random forest classifiers-were applied to predict key wheat quality parameters. These models were selected based on their proven performance in spectral data analysis and their complementary strengths in regression and classification tasks. The integration of machine learning with spectral data has been shown to improve classification accuracy for cereal grains, particularly in regional studies focused on wheat and maize (Priyadarshi Bala *et al.*, 2023).

Validation strategy

Model performance was evaluated using 10-fold cross-validation, ensuring robustness and minimizing overfitting. Due to sample size constraints, no independent test set was used; however, the cross-validation approach provided reliable internal validation and generalization metrics.

RESULTS AND DISCUSSION

As a rapid, non-destructive technique, near-infrared spectroscopy (NIR) has become increasingly popular for evaluating food and feed quality (Zhang *et al.,* 2022). It provides

decision-makers with a helpful tool to protect the quality of grain and associated products during storage and processing. Seventy-five wheat samples from the harvest season were analyzed using FOSS NIR equipment, specifically the InfratecTM 1241. The analyzed quality parameters included protein content, moisture, starch content, gluten and sedimentation index (Table 1).

The protein content ranged from 9.2 to 15.1%, with an average of 12.5% (Fig 2).

The near-infrared (NIR) examination of the starch in the wheat sample yielded a range of 66.8 - 72.0%, with an average value of 69.6% (Fig 3).

The gluten concentration ranged from 18.2% to 39.2%, with an average of 28.2% being the number of gluten occurrences.

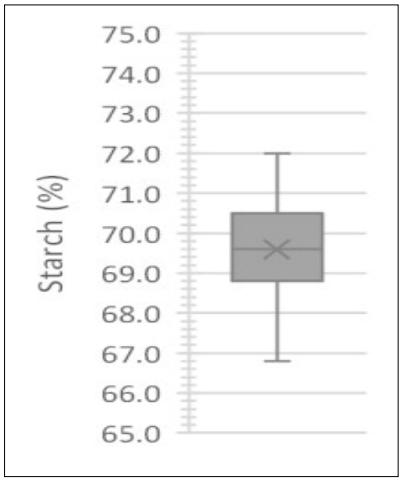


Fig 3: Starch content in wheat samples (%).

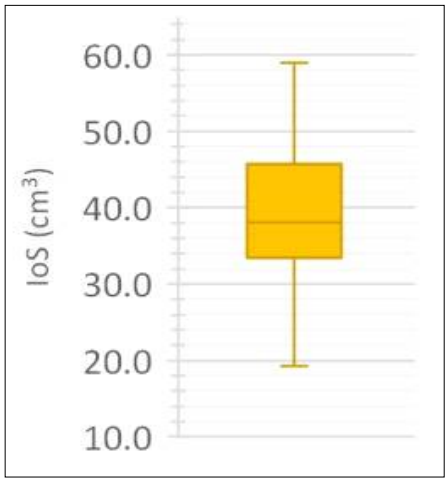


Fig 4: The index of sedimentation (IoS) (cm^3).

The sedimentation index varied from 19.2 to 59.0 cm^3, with an average of 39.35 cm^3 (Fig 4).

Optimal moisture levels are crucial for the proper storage of flour. Elevated flour moisture levels facilitate the proliferation of mold and worms throughout storage. According to EU law, the moisture content of the wheat samples ranged from 9.5% to 12.2%, with none exceeding the maximum threshold of 14.5%. The mean protein content was 12.54±1.27%, with a maximum of 15.1%.

This study employed near-infrared reflectance (NIR) spectroscopy to investigate the physicochemical properties of wheat grain grown and consumed in Albania, including protein, gluten, starch, sedimentation index and moisture content. NIR spectroscopy has proven to be an effective method in the food and agricultural industries for accurately determining protein content in grains, particularly wheat. It plays a crucial role in quality control by offering a precise analytical approach for assessing grain composition. It is also utilized as a rapid, non-destructive method during seed screening in breeding processes (Delwiche, 2021). The accurate evaluation of wheat protein levels is attributed to the strong absorption of N-H bonds in the near-infrared (NIR) spectral range (Caporaso *et al.,* 2018). A NIR monochromator operating in the 400-2500 nm range was used to conduct reflectance analysis on whole grain samples. The partial least squares (PLS) method was employed to establish calibration equations for the qualitative characteristics of whole wheat (Pojic and Mastilovic, 2013). This method provides a significant benefit: the tests are quick and cost-effective.

