A NOVEL APPROACH TO COMBINE NIR AND IMAGE FEATURES FOR NON-DESTRUCTIVE ASSAY OF INDIAN WHEAT VARIETIES

Dr. A. Anne Frank Joe[1], A. Veeramuthu[2*], Dr. K. Ashokkumar[3]

[1]Assistant Professor, Department of Electronics and Instrumentation, Sathyabama Institute of Science and Technology, Tamil Nadu, India

[2]Assistant Professor, Department of Information and Technology, Sathyabama Institute of Science and Technology, Tamil Nadu, India

[3]Associate Professor, Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Tamil Nadu, India

Email: annefrankjoe@gmail.com[1], aveeramuthu@gmail.com[2*] (corresponding author), ashokkumark.cse@gmail.com[3]

DOI: https://doi.org/10.22452/mjcs.sp2022no1.6

ABSTRACT

Near InfraRed Spectroscopy (NIRS) based techniques have evolved tremendously and are being perfected over ages to be applied in a wide variety of applications. This study focuses on the selection of optimum classification algorithms, as an automated variety identifier suitable for wheat grains based on the statistical performance indices for the quality analysis and variety classification of wheat grains. NIRS was used to non-destructively determine protein, carbohydrate, ash and moisture content of wheat grains. Structural analysis focuses on the visualization aspect of the wheat grains such as the shape, size (learnt from the length, width, and height), colour and glossiness of the seed coat. In addition to the spectral information, the image derived characteristics are incorporated into the classification models to further enhance the variety identification of 10 varieties of whole wheat samples UP 262, Samba, RR 21, 343, Super sitwa, Punjab, Ankurkedar, Super 303, Pusa 360, PBW 502. Varietal purity of wheat grains is a significant factor to be considered before the milling process. The results clearly reveal that the proposed selective wavelength-based prediction algorithms and selection of limited individual quality parameters, using improved methods to extract these features has aided with the success of classification performed in this work. The proposed novel approach proves that collaborating the selected spectral features and image features further enhances the effectiveness of this work.

Keywords: Near-infrared spectroscopy, Image analysis, SVM, Random Forrest, LDA

1.0 INTRODUCTION

As a staple food for about 35% of the world's population, wheat provides the richest calories and protein in the diet than any other crop [1][2]. Worldwide the demand for wheat is on the rise owing to its peculiar characteristics of viscoelasticity and adhesiveness offered by gluten proteins, which facilitates the production of processed foods [3]. Grain analysis and maintaining varietal purity is vital at every stage of wheat handling. The speed and accuracy of the key constituent estimation in wheat grain is a prerequisite for consistent and profitable grain trading [4]. Wheat can be analysed in two ways: compositional and structural analysis. The former measures the internal features such as oil, starch, protein etc. and the latter measures the visual appearance like size, texture and colour. The traditional methods of quality evaluation are expert’s sensory valuation or chemical analysis to be more precise. The growing technique of non-destructive analysis using visual and near-infrared optical techniques is quickly replacing the traditional practices [5]. The major difficulty of variety identification comes from the similarity of wheat NIR spectra [6] and also the difficulty to differentiate the appearance of a certain variety. This research work intends to develop an automated, non-destructive, rapid classification algorithm to analyse selected Indian wheat varieties. Chemometric evaluation of the spectra was carried out to minimize

