A NON-DESTRUCTIVE METHOD OF PREDICTION OF CARBOHYDRATE CONTENT IN MILLETS USING HYPERSPECTRAL IMAGING AND MACHINE LEARNING TECHNIQUES

Shrraddhas Mishra
Computer Science Engineering
Vellore Institute of Technology, Chennai

Anshu Garg
Computer Science Engineering
Vellore Institute of Technology, Chennai

***Abstract*—This study introduces a non-destructive method to predict the carbohydrate content in various millet grains using hyperspectral imaging (HSI) combined with advanced machine learning techniques. Millets, known for their resilience in arid conditions and nutritional benefits, include varieties such as Sorghum, Proso, and Pearl Millet. Traditional methods for nutritional analysis of these grains are labor-intensive and inefficient. Hyperspectral imaging provides a more scalable solution by capturing detailed spectral data across a broad wavelength range, from 900 to 1700 nm, with a spectral resolution of 4.9 nm. This spectral information enables precise identification of unique absorption and reflection characteristics, crucial for predicting nutritional content.**

The study implements regression models, including Linear Regression, Ridge Regression, Support Vector Regression, Partial Least Squares Regression, Random Forest, and XGBoost, to estimate carbohydrate content. The effectiveness of these models is evaluated using metrics such as Mean Squared Error (MSE) and R^2 scores, demonstrating the potential of HSI as a rapid and reliable approach for millet analysis. This innovative method paves the way for advancements in agricultural practices, offering a robust framework for crop quality assessment and management through enhanced imaging and data-driven insights.

I. Keywords

Multi Linear Discriminant Analysis(MLDA), Least Square Support Vector Machine(LS-SVM), Partial Least Squares regression(PLSR), Hyper Spectral Imaging(HSI), Near-Infrared Spectral Resolution(NIR-SR).

II. Introduction

Hyperspectral imaging (HSI) is an advanced technique that captures detailed spectral information across the electromagnetic spectrum. Unlike traditional imaging methods, which capture data at limited wavelengths (like RGB), HSI records continuous spectral data for each pixel in an image, spanning from ultraviolet to infrared wavelengths. This allows for precise analysis of materials based on their unique spectral signatures. HSI is used in agriculture for crop health monitoring, medical imaging for tissue analysis, and remote sensing applications such as geological surveys. By leveraging spectral data, HSI enhances our ability to identify substances, map their distribution, and make informed decisions across various fields.

Millets are a group of small-seeded grains that belong to the Poaceae family, commonly known as the grass family. They are cultivated as cereal crops primarily in semi-arid and arid regions of the world, including parts of Asia, Africa, and even in some regions of Europe and North America. Millets are known for their resilience and ability to grow well in harsh environmental conditions such as low rainfall and poor soil fertility. 11 types of millets include:

- Sorghum Millet (Jowar): Commonly known as Jowar in India, sorghum is predominantly cultivated in the states of Maharashtra and Karnataka. It plays a vital role in the agricultural landscape of India.
- Proso Millet (Chena/Barri): Also referred to as broom corn millet, this grain thrives in arid regions across Asia, Australia, Africa, Europe, and North America.
- Pearl Millet (Bajra): Known as Bajra in India, this crop is cultivated in several states, including Gujarat, Rajasthan, Maharashtra, Uttar Pradesh, and Haryana.
- Finger Millet (Ragi): Referred to as Ragi in India, this millet is highly regarded for its nutritional content and is often incorporated into products for infants. It is a significant crop cultivated throughout India.
- Browntop Millet (Korle): Primarily grown in Karnataka and Andhra Pradesh, browntop millet is steadily gaining recognition for its health benefits.
- Foxtail Millet (Kakum/Kangni): Grown predominantly in semi-arid regions, this millet, also known as Italian millet, is well-suited to dry climates.
- Barnyard Millet (Sanwa): Cultivated in various areas of Andhra Pradesh and Uttarakhand, barnyard millet is recognized for its adaptability to diverse growing conditions.
- Little Millet (Moraiyo): This resilient millet demonstrates the ability to withstand both drought and waterlogging. It

is commonly grown in the Eastern Ghats region of India.

- Buckwheat Millet (Kuttu): Known as Kuttu in India, buckwheat is a recognized variety of millet that is rich in nutritional value.
- Amaranth Millet (Rajgira): Referred to as Rajgira, amaranth is an excellent source of protein and dietary fiber, making it an ideal component of a balanced diet.
- Kodo Millet: This millet variety is easily digestible and contains a higher concentration of the amino acid lecithin, enhancing its nutritional profile.