In contrast to conventional laboratory methods, these economical tests do not necessitate solvents or expensive lab equipment, which is especially advantageous for industrial-scale quality control (Cen and He, 2007). The calibration model establishes a connection between spectral data and the target molecule or property. However, creating a NIR calibration model can be challenging due to the complex nature of the samples being analyzed, which often results in numerous interference bands where essential characteristics or components frequently overlap. The NIR methodology has undergone significant advancements due to improvements in NIR devices, chemometric techniques and computer technology (Phuong *et al.,* 2025).

The gluten content is crucial for assessing the quality of wheat flour, as it significantly influences baking results. While protein content can range from 8% to 16%, it does not always indicate baking quality, as flours with similar protein levels but different gluten compositions can produce varied baking outcomes (Schuster *et al.,* 2023). Wheat flour

Table 1: Chemical-physical parameters in wheat during the 2023 harvesting year.

Parameter	Proteins (%)	Humidity (%)	Starch (%)	Gluten (%)	Sedimentation index (cm^3)
Min	9.20	9.50	66.80	18.20	19.20
Max	15.10	12.20	72.00	39.20	59.00
Mean	12.54±1.27	10.86±0.53	69.59±1.21	28.17±4.34	39.35±8.62

mainly consists of protein and glucose, in proportions of 10-12% and 70-75%, respectively. In contrast, carbohydrates and lipids make up minor components, each at roughly 2% (Goesaert *et al.,* 2005). During dough mixing, proteins form a three-dimensional structure that contributes to the strength and chewiness of baked goods.

The protein content varies among commonly eaten cereals. For example, barley contains about 9-12% of its total weight in protein; maize has 8-12%; oats contain 12-15%; rye accounts for 12-15% and wheat comprises 9-16%. The protein levels in this particular wheat variety are higher (12-16%) compared to soft wheat, especially durum wheat (Williams, 2006; Shewry *et al.,* 2009). Various techniques are employed to evaluate protein content (Table 2).

Starch, the primary macromolecule found in plants, is composed of helical chains of glucose linked by α-1,4 bonds, with branches formed by α-1,6 bonds. It is stored as granules in the endosperm of grains. Starch has two main components: amylose, which consists of linear chains of α-(1-4)-linked glucose and amylopectin, the more prevalent component, which is characterized by highly branched polymers (Shewry *et al.,* 2009). Starch analysis using near-infrared reflectance (NIR) spectroscopy of the tested wheat revealed a starch content of 66.8% to 72.0%, with an average of 69.6%. Plant proteins account for nearly 50% of our dietary protein intake, mainly from the top three

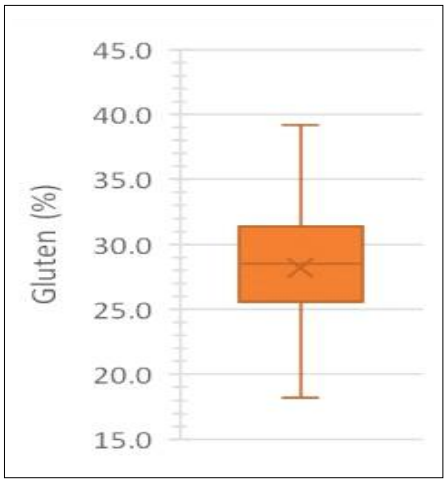


Fig 5: Gluten content in wheat samples (%).

Table 2: Employed techniques for protein content determination in small grains.

Process	Material assayed	Method
Mill	Grain/Flour	NIR
Extract	Protein in solution	Biuret
Hydrolyze	Amino acids	Liquid chromatography
Alkaline digestion	Amide N	Titrate ammonia released
Acid digestion	Ammonia or nitrogen	Kjeldahl method

Source: Shewry *et al.* (2009).

cereal crops: wheat, rice and maize (Shewry *et al.,* 2009). The gluten levels ranged from 18.2% to 39.2%, with an average of 28.2% (Fig 5). The recommended minimum gluten content in wheat flour, evaluated in its wet state, is approximately 24% (Kaushik *et al.,* 2015). Of all the wheat samples analyzed, only 4 (5.3% of the total) had gluten levels below 24%, demonstrating that the wheat harvested in *2023* is of exceptional quality.

The sedimentation index ranged from 19.2 to 59.0 cm^3, with an average of 39.35 cm^3. This test employs a scientific method that provides critical insights into the baking properties of wheat flour. The Zeleny value indicates the degree of sedimentation of flour in a lactic acid solution over a specified period. The Zeleny test evaluates the quality of the baking process. Lower sedimentation rates, associated with higher gluten content and better gluten quality, yield greater Zeleny test values (Hruskova and Famera, 2003).