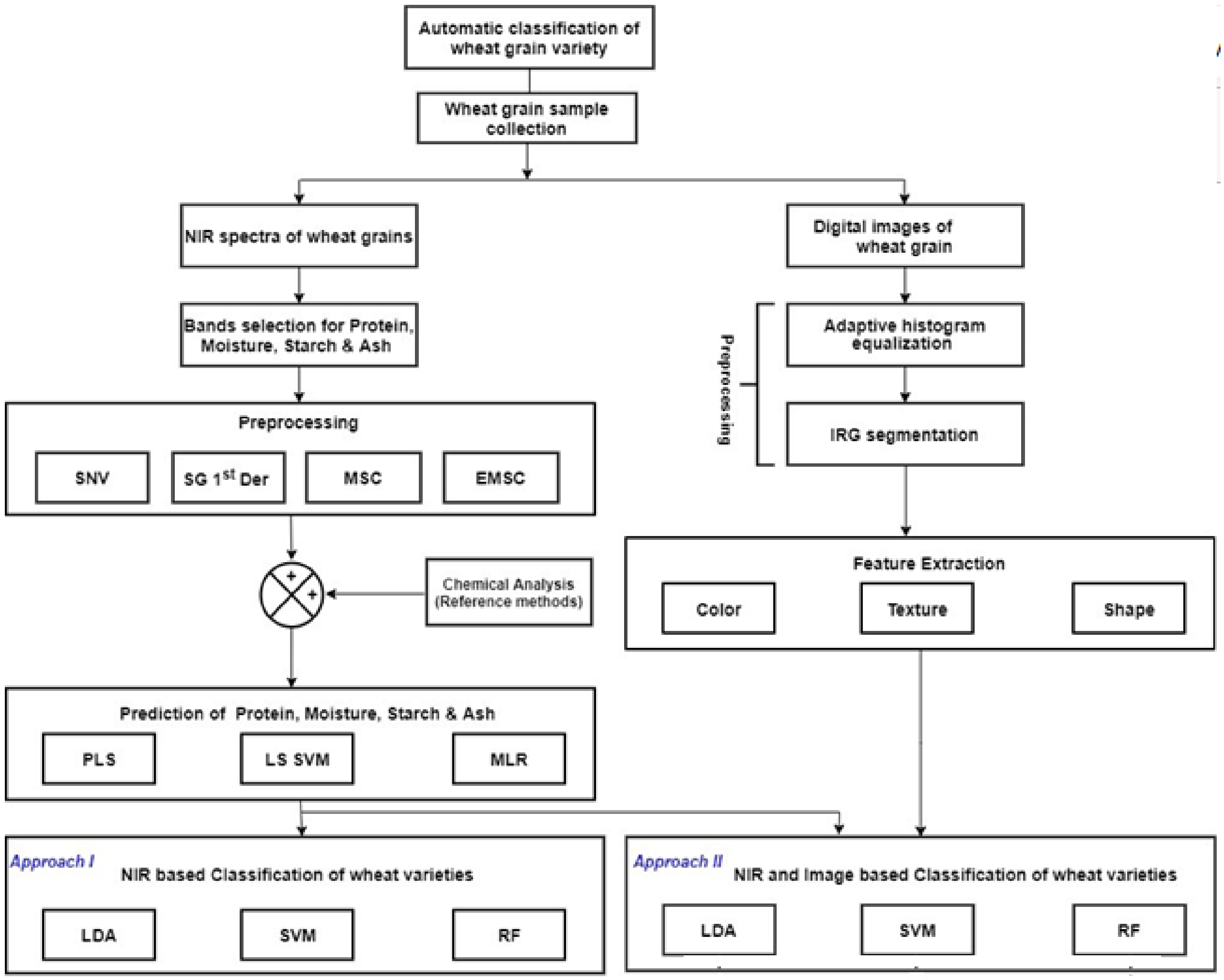

Fig. 1: Flowchart indicating scope of the proposed work

Ankurkedar, Super 303, WR-544/ Pusa Gold of common export quality that are commonly employed for producing the scattering and other inherent interferences by applying Savitzky–Golay (SG) first derivative, Standard Normal Variate (SNV) derivative, Multiplicative Scatter Correction (MSC) and Extended Multiplicative Scatter Correction (EMSC). Selective contributing wavelengths were used to develop models to predict the key components. Aiding the compositional assay of 10 selected wheat varieties, the morphological image features of the wheat grains were employed to compare the results of variety classification models such as Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), and Random Forest.

2.0 MATERIALS AND METHODS

A total of 122 wheat samples from different states of India, cultivated between the year 2015 to 2017 were collected and stored for analysis. 10 known varieties namely UP262, Samba, SONALIKA rr-21, pbw343, GW-173, punjab wheat seeds, Ankurkedar, Super 303, WR-544/ Pusa Gold of common export quality that are commonly employed for producing chapati, noodles and other pastries were among them. FOSS XDS NIR SmartProbeTM Analyzer was used to collect spectral details of wavelength Vs absorbance for the NIR region 400–2500 nm at 0.5-nm data point intervals. A digital camera was used to obtain images of wheat samples. Wheat samples were manually cleaned to remove damaged grains and images of non-touching grains on a black background were obtained. Figure 1 shows detailed scope of the proposed work. Two parallel operations are performed on the wheat grains to capture the structural and the compositional information using Spectral and image analysis respectively. Noise elimination and feature extractions is performed in both the methods and the wheat varieties are classified using the combination of both the spectral and image derived features.

2.1 Processing of Spectra and Multivariate Analysis

Near Infrared spectroscopy has a good reliability in the industry where it had been employed for the examination of grains for about 20 years. A rapid analysis with no sample preparation and no destruction of wheat grains has significant commercial appeal. This phenomenon makes NIR spectroscopy method of estimating the key constituents in wheat grains gain greater recognition and application in the milling and baking industries. There is a treasure of information that can be unveiled from the peaks and troughs of the NIR spectra. The ability to be operated by non-technical personnel with very high precision, reproducibility and near instantaneous results are added advantages.

Grains that are measured using NIR based devices, often exhibit substantial differences in the spectra caused by potential physical variations [6]. The variations in the presented sample would be because of non-homogeneous distribution of the particles, particle size distribution, changes in refractive index, and sample morphology (surface roughness/shape). Pre-processing of spectral data is an integral part that cleans up the variations and consequently building a good and robust multivariate regression and classification model [7]. Results based on the previous work on 165 near infrared reflectance spectra of whole wheat grain samples, showed that EMSC performed better than the trialled MSC, SNV, SG smoothing and SG derivative [8].

2.2 Selection of Wavelengths

The collected spectra were analysed with Principal Component Regression (PCR) with 7 principal components (PC). The PCR model was built with Non-linear Iterative Partial Least Squares algorithm (NIPALS), the data were mean centred before being applied to the model. Cross validation was performed by segregating the data into 10 segments with two samples per segment. This activity is performed to locate the spectral bands (form the entire NIR spectral region, 700nm to 2350nm) that holds the most information about the selected components under study.