III. Literature Review

- Hyperspectral imaging is used for a variety of tasks [1] [2] [3] [4], to harness full potential of hyperspectral imaging, image analysis[5] and processing [6] [7] [8] and machine learning and deep learning techniques [3] [9] are used for the extraction and analysis of data like in [2], Jianchang Ren et. al provided a comprehensive exploration of various methods designed to address the high-dimensional nature of hyperspectral data which included not only traditional techniques such as Principal Component Analysis and Independent Component Analysis but also introduces innovative hybrid approaches that enhance the efficiency and accuracy of hyperspectral data processing. Building upon these techniques, D. Manolakis [10] used various detection algorithms like matched filter , adaptive matched filter, and constrained energy minimization needed for the identification and classification of the materials within hyperspectral data.
- Adding to these advancements, data mining and noise reduction are important for the improvement of the hyperspectral image processing, in [11] Qiong Dai explores the integration of data mining with HSI, resulting in the developments in methods like clustering and regression which enhance the analysis of hyperspectral data for food quality assessment and process optimization.In [12], he focussed on advanced de-noising techniques that improve the quality of hyperspectral images by reducing noise while preserving critical information. The review provides information on methods such as wavelet transforms, sparse representation, and machine learning-based approaches, emphasizing their application in the food industry to enhance the reliability of hyperspectral data for quality control and safety monitoring.
- Relevant to the study presented here, Telmo Adao et. al [13] provided a comprehensive method to integrate unmanned aerial vehicles (UAVs) with hyperspectral imaging technology in turn helping in the development of agriculture and Forestry sector. The paper provides knowledge for advancements in UAV platforms, sensor technologies, and data processing techniques that help enhance the precision and efficiency of hyperspectral imaging in these fields basically providing development in practical applications such as crop monitoring, forest health assessment, and resource management, etc.
- Hyperspectral imaging in the agricultural sector has shown immense potential. In [14], Bing Lu et. al highlights how HSI technology enables precise monitoring of crop health, allowing for early detection of diseases and stress factors that can affect yield and quality, helping us to address the challenges faced in that sector. The rapid advancements HSI technology has created new opportunities for precision agriculture, allowing for comprehensive analysis of crop health, soil characteristics, and weed detection. Image analysis, combined with machine learning techniques, is utilized to develop new methods for identification, inspection, and quality assessment. Moreover, significant advancements have been made in the application of HSI for the analysis and management of grains such as rice and maize.
- HSI has shown considerable potential in the varietal classification of rice seeds. There are a lot of non-destructive variety classification methods used for identifying the varieties using machine – vision and machine learning techniques. One such technique is given by samson Damilola Fabiyi in [15]. The paper explores the combined use of RGB and hyperspectral images for classifying different rice varieties. He comprehensively surveys different computer vision techniques and machine learning methods to integrate hyperspectral imaging with traditional RGB imaging can improve the accuracy and efficiency of varietal classification. The use of advanced image analysis and machine learning techniques in this context enables precise identification and sorting of rice seed varieties, contributing to better crop management and quality control in rice production. Another method for varietal classification or purity identification is provided in [16] by Weihua Liu et. al which uses LASSO Logistic Regression model which demonstrates a highly accurate method for distinguishing pure seeds from contaminants. This advancement is essential for ensuring seed quality and purity, which are critical for achieving optimal crop yields and maintaining the integrity of agricultural production.
- Another kind of seeds HSI has been proven to be efficient in for classification are maize seeds. Chao Xia et al. [17] explored the area of classification of maize seeds using hyperspectral imaging coupled with multilinear discriminant analysis. They enhanced their model by integrating a least-squares support vector machine for better feature extraction. This integration of HSI, MLDA, and LS-SVM markedly improved the accuracy of maize seed classification. Similar research has been conducted for the classification of maize seeds by integrating learning algorithms with HIS by Huan Yang [18]. Integrating machine learning algorithms with HSI yielded highly accurate varietal recognition results. Implementing these advanced algorithms enhances the accuracy of seed variety identification and supports the preservation of genetic diversity and quality in maize production.
- This is essential for breeding programs and commer-

cial agriculture, ensuring improved outcomes. Integrating these cutting-edge techniques into agricultural practices doesn't just boost efficiency, it also supports more sustainable and thoughtful farming methods.

IV. Methodology

A. Problem Statement

Traditional methods for calculating nutritional properties in millets through regression analysis are often labor-intensive, time-consuming, and require extensive manual effort and laboratory analysis. These methods are challenged by scalability and efficiency limitations, making them less suitable for the demands of modern agriculture. There is a critical need for innovative solutions that can streamline these processes, minimize manual intervention, and provide accurate results more swiftly.

Objective: Develop a robust regression model to predict the nutritional contents in 11 different varieties of millets using hyperspectral imaging. The aim is to overcome the limitations of traditional methods by leveraging advanced imaging technology to deliver accurate, efficient, and scalable results.

B. Data Acquisition

1) Sample Preparation: Make 5 gm samples for each millet variety using a weighing scale and store these samples to use them for hyperspectral imaging.
2) Imaging and Data Capture: To begin the data acquisition, first set up the hyperspectral camera and turn it on at least 45 minutes before starting the readings to ensure it stabilizes. Adjust the camera's focus to achieve optimal image clarity. Once the camera is ready, capture a pure black image and a pure white image for calibration purposes. After completing the initial calibration, proceed with the millet readings. Place up to three millet samples on the tray at a time and systematically capture spectral images of each sample. Ensure consistent imaging to maintain reliable data quality throughout the process.
3) Image Segmentation, ROI and spectral data extraction: Region of Interest (ROI) Selection- define regions of interest within the image where you want to extract and analyze spectral data. This can be done manually or using automated methods.
4) Mean Spectrum Calculation: Calculate the mean spectrum by averaging the spectral values across all pixels in the ROI for each wavelength. This involves summing the spectral values for each wavelength and dividing by the number of pixels in the ROI.
5) Model Training: We then train our model using the extracted data and the specific nutritional components of interest (e.g., protein, fat, carbohydrate). Once trained, the model is evaluated on test data to assess its performance. We use regression metrics, such as Mean Squared Error (MSE) and R2 score, to determine the accuracy and effectiveness of the model in predicting the nutritional values.

