# 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]

[1]Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G1 1XQ, U.K.
[2]International Research Institute MICA, Hanoi University of Science and Technology, Hanoi 868 2305, Vietnam

Corresponding author: Samson Damilola Fabiyi (samson.fabiyi@strath.ac.uk)

**ABSTRACT** Inspection of rice seeds is a crucial task for plant nurseries and farmers since it ensures seed quality when growing seedlings. Conventionally, this process is performed by expert inspectors who manually screen large samples of rice seeds to identify their species and assess the cleanness of the batch. In the quest to automate the screening process through machine vision, a variety of approaches utilise appearance-based features extracted from RGB images while others utilise the spectral information acquired using Hyperspectral Imaging (HSI) systems. Most of the literature on this topic benchmarks the performance of new discrimination models using only a small number of species. Hence, it is unclear whether or not model performance variance confirms the effectiveness of proposed algorithms and features, or if it can be simply attributed to the inter-class/intra-class variations of the dataset itself. In this paper, a novel method to automatically screen and classify rice seed samples is proposed using a combination of spatial and spectral features, extracted from high resolution RGB and hyperspectral images. The proposed system is evaluated using a large dataset of 8,640 rice seeds sampled from a variety of 90 different species. The dataset is made publicly available to facilitate robust comparison and benchmarking of other existing and newly proposed techniques going forward. The proposed algorithm is evaluated on this large dataset and the experimental results show the effectiveness of the algorithm to eliminate impure species by combining spatial features extracted from high spatial resolution images and spectral features from hyperspectral data cubes.

**INDEX TERMS** Hyperspectral imaging, rice seed variety, spatio-temporal feature fusion.

## I. INTRODUCTION

Inspecting rice (Oryza sativa) seed variety is a critical procedure for quality assessment in the arable sector [1]–[3]. Ensuring that all seeds in a batch belong to one variety is a significant challenge for professional inspectors and farmers in seedling propagation stations. Varietal contamination can affect rice seed yields by introducing weeds and off-types into the crop while making it susceptible to disease. In turn, both factors contribute to the grade and price of produce, all of which has a high impact on large rice exporting nations such as Thailand, Vietnam and China. In a biotechnology context, where new species are engineered, the price and

quality of product is affected by its originality and the certification and authorisation process is the responsibility of Rice Seed Authenticating Centres. Therefore, these centres have strict and challenging requirements to identify and confirm/authorise new rice accessions while also protecting existing rice varieties with high confidence. To fulfil such strict requirements, seedling propagation stations and plant protection centres, often utilise conventional methods that rely on extracting a sample of rice seeds from a batch and using human visual inspectors to perform manual screening. In this analysis, inspectors accept or reject grains based on their appearance by analysing features such as: shape, length, width and colour. This is a tedious, laborious, time consuming task and requires trained and experienced personnel. Automation of the inspection process would not only permit increased

**TABLE 1.** A survey on current publications.

| Ref. | Number of Varieties | Extracted Features | Sensing modality | Classifier | Performance | Dataset Public | Year |
|---|---|---|---|---|---|---|---|
| Huang et al. [4] | 3 | Shape-based | RGB | Back-Propagation Neural Network (BPNN) | 95.56% | No | 2017 |
| Hong et al. [6] | 6 | Morphological, colour, texture, Gist Features, Scale-Invariant Feature Transform (SIFT) | RGB | Random Forrest (RF), Support Vector Machines (SVM) | 90.54% | No | 2015 |
| Lui et al. [7] | 6 | Colour, Morphological | RGB | Neural network | 84.33% | No | 2005 |
| Peralta et al. [8] | 754 | Shape-based | RGB | Residual error and Average Point distance | 44.8679 and 44.84528% | No | 2016 |
| Kong et al. [9] | 4 | Spectral | HSI | Partial Least Squares Discriminant Analysis (PLS-DA), K-Nearest Neighbor (K-NN), SVM, RF | 80% - 100% | No | 2013 |
| Wang et al. [3] | 3 | Spectral and Morphological | HSI | Principal Component Analysis (PCA), BPNN | 89.18% - 94.45% | No | 2014 |
| Vu et al. [2] | 6 | Spectral and Morphological | HSI | RF, SVM | 84% | No | 2016 |
| Kuo et al. [11] | 30 | Morphological, Colour, Texture | RGB | Sparse coding | 89.1% | No | 2016 |
| On Yang et al. [12] | 5 | Colour | RGB | BPNN | 93.66% | No | 2010 |
| Aznan et al. [13] | 5 | Morphological | RGB | Discriminant function analysis | 96% | No | 2016 |
| Pazoki et al. [14] | 5 | Colour, Morphological, Shape | RGB | Multi-layer perceptron, neuro-fuzzy neural networks | 98.40% - 99.73% | No | 2014 |
| Sun et al. [15] | 4 | Spectral, texture and Morphological | HSI | SVM | 91.67% | No | 2015 |
| Singh et al. [16] | 4 | Colour, Texture, Wavelet | RGB | BPNN | 96.25 - 100% | No | 2016 |

workforce productivity, by redeployment of personnel to more crucial tasks, but also has the potential to increase the consistency of inspections while allowing a larger number of batches/seeds to be screened with confidence.