Regression coefficients can be used to discuss the contributions of individual wavelengths to a calibration model, since a regression coefficient spectrum (Figure 2) shows characteristic peaks and troughs that can indicate which wavelength range is important for the calibration model. The significantly contributing wavelengths identified in the regression coefficients plot (Figure 2) were observed at 1440-1470nm, 1940nm, 970-1020nm, 1200nm, 2200nm, 2100nm, 1450nm, 1540nm,1690nm, 1720nm, 990nm 2336nm and 2345nm. The unveiled characteristic regions holding the maximum information on Moisture, Protein, Starch and Ash molecules closely coincides with the spectral wavelength identified by previous researchers in the wheat spectra [9].

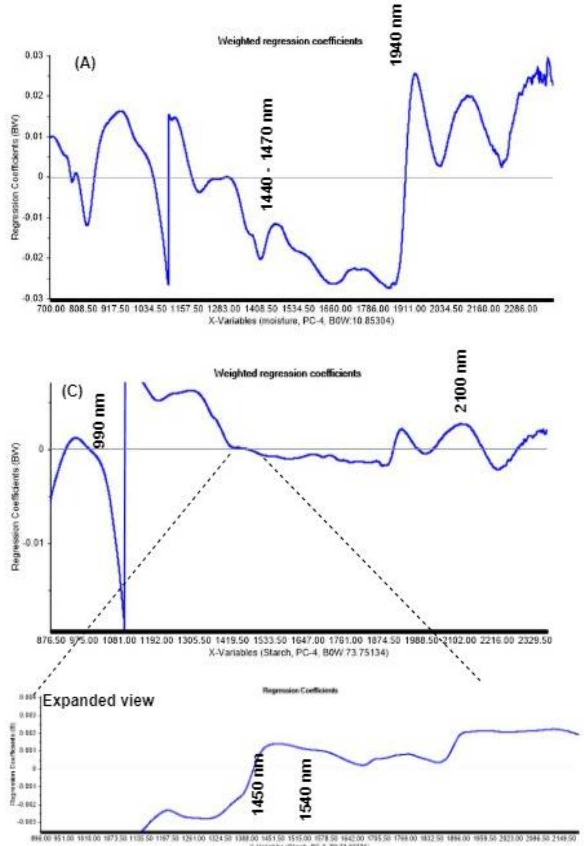

Fig. 2: Regression coefficients of the (A) moisture, (B) Protein and (C) Starch (D) Ash for the PCR models.

The chief challenge in decoding the wheat NIR spectra for variety identification lies on the similarity of the spectral pattern across varieties. Spectral differences between wheat of different varieties are very slight in comparison with the variations due to the effect of granularity [10]. Chemometric modeling is used for multivariate analysis. Prediction models such as PLS, MLR and LS SVM are used to statistically relate the absorbance value of the NIR spectrum with the constituent concentration and confident results were inferred from the R^2 coefficient of 0.976 to 0.999 and RMSECV of 0.029 to 0.405 [11]. The results indicate feasibility of practical application of NIRS in wheat quality profiling. The above mentioned four constituent concentrations in association with extracted image features is used to variety identify the samples.

2.3 Processing of Digital Images

Wheat image processing offers an objective and quantitative evaluation of morphological characteristics to mathematically extract useful information from the grain images. The geometric, shape, color and texture measurements of wheat grains in combination with the spectral features are used to create an enhanced variety identification model [12]. The acquired images were filtered for brightness corrections, grayscale transformations, focus adjustments, contrast enhancement and noise removal [13][14]. An Improved Region Growing (IRG) method is executed on the images to segment the individual wheat grains. The additional constraint of direction matching along with the traditional intensity matching based region growing method preserves the shape of the grain more accurately with not much significant holes in the segmented image. The descriptive features such as color, shape and texture were used to study the profile of the wheat variety. To quantify the texture of the wheat grains, descriptors namely: GLCM - Gray Level Co-occurrence Matrix and EOTM - Edge Oriented

Text on Map were investigated. In EOTM, to obtain maximum information of edge, orientation of edge is derived along vertical, horizontal and diagonals.

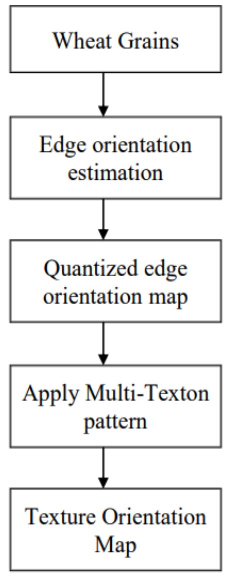

Fig. 3: Framework of the edge-oriented text on map (EOTM)

The captured wheat images are combination of irregular and uniform texture patterns. In such a case, the chances of getting keen information are less with two-directional edge information. In order to eliminate these issues, an edge oriented texton map (EOTM) is suggested.

In EOTM, to obtain maximum information of edge, orientation of edge is derived along vertical, horizontal and diagonals. Figure 3 shows the steps involved in obtaining the EOTM. Canny operator is used here to extract vertical and horizontal information of edge, while the Sobel operator is worked along both diagonals for edge detection [15].