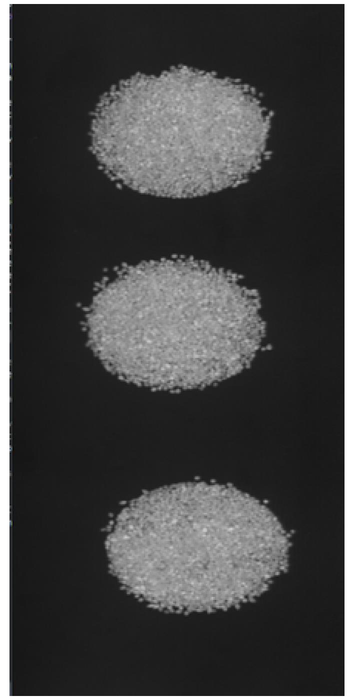


Fig. 1. Banyard at 1127.65nm

6) Data Storage: Securely store spectral data and use appropriate data management systems for the organization.
7) NIR Hyperspectral images of millet seeds were acquired were acquired in reflectance mode using laboratory - based push broom hyperspectral imaging system. Spectral Range: 900- 1700 (nm)
8) Spectral Resolution: 4.9 nm (Spectral resolution refers to the ability of an imaging system, such as a hyperspectral camera or spectrometer, to distinguish between different wavelengths of light. It indicates how finely an instrument can divide the spectrum of light into its component wavelengths.) Total Spectral Wavelengths: 168 For each millet 120 images were taken Millet varieties: 11 Total Hyperspectral images captured: 120* 11= 1320

C. Machine Learning Models Used

1) Linear Regression: Linear regression is a fundamental statistical technique employed to examine the relationship between a dependent variable and one or more independent variables. Its primary aim is to identify a linear equation that accurately reflects this relationship, facilitating both analysis and prediction. (till prediction or analysis)

2) Ridge Regression: It is also known as Tikhonov regularization, is a technique used to address some of the limitations of ordinary least squares (OLS) regression, particularly when dealing with multicollinearity (highly correlated independent variables) and overfitting. Ridge regression introduces a regularization term to the loss function, which helps to shrink the regression coefficients and reduce their variance. Ridge regression is a form of regularization that helps reduce overfitting by penalizing large coefficients. It's useful when you have a large

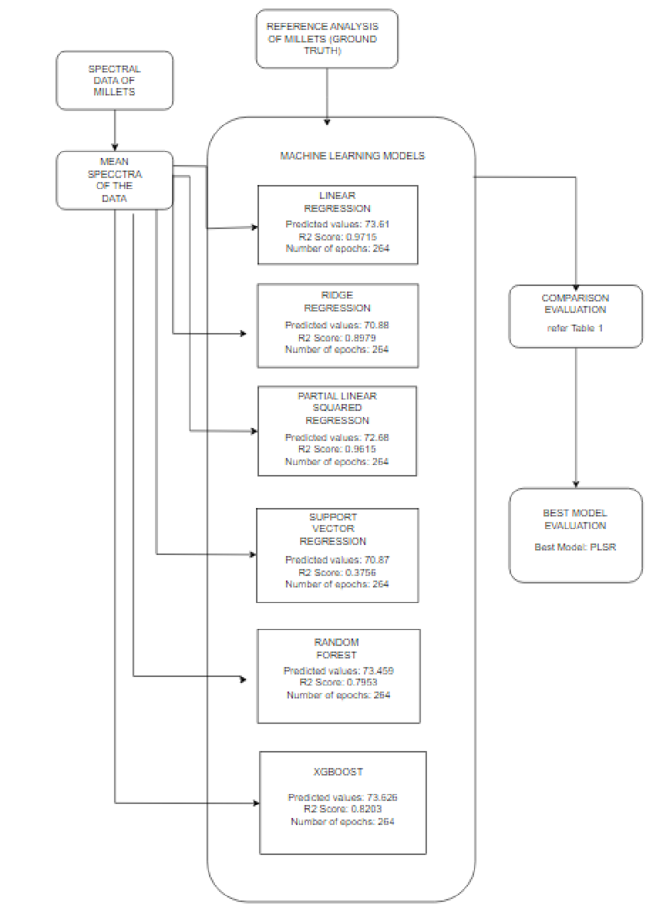


Fig. 2. Architecture Diagram

number of correlated features, as it stabilizes the predictions and avoids overfitting.

3) Support Vector Regression: Support Vector Regression (SVR) is a variant of Support Vector Machine (SVM) tailored for addressing regression challenges. Unlike traditional regression methods that focus on minimizing the discrepancy between predicted and actual values, SVR seeks to establish a line that remains within a defined margin of tolerance. This methodology emphasizes the development of a more robust and precise prediction model while allowing for minor deviations.

4) Partial Least Squares Regression: Partial Least Squares Regression (PLSR) is a sophisticated statistical technique utilized to model complex relationships involving multiple dependent (response) variables and independent (predictor) variables. By integrating elements of principal component analysis (PCA) and multiple regression, PLSR proves particularly effective in scenarios where predictors exhibit high collinearity or when the dataset comprises more predictors than observations. This method is particularly well-suited for hyperspectral data, as it adeptly manages multicollinearity (highly correlated bands) and demonstrates strong performance, particularly in instances where the number of features (bands) exceeds the number of samples.(till observations)

5) Random Forest: Random Forest handles non-linearity and interactions between bands very well and is robust against overfitting due to its ensemble nature (multiple decision trees). It provides feature importance, which can help identify the most important spectral bands for predicting carbohydrate content.