In the literature, a wide range of computer vision approaches have been proposed to automate the non-destructive inspection of rice seeds [4] that commonly rely on conventional RGB cameras [5]–[8]. These approaches extract appearance-based features from the seeds such as shape descriptors, texture or colours and train models using techniques from machine learning to discriminate between species. The advent of Hyperspectral Imaging (HSI) provides an alternative sensing modality with improved discrimination performance [9] and more recent approaches combine appearance and spectral features [3], [10], [2] to increase model robustness even further.

Unlike conventional RGB imaging which measures reflectance in 3 spectral regions (Red, Green, Blue), hyperspectral imaging offers increased spectral resolution in hundreds or thousands of spectral bands. However, compared to RGB cameras, HSI hardware, typically offer reduced pixel density (∼150 DPI) mainly due to spatial binning and to improve the robustness of acquired data. The reduced resolution leads to reduce fidelity in the appearance-based features especially when discriminating small objects such as rice seeds. This observation motivates the proposed approach,

to build a system that combines both high spatial resolution (RGB) and high spectral resolution (HSI), sensing modalities combined with intelligent data processing techniques to perform rice seed screening automatically.

Table 1 lists a number of existing techniques that have been proposed in the literature in recent years. The table allows a quick comparison of existing techniques based on factors including: the number of species considered in the studies, the features extracted, the sensing modalities utilised, the algorithmic approaches taken to process the data, and, their reported performance. It can be noted that the majority of the techniques presented are applied to RGB images and almost all methods are evaluated using only asmall number of different rice seed species with varying degrees of accuracy. From the review in Table 1, it is not clear whether the differences in performance between existing techniques is caused by superior algorithms and the effectiveness of feature descriptors used, or, if this is simply due to differences in the inter-class or intra-class variation of species used in each study. That said, Kue *et al.* [11] and Peralta *et al.* [8] used relatively large number of species in their study having evaluated their method using 30 and 754 species and reported accuracy of up to 89.1% and 44.87% respectively. Their reported performance is not as strong as some of the other methods listed in the table, and this supports the hypothesis that other techniques do not necessarily utilise better

algorithms or feature descriptors. Instead, it is possible that these techniques are evaluated in less challenging datasets with a smaller number of species which exhibit favourable intra-class and inter-class variation. However, this hypothesis is difficult to confirm or reject as the datasets used in each paper are not publicly available.

The contributions of this paper are as follows:

1. A novel rice seed inspection system that combines a conventional RGB and hyper-spectral imaging system is proposed.

2. An innovative framework to fuse spatial and spectral data is developed and it is shown that the combined features improve discrimination ability and classification performance.

3. The performance of the proposed algorithm and system is evaluated in a large, diverse dataset of 90 rice seed varieties with 96 seeds per variety. Experiments show that varying the number of rice seeds species in the datasets can impact the classification performance and recommends that the similarity of rice seeds varieties be assessed.

4. The large dataset evaluated in this paper is made publicly available[1] to the community to assist in the benchmarking of proposed algorithms and features.

The remainder of this paper is organised as follows: Section II describes in detail, the related work. A description of the: system setup; datasets; data processing; feature extraction; classification; and species discriminant analysis technique adopted are presented in Section III. Section IV presents the experimental results and evaluates the performance of the proposed techniques on the collected dataset. Finally, Section V concludes the work and suggests future research directions.

## II. RELATED WORK

Machine vision systems have been proposed for a range of food quality assessment tasks [10], [17]–[19]. Research has focused on combining image analysis and machine learning techniques to create new methods to perform automatic inspection and qualification. Relevant to the study presented here are rice seeds (polished) quality control or cultivar classification tasks which are specifically explored in [1]. In [1], Y. Ogawa comprehensively surveys computer vision techniques, physical property measurements, compound content and distributions of rice grains for seed quality control. Related work which specifically address non-destructive variety classification are listed in Table 1.

Rice seed classification using an automatic machine-vision system usually consists of several key stages. The most important of which include image data collection, feature extraction (such as shape, size, colour, and orientation etc.) and feature representations via models using pattern recognition algorithms or multivariate analysis techniques.