Prolonged exposure can lead to toxic effects on specific organs. The issue of mycotoxin contamination in Albania, particularly regarding cereals, has been stressed (Topi *et al.,* 2017; Topi *et al.,* 2023; Mato *et al.,* 2024; Topi *et al.,* 2024). However, NIR practitioners have primarily focused on other contaminants, such as aflatoxins in corn and deoxynivalenol in wheat, even though their natural concentrations often remain below the detection limits of NIR reflection or transmission spectroscopy at levels of one part per million or less (Levasseur-Garcia, 2018; Mato *et al.,* 2024).

Maintaining proper moisture levels is crucial for flour storage, as excessive moisture can lead to mold and worm infestations. EU regulations indicate that the moisture levels in sampled wheat ranged from 9.5% to 12.2%, remaining below the maximum allowable level of 14.5%. The average protein content was 12.54±1.27%, with a high of 15.1%. The presence of mycotoxins in grains raises significant concerns for global food safety. The presence of both regulated and unregulated mycotoxins poses ongoing health risks, especially since grains such as wheat and maize are dietary staples for many.

The integration of near-infrared (NIR) spectroscopy with machine learning models enabled rapid, non-destructive assessment of wheat quality traits across diverse Albanian regions. Protein content ranged from 9.6% to 15.0%, gluten from 19.4% to 37.5% and starch from 66.3% to 71.5%, reflecting consistent nutritional profiles. Sedimentation index values (21.1-57.8 cm^3) and moisture levels (9.8-12.5%) confirmed compliance with EU food safety standards, underscoring the reliability of the analytical pipeline.

Among the tested algorithms, support vector machines (SVM) demonstrated the highest predictive accuracy for protein content and sedimentation index, particularly in handling nonlinear relationships within the spectral data. This aligns with recent findings by Wang *et al*. (2025), who applied a starfish-optimized SVR model to predict sedimentation value and falling number using a

portable NIR spectrometer, achieving high precision and field applicability. Partial least squares regression (PLSR) offered robust performance across multiple traits and maintained interpretability, consistent with its widespread use in cereal quality modeling as reviewed by Du *et al*. (2022), who emphasized its balance between simplicity and predictive power in NIR applications. Random forest classifiers showed strong classification capabilities but were more sensitive to spectral noise and required careful tuning.

These findings have practical implications for stakeholders across the agri-food value chain. For farmers, the ability to rapidly assess grain quality at harvest supports informed decisions on storage and market timing. Applications of NIR in feed quality assessment have also been reported, supporting its broader utility across both human and animal nutrition domains (Rathore and Bala, 2021). Processors can use model outputs to optimize blending strategies and ensure product consistency, while policymakers may leverage such tools to enforce quality standards and support regional traceability initiatives.

Despite its promise, the approach presents limitations. NIR instrumentation remains costly for small-scale producers and model calibration requires domain-specific expertise and periodic updates to maintain accuracy. Indian research has emphasized the importance of calibration and regional variability in deploying NIR-based models for grain quality control (Priyadarshi and Bhardwaj, 2025). Additionally, expanding the dataset-both in sample size and varietal diversity-is essential to support more complex machine learning architectures and improve generalizability. These challenges echo concerns raised in recent literature (Du *et al*., 2022; Abbaspour-Gilandeh *et al*., 2024) and highlight the need for scalable, context-sensitive deployment strategies.

Overall, the integration of NIR spectroscopy with machine learning offers a scalable, data-driven solution for modernizing cereal quality control in Albania and similar agricultural contexts. Future work should focus on developing portable systems, expanding regional datasets, and fostering interdisciplinary collaboration to bridge analytical precision with field-level applicability.

CONCLUSION

This study demonstrates the effectiveness of integrating near-infrared spectroscopy with machine learning models for rapid, non-destructive assessment of wheat quality traits. The approach yielded reliable predictions for protein content, gluten levels, starch composition and sedimentation index, confirming compliance with EU food safety standards across samples from diverse Albanian regions. Support vector machines (SVM) and partial least squares regression (PLSR) emerged as the most robust algorithms, offering complementary strengths in predictive accuracy and operational simplicity. Their application in agri-food systems holds promise for enhancing quality control, optimizing processing decisions and supporting regulatory oversight. However, the scalability of this approach depends on several practical factors. NIR instrumentation remains costly for small-scale producers and model calibration requires technical expertise and periodic updates. Additionally, expanding the dataset-both in size and varietal diversity-is essential to support more complex machine learning architectures and improve generalizability. Despite these limitations, the integration of NIR and machine learning represents a scalable, data-driven solution for modernizing cereal quality control in Albania and similar agricultural contexts. Future research should focus on developing portable systems, refining calibration protocols, and fostering interdisciplinary collaboration to ensure field-level applicability and long-term sustainability.