Diagonal edges are extracted using equation 1. Let GD1 and GD2 represent gradient information obtained by the 45◦ and 135◦ diagonal Sobel filter respectively. When the unit vectors are given as $\widehat{\mathrm{SD}}_1$ and $\widehat{\mathrm{SD}}_2$, the angle between them is calculated as:

$$\cos\,(\widehat{SD_1,SD_2}) = \frac{SD_1 \,.SD_2}{|SD_1|\,.|SD_2|} \qquad 1$$

Diagonal edge orientation map $\mathrm{Orimap}_{SD_1,SD_2}$ is expressed in equation 2 as:

$$\mathrm{Orimap}_{SD_1,SD_2} = \arccos\left[\cos\,(\widehat{SD_1,SD_2})\right]$$
$$= arccos\left[\frac{SD_1 \,.SD_2}{|SD_1|\,.|SD_2|}\right] \qquad 2$$

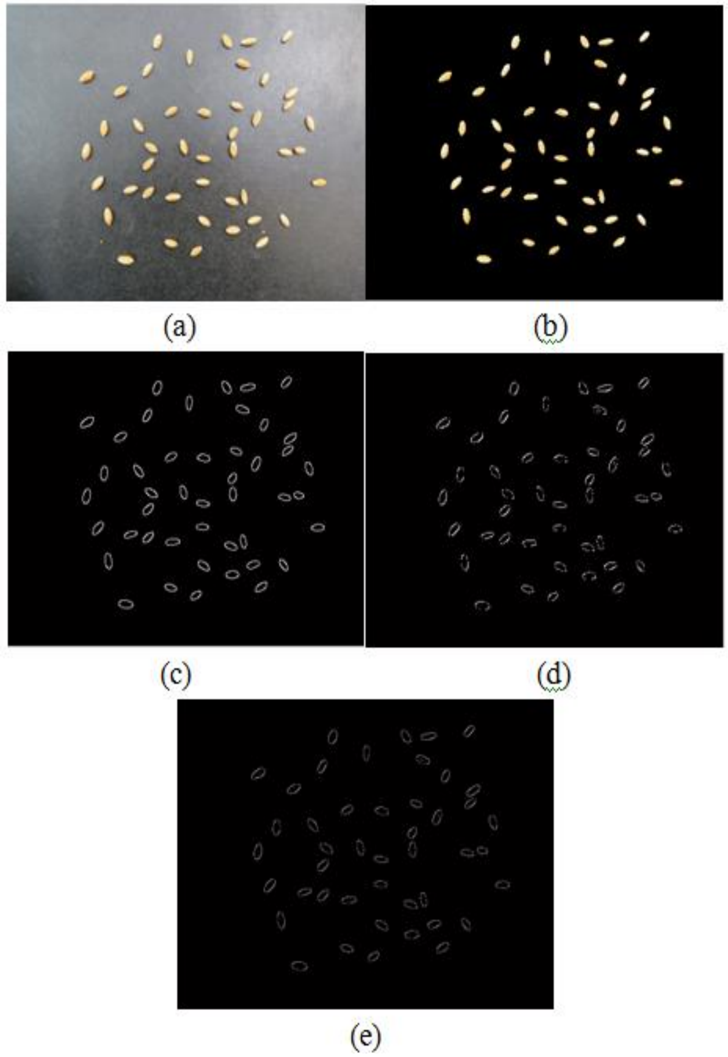

Fig. 4: (a) Input Image (b) Segmented Image (c) Diagonal edge orientation map (d) Horizontal and vertical edge orientation map (e) Quantized orientation map

where $SD_1 . SD_2 = GD_1 + GD_2$ and $|SD_1| . |SD_2| = \sqrt{GD_1^2 + GD_2^2}$. Horizontal and vertical edge orientation map $\mathrm{Orimap}_{CH_1,CH_2}$ is captured using the detector called as Canny edge detector. By applying non-maximum suppression, Canny operator avoids false edge detection. Furthermore, this operator works a preprocessing step with the application of Gaussian filter (with μ = 0 and σ = 1) reduces the content of noise to maintain reliable edge information. Likewise, Sobel operator is applied along the diagonals to extract the high change rate with susceptibly lower noise.

$$Q_\theta(x, y) = Avg\left(\mathrm{Orimap}_{SD_1,SD_2} + \mathrm{Orimap}_{CH_1,CH_2}\right) * M_0 /180 \quad 3$$

The orientation is quantized uniformly into M bins. Equation 3 is used to obtain the quantized edge orientation map, denoted as $Q_\theta(x, y) = \alpha, \alpha\varepsilon\{0,1, \ldots, M_0 - 1\}$ and M_0 = 8.

Figure 4 illustrates the application of a 2*2 texton mask on a segmented image of the wheat variety ‘Punjab seed’ in horizontal, vertical and diagonal direction and the resulting orientation map that is obtained.