6) XGBoost: XGBoost is an advanced boosting algorithm that often outperforms other models in regression tasks, especially when dealing with large amounts of data. It combines many weak learners (decision trees) to make stronger predictions. It handles missing data and outliers well and is efficient for high-dimensional data.

V. Results

A. Linear Regression

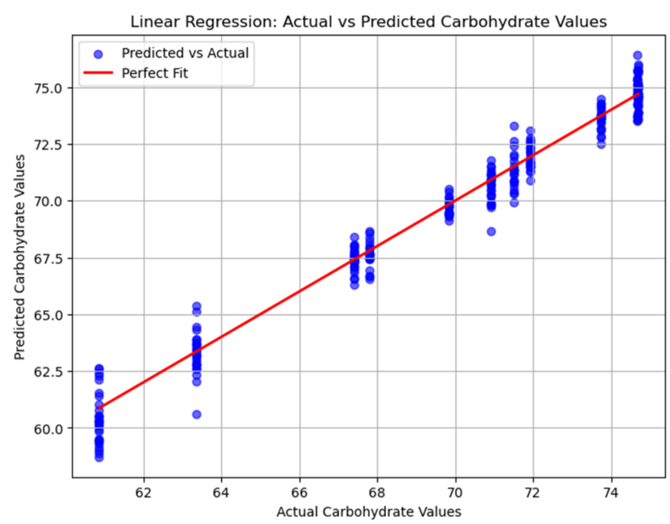


Fig. 3. Line Plot: Actual vs Predicted Carbohydrate Values

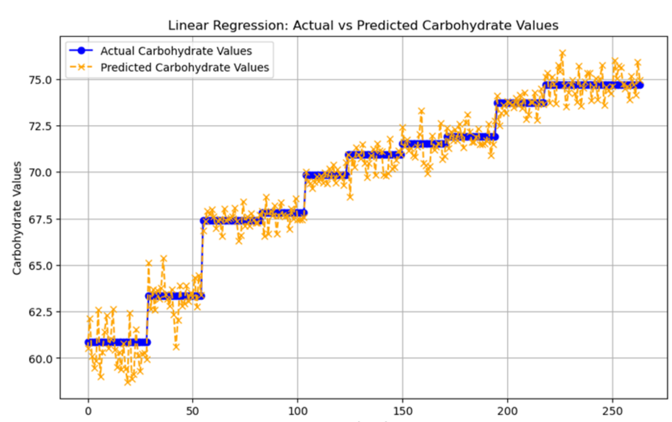
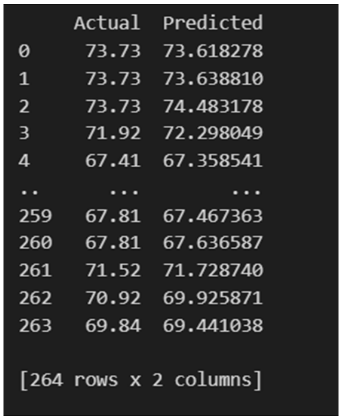


Fig. 4. Scatter Plot: Actual vs Predicted Carbohydrate Values



	Actual	Predicted
0	73.73	73.618278
1	73.73	73.638810
2	73.73	74.483178
3	71.92	72.298049
4	67.41	67.358541
..	...	...
259	67.81	67.467363
260	67.81	67.636587
261	71.52	71.728740
262	70.92	69.925871
263	69.84	69.441038

[264 rows x 2 columns]