ACKNOWLEDGEMENT

The present study received no funding.

Disclaimers

The views and conclusions expressed in this article are solely those of the authors and do not necessarily represent the views of their affiliated institutions. The authors are responsible for the accuracy and completeness of the information provided, but do not accept any liability for any direct or indirect losses resulting from the use of this content.

Informed consent

No animal or human materials were used in the experiments.

Conflict of interest

The authors declare that there are no conflicts of interest regarding the publication of this article. No funding or sponsorship influenced the design of the study, data collection, analysis, decision to publish or preparation of the manuscript.

REFERENCES

Abbaspour-Gilandeh, Y., Abbaspour-Gilandeh, Y., Tahmasebi, M., Kaveh, M., El-Mesery, H., Szymanek, M., Sprawka, M. (2024). VIS/NIR spectroscopy as a non-destructive method for evaluation of quality parameters of three bell pepper varieties based on soft computing methods. *Applied Sciences*. **14:** 10855.

Barbin, D.F., ElMasry, G., Sun, D.W. and Allen, P. (2013). Non-destructive determination of chemical composition in intact and minced pork using near-infrared hyperspectral imaging. *Food Chemistry*. **138(2-3):** 1162-1171. https://doi.org/10.1016/j.foodchem.2012.11.120.

Caporaso, N., Whitworth, M.B. and Fisk, I.D. (2018). Near-infrared spectroscopy and hyperspectral imaging for non-destructive quality assessment of cereal grains. *Applied Spectroscopy Reviews*. **53/8:** 667-687. https://doi.org/10.1080/05704928.2018.1425214.

Cen, H. and He, Y. (2007). Theory and application of near-infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science and Technology*. **18:** 72-83.

Chadalavada, K., Chadalavada, K., Anbazhagan, K., Anbazhagan, K., Ndour, A., Choudhary, S., Palmer, W., Flynn, J., Mallayee, S., Pothu, S., Jones, C., Kholová, J. and Kholová, J. (2022). NIR instruments and prediction methods for rapid access to grain protein content in multiple cereals. *Sensors*. **22:** 3710.

Delwiche, S.R. (2021). Ergot alkaloids mycotoxins in cereals and cereal-derived food products: Characteristic, toxicity, prevalence and control strategies. *Agronomy*. **11/5:** 931. https://doi.org/10.3390/agronomy11050931.

Diku, A. (2011). The Impact of Climate Change on Food Production and Selected Crop Yields in Albania. In: The Impacts of Climate Change on Food Production in the Western Balkans Region [Ivanyi, Z. (ed.)]. Report: Regional Environmental Center for Central and Eastern Europe. pp. 1-23.

Du, Z., Tian, W., Tilley, M., Wang, D., Zhang, G., Li, Y. (2022). Quantitative assessment of wheat quality using near-infrared spectroscopy: A comprehensive review. *Comprehensive Reviews in Food Science and Food Safety*. **21:** 2956-3009. https://doi.org/10.1111/1541-4337.12958.

EC (2008). Commission Regulation (EC) 687/2008 establishes procedures for taking over cereals by intervention agencies and lays down methods of analysis for determining the quality of cereals. 19.7.2008. *Official Journal of the European Union*. **L:** 192/20.

Gao, X., Lu, B., He, K., Tang, X. (2021). Effects of different pretreatment and modeling methods on soil moisture content detected by near-infrared spectroscopy. American Society of Agricultural and Biological Engineers Annual International Meeting, ASABE 2021. 2100335. doi: 10.13031/aim.202100335.

Giraldo, P., Benavente, E., Manzano-Agugliaro, F., Gimenez, E. (2019). Worldwide research trends on wheat and barley: A bibliometric comparative analysis. *Agronomy*. **9:** 352.

Goesaert, H., Brijs, K., Veraverbeke, W.S., Courtin, C.M., Gebruers, K. and Delcour, J.A. (2005). Wheat flour constituents: How they impact bread quality and how to impact their functionality. *Trends in Food Science and Technology*. **16:** 12-30.