From the texture orientation map many valuable quantitative first order features of statistical data can be derived, namely, variance (E_4), kurtosis (E_2), skewness (E_3), entropy (E_5), mean (E_1) and skewness (E_3) and are presented in equation 4 – 8.

$$E_1 = \sum_{k=1}^{M} kQp(k) \quad (4)$$

$$E_2 = \sigma^{-4}.\sum_{k=1}^{M}(k - E_1)^4.Qp(k) - 3 \quad (5)$$

$$E_3 = \sigma^{-3}.\sum_{k=1}^{M}(k - E_1)^3.Qp(k) \quad (6)$$

$$E_4 = \sum_{k=1}^{L}(k - E_1)^2 Qp(k) \quad (7)$$

$$E_5 = -\sum_{k=1}^{M} Qp(k)\log_2 Qp(k) \quad (8)$$

3.0 RESULTS AND DISCUSSION

Two set of features are derived, one from the NIR spectral data and the other from the images of wheat grains. The classifiers such as Random Forest, Linear Discriminant Analysis and Support Vector Machine are applied to work on the first set of NIR derived features and then executed with additional features being added from the morphological derived parameters to perform wheat classification. The performance of the classifiers is compared.

3.1 Linear Discriminant Analysis (LDA)

A method used for pattern recognition works on linear combinations of variables which best defines the data[16].Linear Discriminant Analysis Multi-classification resembles analysis of variance (ANOVA) and regression analysis, where one dependent variable is expressed as a linear combination of other features under consideration. LDA has a categorical dependent variable and continuous independent variables that are assumed to be normally distributed, which is a fundamental assumption of the discriminant analysis. Multi-class LDA is based on the analysis of two scatter matrices: within-class scatter matrix and between-class scatter matrix.
x_1…..x_n are the set of samples, and their class labels y_1…y_n
S_w is the **within-class scatter matrix** calculated as:

$$S_w = \sum_j (x_j - \mu_j)(x_j - \mu_j)^T \quad (9)$$

Here j is the no. of classes, in our case it is 10 different varieties of wheat. xj is the vector value per occurrence in each class. It is desirable to have $(x_j - \mu_j)$ as a small value that would indicate closer datapoints within the class, the simpler it is to discriminate this class from the others.

The between-class scatter matrix is computed by the following equation.

$$S_b = \sum_j N_j(\mu_j - \mu_3) * (\mu_j - \mu_3)T \quad (10)$$

µ3 is the mean of all data in data set. µj is the mean vector of class j.

Then, multi-class LDA can be formulated as an optimization problem to find a set of linear combinations that magnifies the ratio between S_b and S_w as expressed in equation 9 and 10.

Table 1: Performance of LDA classifier

Pre-processing Methods	Prediction Models	Accuracy (model worked with NIR derived features)	Accuracy (model worked with NIR+Image derived features)
SG	LS SVM	96.8	97.2
SNV		90.94	91.49
MSC		95.97	97.78
EMSC		98.39	99.6
SG	MLR	95.74	95.74
SNV		95.4	96.5
MSC		96.1	96.8
EMSC		98.7	99.4
SG	PLS	96.9	97.8
SNV		94.32	95.37
MSC		97.9	99.1
EMSC		98.1	100

Performance of LDA classifier worked with NIR features and NIR+Morphological features is shown in Table 1. Figure 5 displays the performance of LDA classifier graphically.

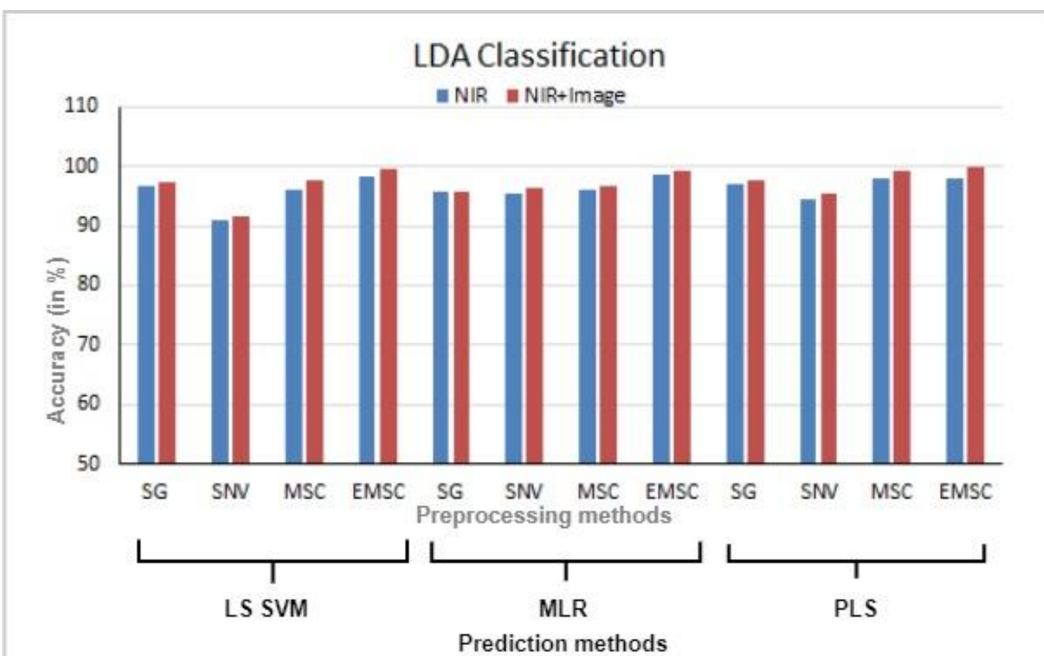

Fig. 5: Classification performance (Accuracy in %) for 12 LDA models worked with spectral data and combination of spectral with image derived features for wheat grains