Fig. 5. Mean Squared Error and R2 Score

B. Ridge Regression

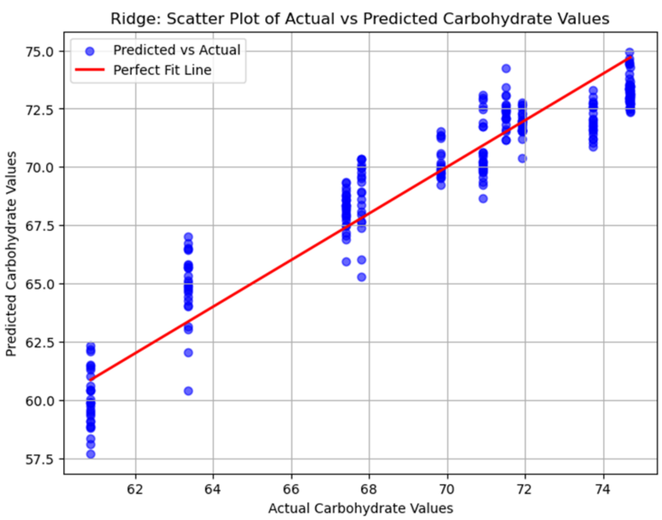


Fig. 6. Line Plot: Actual vs Predicted Carbohydrate Values

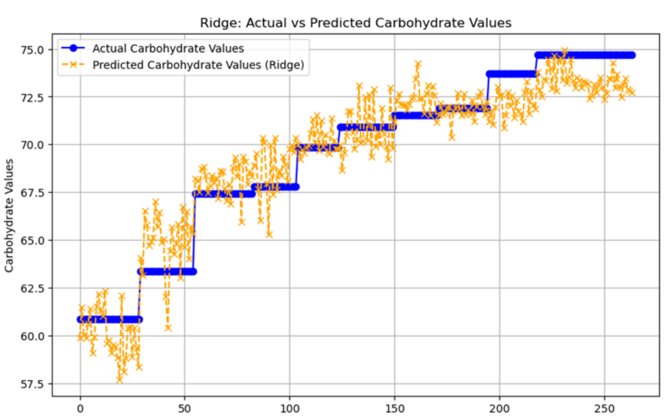
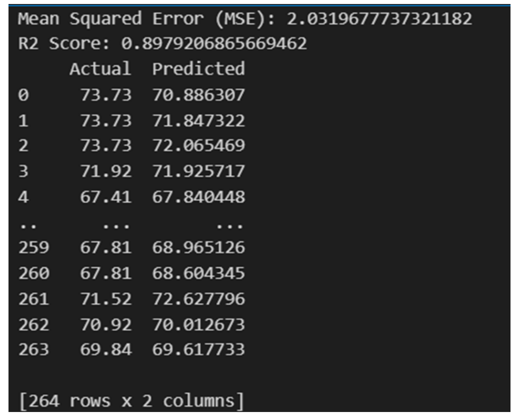


Fig. 7. Scatter Plot: Actual vs Predicted Carbohydrate Values



Mean Squared Error (MSE): 2.0319677737321182
R2 Score: 0.8979206865669462

	Actual	Predicted
0	73.73	70.886307
1	73.73	71.847322
2	73.73	72.065469
3	71.92	71.925717
4	67.41	67.840448
..	...	...
259	67.81	68.965126
260	67.81	68.604345
261	71.52	72.627796
262	70.92	70.012673
263	69.84	69.617733

[264 rows x 2 columns]

Fig. 8. Mean Squared Error and R2 Score

C. Support Vector Regression

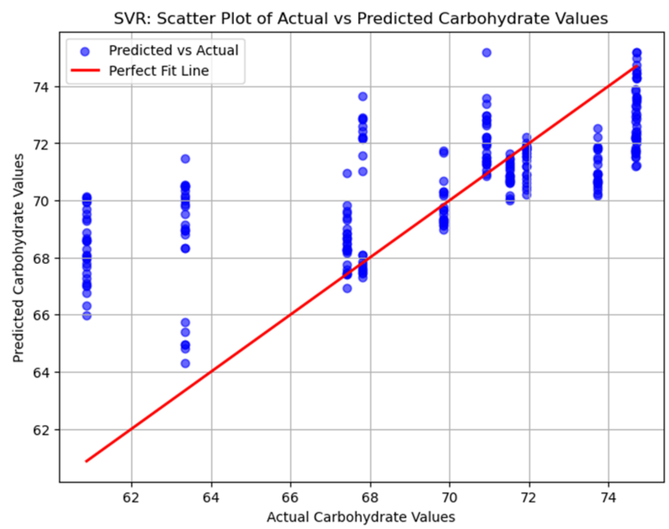


Fig. 9. Line Plot: Actual vs Predicted Carbohydrate Values

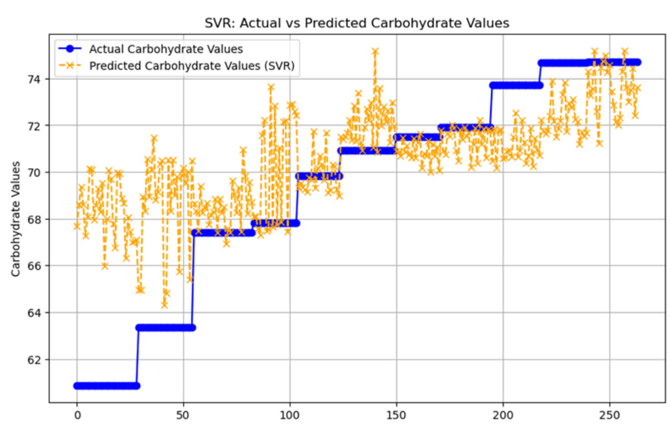


Fig. 10. Scatter Plot: Actual vs Predicted Carbohydrate Values

H

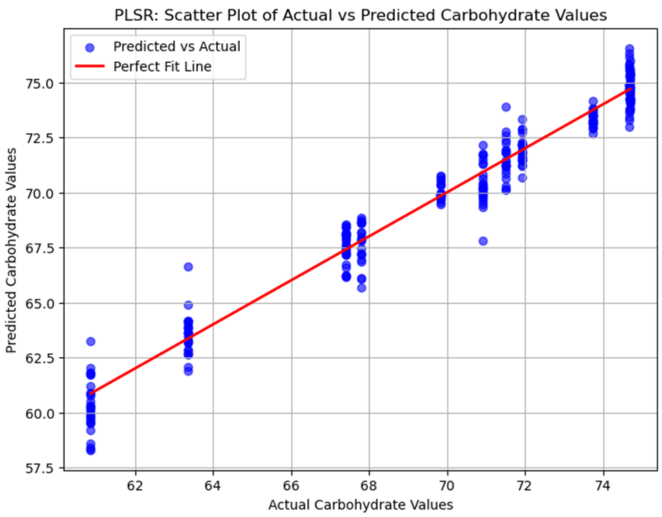


Fig. 12. Line Plot: Actual vs Predicted Carbohydrate Values

```
     Actual  Predicted
0     73.73  70.878576
1     73.73  70.728279
2     73.73  70.961158
3     71.92  71.525862
4     67.41  67.760771
..      ...        ...
259   67.81  72.170922
260   67.81  71.036219
261   71.52  70.669851
262   70.92  71.966693
263   69.84  69.120507

[264 rows x 2 columns]
Mean Squared Error (MSE): 12.428822059479657
R² Score: 0.3756172519001091
```