Hruskova, M. and Famera, O. (2003). Prediction of wheat and flour Zeleny sedimentation value using NIR technique. *Czech Journal of Food Sciences*. **21:** 91-96.

Javid Iqbal, M., Shams, N., Fatima, K. (2022). Nutritional Quality of Wheat. In: Wheat, (Eds. M-ur-R. Ansari) IntechOpen, Rijeka, Croatia. doi: 10.5772/intechopen.104659.

Johnson, J.B., Walsh, K.B. and Naiker, M. (2023). The use of infrared spectroscopy for quantification of bioactive compounds in food: A review. *Molecules*. **28/7:** 3215. https://doi.org/10.3390/molecules28073215.

Kaushik, R., Kumar, N., Sihag, M.K., Ray, A. (2015). Isolation, characterization of wheat gluten and its regeneration properties. *Journal of Food Science and Technology*. **52:** 5930-5937. doi: 10.1007/s13197-014-1690-2.

Kays, S.E. (2015). Analysis of Cereal Food Products. In: Near-Infrared Reflectance Spectroscopy in Agriculture. [Roberts, C.A., Workman Jr., J., Reeves III, J.B. (eds)]. **44.** doi: 10.2134/agronmonogr44.C16.

Khalid, A., Hameed, A. and Tahir, M.F. (2023). Wheat quality: A review on chemical composition, nutritional attributes, grain anatomy, types, classification and function of seed storage proteins in bread making quality. *Frontiers in Nutrition*. **10:** 1053196. https://doi.org/10.3389/fnut.2023.1053196.

Levasseur-Garcia (2018). Updated overview of infrared spectroscopy methods for detecting mycotoxins on cereals (corn, wheat and barley). *Toxins*. **10/1:** 38. https://doi.org/10.3390/toxins10010038.

Liakos, K.G., Busato, P., Moshou, D., Pearson, S. and Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*. **18:** 2674. doi: 10.3390/s18082674.

Martínez-Martín, I., Hernández-Jiménez, M., Revilla, I. and Vivar-Quintana, A. (2023). Prediction of mineral composition in wheat flours fortified with lentil flour using NIR technology. *Sensors*. **23:** 1491.

Mato, L., Damani, Z., Spahiu, J., Halimi, E., Seiti, B., Topi, D. (2024). High prevalence of mycotoxigenic fungi and aflatoxin B1 contamination in corn and wheat grains grown to albania: Implications for food safety. *Journal of Food Quality and Hazards Control*. **11:** 59-68.

Ozaki, Y., Genkawa, T. and Futami, Y. (2017). Near-infrared Spectroscopy. In: Encyclopedia of Spectroscopy and Spectrometry (3rd edition), [J.C. Lindon, G.E. Tranter, D.W. (eds)] Koppenaal, Academic Press. pp. 40-49. doi: 10.1016/B978-0-12-409547-2.12164-X.

Phuong, N.T.H., Van, H.N., Thanh, X.N.T. and Ngoc, P.N. (2025). Advances in machine learning framework for near-infrared spectroscopy: A taxonomic review on food quality assessment. *Informatica*. **49:** 121-148. doi: 10.31449/inf.v49i11.7482.

Pojic, M.M. and Mastilovic, J.S. (2013). Near-infrared spectroscopy-advanced analytical tool in wheat breeding, trade and processing. *Food and Bioprocess Technology*. **6:** 330-352.

Priyadarshi Bala, M., Sharma, A., Chaturvedi, K. K., Bhardwaj, R., Lal, S.B., Farooqi, M.S., Kumar, S., Mishra, D.C. and Singh, M. (2023). Comparing various machine learning algorithms for sugar prediction in chickpea using near-infrared Spectroscopy. *Legume Research*. **46:** 251-256. doi: 10.18805/LR-4931.

Priyadarshi Bala, M. and Bhardwaj, R. (2025). Analysis of phosphorus content in chickpea samples using near-infrared spectroscopy and machine learning models. *Legume Research*. 1-9. doi: 10.18805/LR-5331.

Qiu, Z., Feng, L., Li, W. and He, Y. (2005). Application of near-infrared spectroscopy in non-destructive quality inspection of agricultural products. Proceedings of SPIE-The International Society for Optical Engineering, Proc. SPIE 6045, MIPPR 2005: Geospatial Information, Data Mining and Applications. 60452E. doi: 10.1117/12.651581.