3.2 SVM Multi Classification with Polynomial Kernel

Classification algorithms demand a large collection of data to perform well, Support Vector Machines (SVM) is a classification algorithm that works dependably well even with a limited amount of data. For a nonlinear classification, based on the general principle, SVM maps the original feature space (data points) to a higher-dimensional feature space H, in which the data is linearly separable. i.e. where a separating hyper-plane with maximal margin is constructed by the function X->φ (X) € H. With this new presentation the data are considered as linearly separable. i.e. the function that is usually defined as [17]

(x) = sign ($\sum yi.\ \alpha i\langle x.\ xi\rangle NSi$=1 + b) where αi is the Lagrange multiplier,

$\langle x.\ xi\rangle$ is now replaced with $K\langle x,\ xi\rangle$ = $\langle \emptyset(x),\ \emptyset(xi\ )\rangle$ i.e., the kernel function equals to the scalar product of the non-linearly mapped original features[18].

The new decision function to classify the data are: $f(x)$ = sign ($\sum yi.\ \alpha i\ .\ K\langle x,\ xi\rangle NSi$=1 + b).

Performance of 12 models (4-Preprocess methods and 3-Prediction methods) using SVM classifier worked with NIR features and NIR+Morphological features is shown in Table 2. Figure 6 displays the performance of SVM classifier graphically.

Table 2: Performance of SVM classifier

Pre-processing Methods	Prediction Models	Accuracy (model worked with NIR derived features)	Accuracy (model worked with NIR+Image derived features)
SG	LS SVM	85.11	95.74
SNV		82.98	95.74
MSC		95.74	95.74
EMSC		95.74	95.74
SG	MLR	93.62	95.74
SNV		100	95.74
MSC		100	95.74
EMSC		100	97.9
SG	PLS	100	97.87
SNV		95.74	95.74
MSC		97.87	97.87
EMSC		100	97.9

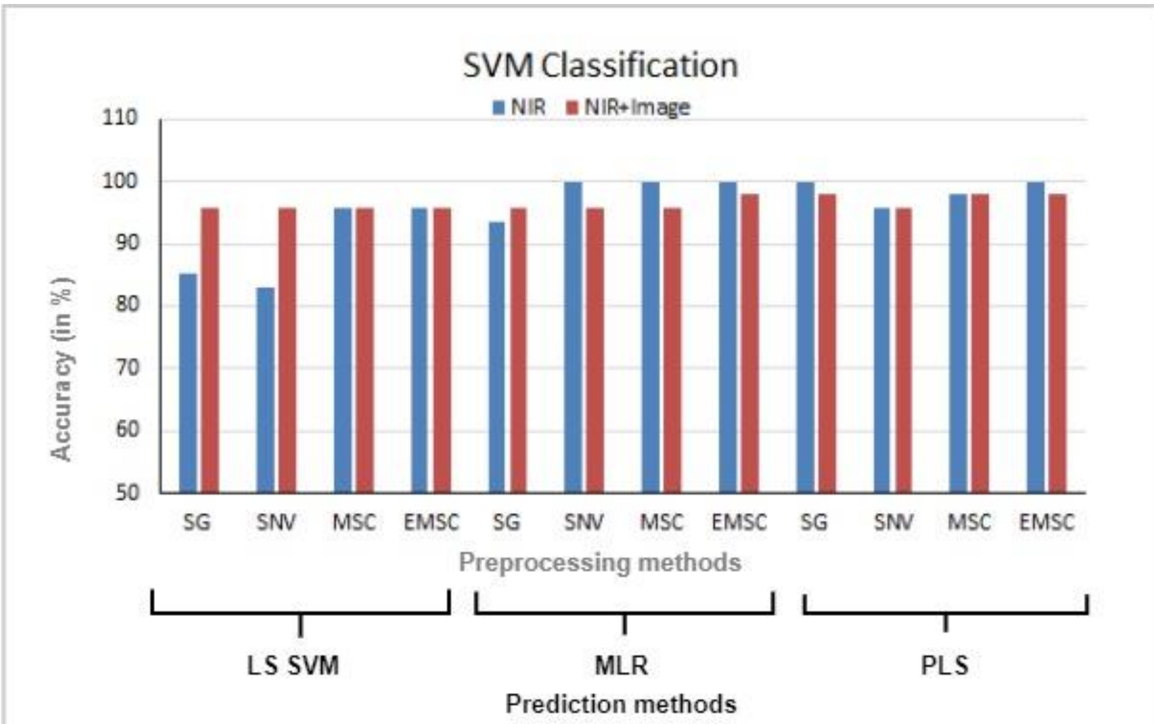

Fig. 6: Classification performance (Accuracy in %) for 12 SVM models worked with spectral data and combination of spectral with image derived features for wheat grains

3.3 Random Forest

Just as the name describes, in random forest algorithm, the model creates an entire forest of random uncorrelated decision trees to arrive at the best possible answer. Random Forest is a flexible, easy to use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time [19]. It is also one of the most used algorithms, because it’s simplicity while applying on a vast data and its appeal increases since it can be used to build predictive models for both classification and regression problems[20][21].