The appearance-based approaches often utilise morphological, colour, and textural traits, or a combination of them. As early as 1986, Lai et al. [20] proposed the use of image analysis to determine the physical dimensions of and manually classify cereal grains. Similarly, Sakai *et al.* [21] demonstrate the use of two-dimensional image analysis to extract dimensions and shape factors of 4 varieties of polished rice grains and manually classify them. More recent work [6], [7] and [12] focuses on rice seed variety classification tasks. Commonly, shape descriptors of the seed samples are extracted and classifiers such as Random Forests [6], Neural Networks [7] or Cubic B-Splines shape model [8] are trained.

Huang *et al.* [4] proposed a detailed analysis of shape descriptors that goes beyond features that are more commonly used in the literature, such as chaff tip (width, height) and depth of concavities of rice kernels. Their work shows promising results in separating visually similar species but their evaluation is limited to only 3 varieties. Kuo *et al.* [11] utilise multi-focus image fusion to study 30 varieties of rice seed using sparse representation classification and obtain accuracy of 89.1% with a standard derivation of 7.0%. Although, the authors briefly recognise that the majority of the literature uses a limited number of species, they do not illustrate the effect that this may have in discrimination ability. Their approach focuses on detailed Region-Of-Interest (e.g., sterile lemmas) on the grains.

Recently, HSI techniques have been used in food and agriculture engineering. Wang *et al.* [3] used VIS/NIR (400-1000 nm) spectral information to discriminate 3 rice varieties. The authors used a combination of: degree of chalkiness; shape features; and spectral features which are all extracted from the acquired HSI images. The dimensionality of the spectral features was reduced using Principal Component Analysis (PCA) and the resulting principal components were used to train an Artificial Neural Network, with classification accuracy of 94.45% achieved.

In [22], the authors discovered that a combination of the Least squares support vector machine (LS-SVM) regression method and VIS/NIR spectroscopy at range 325-1075 nm provides a realisable technique to monitor the nitrogen status in rice. More recently, an HSI system has been used in [9] for identifying four rice seed cultivars. By utilizing the full spectral range of their system 1,039-1,612 nm, the authors, [9], achieved results of up to 100% accuracy with a Random Forest (RF) classifier. However, four cultivars in [9] were hybridized from other species, therefore, it is unclear how the inter/intra class varies among them. Recently, work in [2] and [15] explored different feature combination schemes: spectral and texture features; morphological, texture and spectral features; and morphological and texture features while seeking the optimal feature combination. The highest accuracy in [15] was obtained when using combined spectral, morphological and texture features (91.67%, with four polished rice species). A classification accuracy of 84% was achieved in [2] utilising a combination of spectral and spatial features on a dataset of six rice seed species.

---

When comparing the existing techniques in the literature, one can see some shortcomings. Firstly, the performance of related techniques as listed in Table 1 varies significantly and there are no benchmark evaluations or common datasets to compare these works together. Secondly, although the results of these studies have been promising, most researchers have evaluated their methods on a relatively limited number of different rice seed species. In practical applications, an automated system faces challenges of various rice seeds. For instance, the authors in [8] also noted a very large number of rice accessions (120,000+ accessions) which are available in the TT Chang Genetic Resources Centre at the International Rice Research Institute (IRRI), Philippines. They argue that characterising shapes of the cultivated rice accessions greatly helps in authenticating new rice seeds. Kuo *et al.* [11] also note that rice grains of hundreds of varieties are cultivated and hence the demand for robust non-destructive authentication approaches of large number of grain seed varieties must be addressed to make solution feasible for the rice seed industry. In this study, the dataset consists of ninety rice seed varieties and is made publicly available. Both the high spatial resolution and full-band wavelengths of spectral information are fundamental resources to develop and deploy robust classifiers for performing variety classification of different rice seed species and other relevant tasks in rice seed quality inspections.
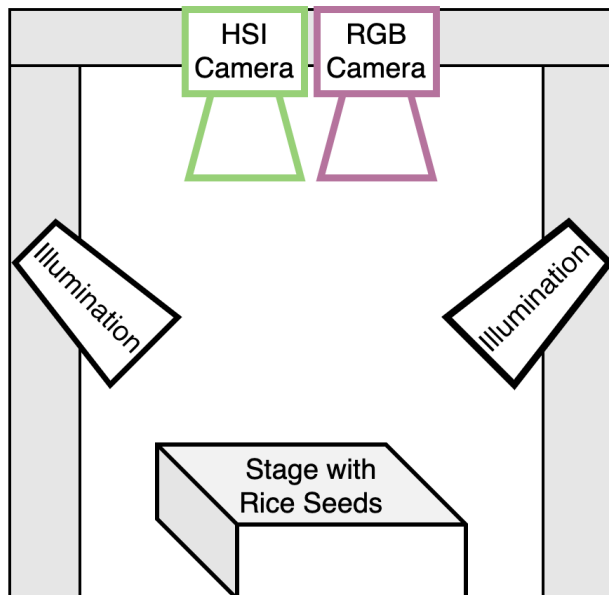