Rathore, M., Prakash, H.G. and Bala, S. (2021). Evaluation of the nutritional quality and health benefits of chickpea (*Cicer arietinum* L.) by using new technology in agriculture (Near Infra-red spectroscopy-2500). *Asian Journal of Dairy and Food Research*. **40(1):** 123-126. doi: 10.18805/ajdfr.DR-1582.

Samuel, A.L. (2000). Some bodies in machine learning using the game of checkers. *IBM Journal of Research and Development*. **44(1/2):** 206-226. https://doi.org/10.1147/rd.441.0206.

Sanjeevannavar, M., Banapurmath, N., Banapurmath, N., Kumar, V., Sajjan, A., Vadlamudi, C., Krishnappa, S., Kamangar, S., Rahmath, U. and Yunus Khan, T. (2023). Machine learning prediction and optimization of performance and emissions characteristics of IC engine. *Sustainability*. **15:** 13825.

Schuster, C., Huenb, J., Scherf, K.A. (2023). Prediction of wheat gluten composition *via* near-infrared spectroscopy. *Current Research in Food Sciences*. **6:** 100471.

Shewry, P.R., D'Ovidio, R., Lafiandra, D., Jenkins, J.A., Mills, E.N.C. and Békés, F. (2009). Wheat Grain Proteins, In: Wheat (Fourth Edition), [Khan, K. and Shewry, P.R. (Eds)]. *AACC International Press*. pp. 223-298. doi: 10.1016/B978-1-891127-55-7.50015-X.

Sirisomboon, P. (2018). NIR spectroscopy for quality evaluation of fruits and vegetables. *Materials Today: Proceedings*. **5. Part 3:** 22481-22486.

Topi, D., Babic, J., Jakovac-Strajn, B. and Tavcar-Kalcher, G. (2023). Incidence of aflatoxins and ochratoxin A in wheat and corn from Albania. *Toxin*s. pp. 15. doi: 10.3390/toxins 15090567.

Topi, D., Babiè, J., Pavšiè-Vrtaè, K., Tavèar-Kalcher, G. and Jakovac-Strajn, B. (2017). Occurrence of ergot alkaloids in wheat from Albania. *Food Additives and Contaminants. Part A*. **34:** 1333-1343. doi: 10.1080/19440049.2017.1307528.

Topi, D., Damani, Z., Babiè, J., Jakovac-Strajn, B. and Tavèar-Kalcher, G. (2024). The presence of some minor aspergillus and penicillium unregulated mycotoxins in main cereals cultivated in Albania. *Molecules*. **29:** 5292. doi: 10.3390/ molecules29225292.

UNE-EN 15948 (2015). Cereals-Determination of moisture and protein-Method using Near-Infrared Spectroscopy in whole kernels. *European Committee for Standardization (CEN), Burssels*. pp. 40.

Venkatesan S., Masilamani, P., Janaki, P., Eevera, T., Sundareswaran, S., Rajkumar, P. (2020). Role of near-infrared spectroscopy in seed quality evaluation: A review. *Agricultural Reviews*. **41(2):** 106-115. doi: 10.18805/ag.R-1960.

Wang, Y., Zhang, C., Li, X., Xing, L., Lv, M., He, H., Pan, L. and Ou, X. (2025). Machine-learning-algorithm-assisted portable miniaturized NIR spectrometer for rapid evaluation of wheat flour processing applicability. *Foods*. **14:** 1799. https://doi.org/10.3390/foods14101799.

Williams, P. (2006). Grains and Seeds. In: Wiley-Interscience, Hoboken, New Jersey, [Ozaki, Y., McClure, W.F. and Christy, A.A. (eds)]. USA. pp. 165-217.

Xu, X., Xie, L. and Ying, Y. (2019). Factors influencing near-infrared spectroscopy analysis of agro-products: A review. *Frontiers of Agricultural Science and Engineering*. **6:** 105-115. doi: 10.15302/J-FASE-2019255.

Xu, Y., Kong, T., Ma, Y. and Zheng, M. (2025). Near-infrared spectroscopy: Application in ensuring food quality and safety. *Analytical Methods*. **17:** 3381-3406.

Zhang, S., Liu, S., Shen, L., Chen, S., He, L., Liu, A. (2022). Application of near-infrared spectroscopy for the non-destructive analysis of wheat flour: A review. *Current Research in Food Sciences*. **5:** 1305-1312.