Performance of 12 models (4-Preprocess methods and 3-Prediction methods) using Random Forest classifier worked with NIR features and NIR+Morphological features is shown in Table 3. Figure 7 displays the performance of SVM classifier graphically.

Table 3: Performance of Random Forest classifier

Pre-processing Methods	Prediction Models	Accuracy (model worked with NIR derived features)	Accuracy (model worked with NIR+Image derived features)
SG	LS SVM	82.98	87.23
SNV		87.23	87.23
MSC		91.49	97.87
EMSC		93.62	93.62
SG	MLR	95.74	95.74
SNV		91.49	91.49
MSC		100	95.74
EMSC		91.49	93.62
SG	PLS	100	95.74
SNV		87.23	93.62
MSC		100	95.74
EMSC		91.49	91.49

The overall accuracy of all the 36 models worked with NIR spectral features and NIR+Morphological features is shown in the Figure 5. Accidental mislabeling is a subject of concern to food processors. The wheat varieties differ mutually from a physio chemical perspective and in terms of quality features. Punjab variety wheat grains are more valued by the trade because they are considered to have a fine color and perfect hardness which suits the chapati making variety and is more appreciated by the consumers, Samba variety is suitable for halwa making because of the perfect hardness and color. Commercially available wheat flour is mostly a blend of chosen varieties.

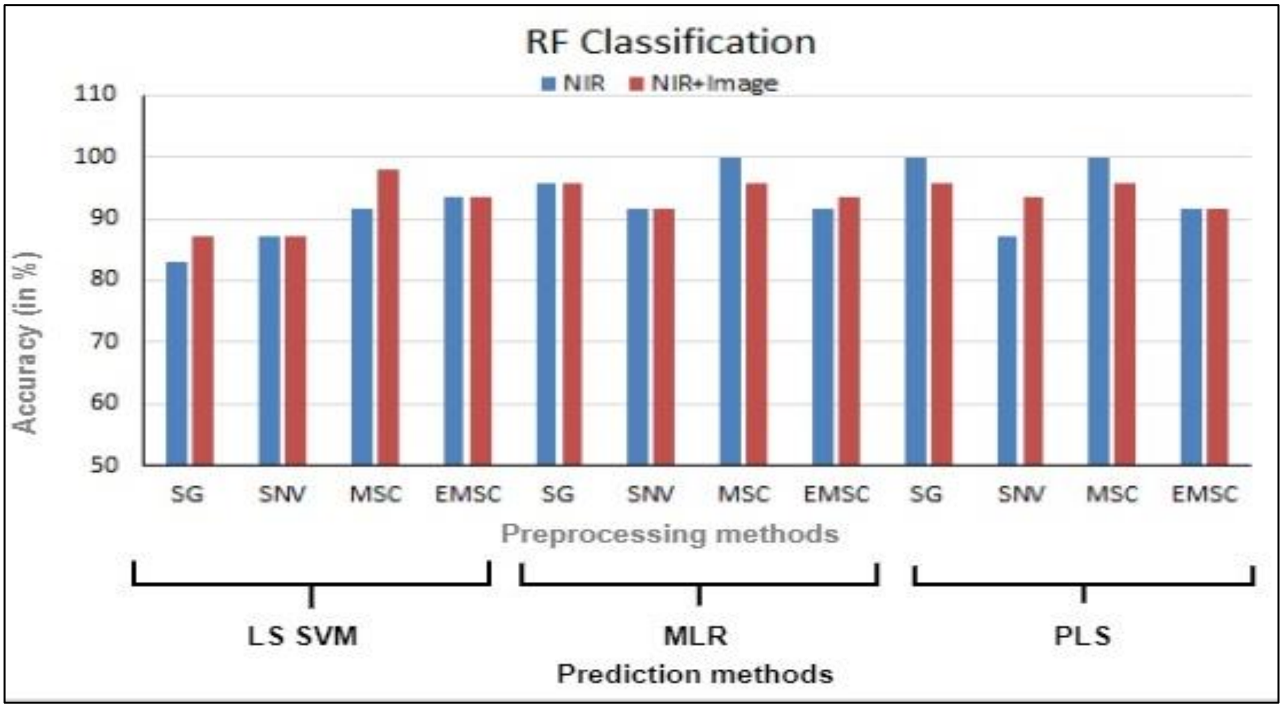

Fig. 7: Classification performance (Accuracy in %) for 12 RF models worked with spectral data and combination of spectral with image derived features for wheat grains

4.0 CONCLUSION

The current techniques of grain analysis using chemical laborious methods that is expensive on a per test basis and time consuming apart from the greater truth of being destructive is also not suitable for analyzing grains. Also, the literature reviewed reflects that although NIRS has been widely applied in the grain analysis industry worldwide, the research findings in India shows that exploratory work has been done avidly in the biomarking of the genetics of the wheat grain, studying the insect infestation and also the suitability of wheat for chapatti making in particular. Image processing on the other hand has been presently explored for the analysis of various agricultural products. The novel technique proposed here picks the features from spectral as well as images with increased accuracy of processing and providing confident results. The derived results indicate that LS-SVM aided with extended multiplicative scatter correction measured as a significant tool for analyzing the quality of wheat grain with a high accuracy. The performance comparisons of proposed method with existing methods are depicted through graphs and bar charts.