Fig. 11. Mean Squared Error and R2 Score

D. Partial Least Squares Regression

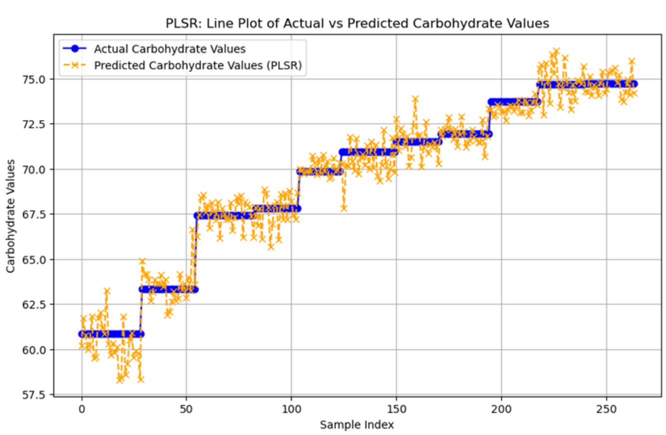


Fig. 13. Scatter Plot: Actual vs Predicted Carbohydrate Values

```
Mean Squared Error (MSE): 0.765455966879191
R2 Score: 0.961546034060005
     Actual  Predicted
0     73.73  72.682931
1     73.73  73.541151
2     73.73  74.136607
3     71.92  72.869120
4     67.41  68.115174
..      ...        ...
259   67.81  67.210833
260   67.81  68.575171
261   71.52  72.238893
262   70.92  70.217397
263   69.84  69.466146

[264 rows x 2 columns]
```

Fig. 14. Mean Squared Error and R2 Score

E. Random Forest

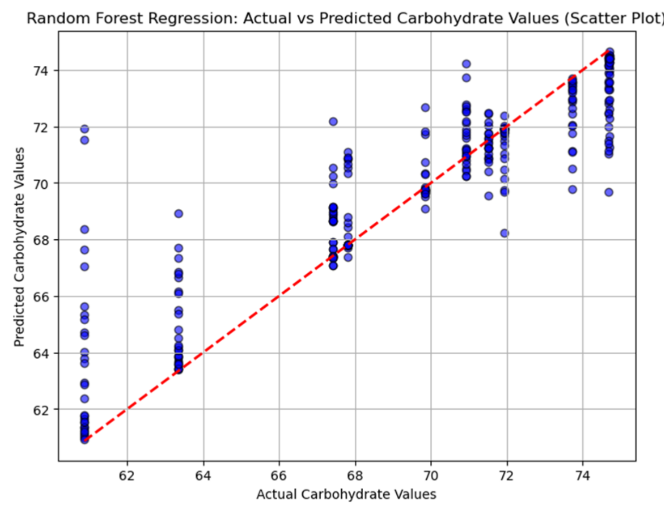


Fig. 15. Line Plot: Actual vs Predicted Carbohydrate Values

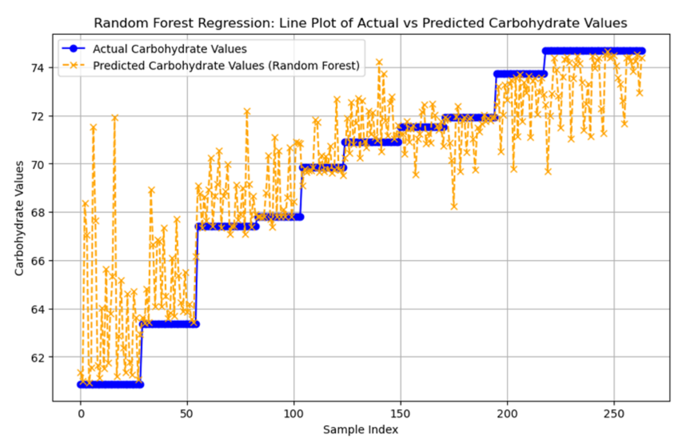


Fig. 16. Scatter Plot: Actual vs Predicted Carbohydrate Values

```
Mean Squared Error (MSE): 4.074810839303184
R2 Score: 0.7952950345852918
     Actual  Predicted
0     73.73   73.45988
1     73.73   72.88468
2     73.73   73.57352
3     71.92   68.23540
4     67.41   67.41000
..      ...        ...
259   67.81   70.68040
260   67.81   68.10692
261   71.52   70.40608
262   70.92   71.77808
263   69.84   69.62508

[264 rows x 2 columns]
```