With reference to the NIRS data, although a number of wavelength regions contribute to the functional groups, selected wavelengths have proved to produced reliable results. Further work can be done to narrow down on the fingerprint wavelength region which can lead to development of handheld instrument to be used in the various stages of grain handling system.

REFERENCES

[1] A. Breiman and D. Graur, “Wheat evolution”, *Israel Journal of Plant Sciences*, Vol.43, No. 2, 1995, pp. 85-98.

[2] S. Nagarajan, “Quality characteristics of Indian wheat”, *Future of flour (Eds L Popper .W Schäfer .W Freund)*, 2006, pp.79-86.

[3] B. Belderok, H. Mesdag, DA. Donner, “Bread-Making Quality of Wheat”, *Springer*, New York, 2000.

[4] G. Downey, “Estimation of moisture in undried wheat and barley by near infrared reflectance”, *Journal of the Science of Food and Agriculture*, Vol. 36, No. 10, 1985, pp.951-958.

[5] J. T. Alander, V. Bochko, B. Martinkauppi, S. Saranwong, & T. Mantere, “A review of optical nondestructive visual and near-infrared methods for food quality and safety”, *International Journal of Spectroscopy*, 2013, pp. 1-36.

[6] D. Bertrand, P. Robert, and W. Loisel, “Identification of some wheat varieties by nearinfrared reflectance spectroscopy”, *J. Sci. Food Agric.*, 36: 1120-1124. doi:10.1002/jsfa.2740361114, 1985.

[7] Sabri, Mohd Shafiq Amirul, et al. "Analysis of Near-infrared (NIR) spectroscopy for chlorophyll prediction in oil palm leaves." *Bulletin of Electrical Engineering and Informatics*, 8.2, 2019, pp.506-513.

[8] A. Anne Frank Joe, and A. Gopal, “A Study on Various Preprocessing Algorithms Used For NIR Spectra”, *Research Journal of Pharmaceutical, Biological and Chemical Sciences*, Vol. 7, No. 4, 2016, pp. 2752- 2757.

[9] A. Anne Frank Joe, and A. Gopal, “Identification of spectral regions of the key components in the near infrared spectrum of wheat grain”, *International Conference on circuits Power and Computing Technologies [ICCPCT]*, ISBN: 9781509049684, 978-1-5090-4966-0/17/ 2017, 2017.

[10] P. C. Williams, B. N. Thompson, “Influence of whole meal granularity on analysis of HRS wheat for protein and moisture by near infrared reflectance spectroscopy (NRS)”, *Cereal Chem.* 1978, 55, 10141037.

[11] A. Anne Frank Joe, and A. Gopal, “Performance evaluation of chemometric prediction models- Key components of wheat grain”, *Journal of Scientific and Industrial Research (JSIR)*, ISSN:0975-1084 (Online); 0022-4456 (Print), Vol. 79, 2020, pp.148-152.

[12] N. J. Ibrahim, M. Y. Idna Idris, M. Y. Mohd Yusoff, N. N. Abdul Rahman, & M. I. Dien, “Robust Feature Extraction Based on Spectral and Prosodic Features for Classical Arabic Accents Recognition” *Malaysian Journal of Computer Science*, 2019, pp.46–72.

[13] J. A. Stark, “Adaptive image contrast enhancement using generalizations of histogram equalization”, *IEEE Transactions on image processing*, Vol.9, No. 5, 2000, pp. 889-896.

[14] S. Kamdi, and R. K. Krishna, “Image segmentation and region growing algorithm”, *International Journal of Computer Technology and Electronics Engineering (IJCTEE)*, Vol. 2, No. 1, 2012, pp. 103-107.

[15] M. R. Mustaffa, N. Mohd Norowi, & S. M. Yee, (2020). “Content-Based Feature Fusion Representation for Marine Invertebrates”, *Malaysian Journal of Computer Science*, Vol. 33, No. 3, 2020, pp.170–187.

[16] K. Ruoff, W. Luginbühl, S. Bogdanov, J. O. Bosset, B. Estermann, T. Ziolko, & R. Amadó, “Authentication of the botanical origin of honey by near-infrared spectroscopy”, *Journal of agricultural and food chemistry*, 54 18, 2006, 6867-72.

[17] B. Yekkehkhany, A. Safari, S. Homayouni, & M. Hasanlou, “A comparison study of different kernel functions for SVM-based classification of multi-temporal polarimetry SAR data”, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(W3), 2014, pp.281-285.

[18] A. M. Martínez, and A. C. Kak, "PCA versus LDA", *IEEE transactions on pattern analysis and machine intelligence*, Vol.23, No. 2, 2001, pp. 228-233.

[19] L. Breiman, "Random Forests, Machine learning", *Kluwer Academic Publishers*, Vol.45, No. 1, 2001, pp. 5-32.

[20] W. J. JI, X. Li, C. X. LI, Y. Zho, and Z. Shi, "Using different data mining algorithms to predict soil organic matter based on visible-near infrared spectroscopy", *Spectroscopy and spectral analysis*, Vol.32, No. 9, 2012, pp. 2393-2398.

[21] A. Liaw, and M. Wiener, "Classification and regression by Random Forest", *R news*, Vol.2, No. 3, 2002, pp. 18-22.