Fig. 17. Mean Squared Error and R2 Score

F. XGBoost

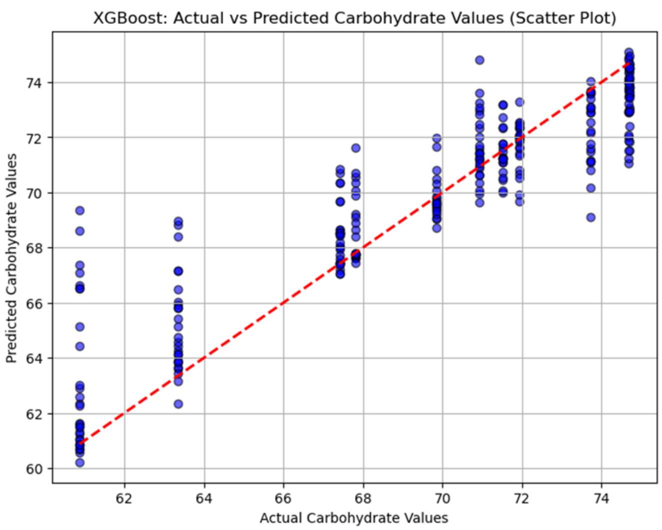


Fig. 18. Line Plot: Actual vs Predicted Carbohydrate Values

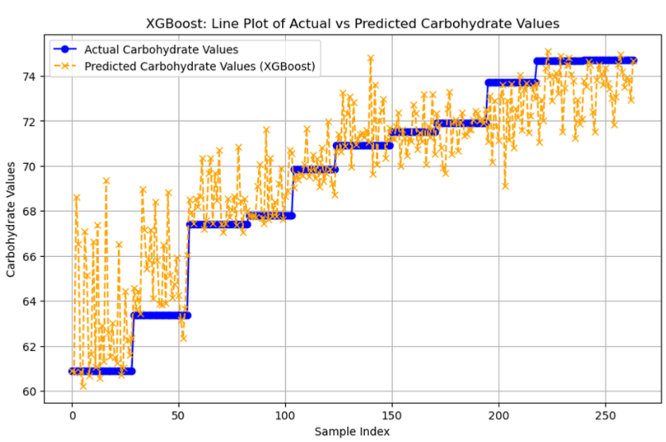


Fig. 19. Scatter Plot: Actual vs Predicted Carbohydrate Values

```
Mean Squared Error (MSE): 3.576331938939567
R² Score: 0.820336959249541
     Actual  Predicted
0     73.73  73.625999
1     73.73  73.007675
2     73.73  73.568878
3     71.92  69.669876
4     67.41  67.182922
..      ...        ...
259   67.81  69.905251
260   67.81  68.384232
261   71.52  71.753654
262   70.92  70.061356
263   69.84  69.337006

[264 rows x 2 columns]
```

Fig. 20. Mean Squared Error and R2 Score

REGRESSION MODEL	MEAN SQUARED ERROR	R2 SCORE
Linear	0.567	0.9715
Ridge	2.03	0.8979
Support Vector	12.43	0.3756
PLSR	0.7654	0.9615
Random Forest	4.07	0.7953
XGBoost	3.576	0.8203

TABLE I
COMPARISON OF VARIOUS REGRESSION MODELS

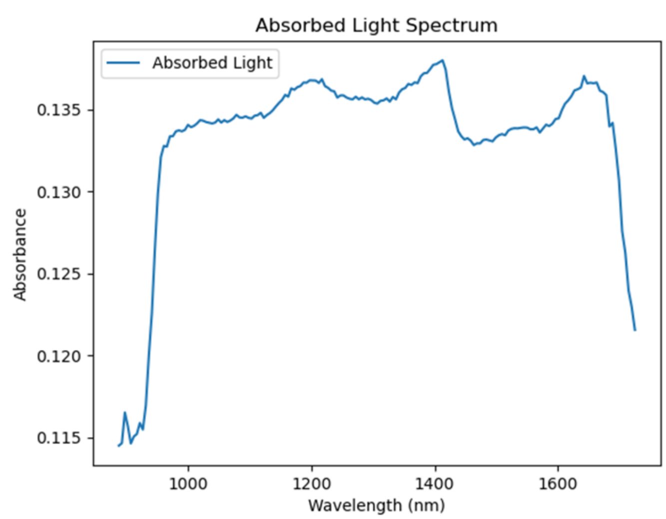


Fig. 21. Absorbed Light Curve

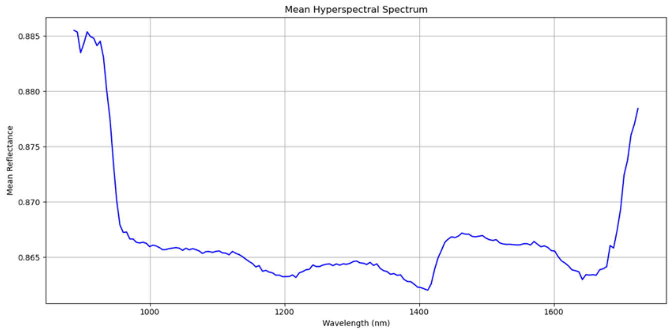


Fig. 22. Spectral Reflectance Curve

VI. Conclusion

The provided plot depicts a spectral reflectance curve, illustrating how the millet sample reflects light across various wavelengths, measured in nanometers. The x- axis represents the wavelength, while the y-axis shows the reflectance values. Key features of the plot include noticeable downward peaks around the 920nm ,1400 nm and 1700nm. These variations highlight specific absorption features of the millet. The overall trend indicates that reflectance decreases as the wavelength increases, which can help identify particular chemical bonds or functional groups in the millet samples. Some noise is evident in the plot, suggesting the need for data preprocessing to smooth the curve for more precise analysis. This spectral data is crucial for identifying the unique wavelengths absorbed or reflected by different millet varieties, aiding in the development of regression models. Additionally, it can correlate specific spectral features with the chemical composition or other properties of the millet, providing valuable insights for agricultural and nutritional studies.

References

[1]] Hyperspectral Image Processing for Automatic Target Detection Applications Dimitris Manolakis, David Marden, and Gary A. Shaw

[2] J. Ren, J. Zabalza, S. Marshall and J. Zheng, "Effective Feature Extraction and Data Reduction in Remote Sensing Using Hyperspectral Imaging [Applications Corner]," in IEEE Signal Processing Magazine, vol. 31, no. 4, pp. 149-154, July 2014, doi: 10.1109/MSP.2014.2312071. keywords: Principal component analysis;Covariance matrices;Feature extraction;Remote sensing;Memory management;Support vector machines;Hypercubes

[3] Deep Learning Applications for Hyperspectral Imaging: A Systematic Review Akin Ozdemir1 , Kemal Polat2

[4] Hyperspectral Image Processing and Analysis System (HIPAS) and Its Applications Blng Zhang, Xlanglun Wang, Jlangul Uu, Lanfen Zheng, and Qlngxl Tong

[5] Data handling in hyperspectral image analysis : James Burger a, Aoife Gowen b

[6] Hyperspectral document image processing: Applications, challenges and future prospects Rizwan Qureshi a , b , Muhammad Uzair c , Khurram Khurshid b , Hong Yan a

[7] Recent advances in techniques for hyperspectral image processing : Antonio Plaza a, Jon Atli Benediktsson b, Joseph W. Boardman c, Jason Brazile d, Lorenzo Bruzzone e,Gustavo Camps-Valls f, Jocelyn Chanussot g, Mathieu Fauvel g,b, Paolo Gamba h, Anthony Gualtieri i,j,Mattia Marconcini e, James C. Tilton i, Giovanna Trianni h

[8] Advances in Feature Selection Methods for Hyperspectral Image Processing in FoodIndustry Applications: A Review : Qiong Daia, Jun-Hu Chenga, Da-Wen Suna, Xin-An Zenga

[9]] H. Petersson, D. Gustafsson and D. Bergstrom, "Hyperspectral image analysis using deep learning — A review," 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), Oulu, Finland, 2016, pp. 1-6, doi: 10.1109/IPTA.2016.7820963. keywords: Feature extraction;Machine learning;Hyperspectral imaging;Neural networks;Kernel;Hyperspectral Imaging (HSI);Deep Learning;Auto-Encoder (AE);Stacked Auto-Encoder (SAE);Convolutional Neural Network (CNN);feature representation

[10] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," in IEEE Signal Processing Magazine, vol. 19, no. 1, pp. 29-43, Jan. 2002, doi: 10.1109/79.974724. keywords: Detection algorithms;Hyperspectral imaging;Composite materials;Signal processing algorithms;Chemical sensors;Reflectivity;Laboratories;Sensor phenomena and characterization;Atmosphere;Remote sensing

[11] Recent Advances in Data Mining Techniques and Their Applications in Hyperspectral Image Processing for the Food Industry : Qiong Dai, Da-Wen Sun, Zhenjie Xiong, Jun-Hu Cheng, and Xin-An Zeng

[12] Recent Advances in De-Noising Methods and Their Applications in Hyperspectral Image Processing for the Food Industry : Qiong Dai, Da-Wen Sun, Jun-Hu Cheng, Hongbin Pu, Xin-An Zeng, and Zhenjie Xiong

[13] Adão, T.; Hruška, J.; Pádua, L.; Bessa, J.; Peres, E.; Morais, R.; Sousa, J.J. Hyperspectral Imaging: A Review on UAV-Based Sensors, Data Processing and Applications for Agriculture and Forestry. Remote Sens. 2017, 9, 1110.

[14] Lu, Bing, Phuong D. Dao, Jiangui Liu, Yuhong He, and Jiali Shang. 2020. "Recent Advances of Hyperspectral Imaging Technology and Applications in Agriculture" Remote Sensing 12, no. 16: 2659.

[15] Varietal Classification of Rice Seeds Using RGB and Hyperspectral Images SAMSON DAMILOLA FABIYI 1, (Student Member, IEEE), HAI VU 2, CHRISTOS TACHTATZIS 1, PAUL MURRAY 1, DAVID HARLE 1, (Member, IEEE), TRUNG KIEN DAO 2, IVAN ANDONOVIC 1, (Senior Member, IEEE), JINCHANG REN 1, (Senior Member, IEEE), AND STEPHEN MARSHALL 1

[16] Liu,W.; Zeng, S.;Wu, G.;Li, H.; Chen, F. Rice Seed Purity Identification Technology Using Hyperspectral Image with LASSO Logistic Regression Model. Sensors 2021, 21, 4384.

[17] Maize seed classification using hyperspectral image coupled with multilinear discriminant analysis : Chao Xia1, Sai Yang1, Min Huang*1, Qibing Zhu1, Ya Guo1, Jianwei Qin2

[18] Recognition of maize seed varieties based on hyperspectral imaging technology and integrated learning algorithms : Huan Yang1,2,3, Cheng Wang1,2,3, Han Zhang1,2, Ya'nan Zhou1,2 and Bin Luo1,2