**FIGURE 1.** A schematic diagram of the HSI and CCD cameras setup for data acquisition.

## III. METHODS

### A. SYSTEM SETUP

A schematic diagram of the system used for acquiring images of rice seed samples is shown in Fig. 1. It is comprised of a high resolution RGB camera and a hyperspectral imaging system. The digital camera used was a Fujifilm

X-M1 with a 35mm/F2.0 lens to collect RGB images at $4,896 \times 3,264$ pixels. For gathering HSI data, a Visible - Near Infrared (VIS/NIR) range HSI system was used which consisted of a Specim V10E Imaging Spectrograph and Hamamatsu ORCA-05G CCD camera. The HSI system was configured to capture hyperspectral image data cubes in which each pixel contains a spectrum of reflected light at 256 discrete wavelengths ranging from $\sim (385 - 1000)$ nm. Two halogen bulbs were used for illumination and these were accurately positioned to provide balanced lighting across the scene. To ensure stability, the halogen bulbs were switched on and allowed to reach constant operating temperature before the data were acquired in a dark room to minimise any other sources of illumination variance.

The Fujifilm RGB digital camera was set to operate in manual mode with an ISO of 400 and a shutter speed of 16 ms. The acquired images were saved in JPEG format and no automatic adjustment (e.g., white balance) was applied.

The HSI system is a push-broom system and hence captures HSI data in a line scan fashion. For this reason, a motorised translational stage was positioned directly beneath the imager to allow scanning. To accurately and consistently collect the data in a repeatable way, three key parameters of the HSI system needed to be adjusted and were selected as follows:

- The exposure time of the camera (e.g. 500 ms) versus the speed of movement of the translational stage (5 mm/s), which was calibrated in order to avoid spatial distortions;
- A trade-off between the exposure time and the aperture of camera (f=18) to ensure a suitable light intensity;
- The height between the lens and the stage which was set so that camera field of view captures the entire area containing all seeds in each data cube.

Prior to imaging rice seeds, images of a flat checkerboard patterns were collected for calibrating planar and lens distortion effects for the RGB and HSI sensors using the traditional camera calibration approach described in [23]. This procedure was performed once and the camera calibration parameters; i.e. rotation and transformation parameters, were stored in a XML file for aligning and registering the images acquired from both systems.

### B. DATASET DESCRIPTION AND PROCESSING

Ninety known rice seed varieties were provided for this study by the National Center of Protection of New Varieties and Goods of Plants (NCPNVGGP) in Vietnam. These rice seed varieties were chosen since they are frequently planted in Vietnam to cultivate rice for consumption and exportation. The selected samples were manually screened in the traditional way by experienced technical staff at NCPNVGGP to ensure that each sample population contained only seeds which belong to the 90 species to be analysed. A single kernel from each of the 90 varieties is shown in Fig. 2.

For each of the 90 species considered in this work, 96 individual rice seeds were provided. The kernels from each of the 90 species were divided to 2 batches each containing
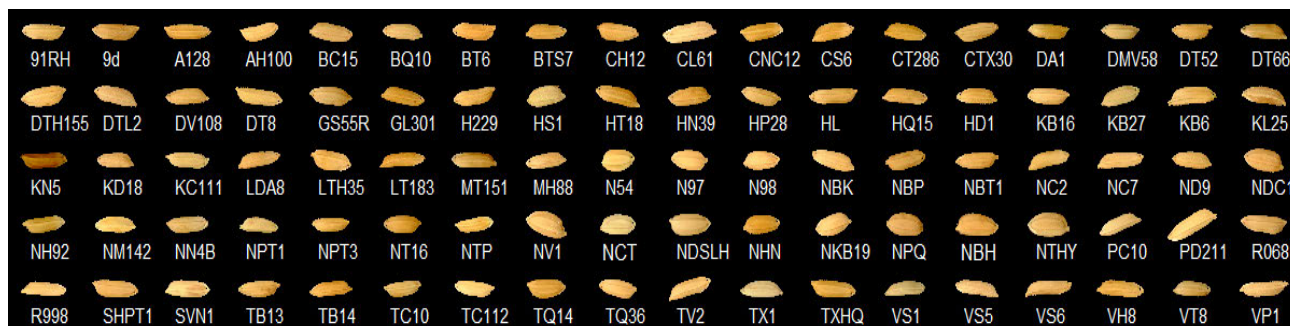
**FIGURE 2.** Photos of rice seed samples of 90 species. The short name of each species is given beneath each kernel.

48 individual rice seed samples. Each set of 48 seeds was then positioned on a white sheet of paper in a (6 × 8) matrix structure and imaged using the HSI system and the digital RGB camera. Thus, each of the 90 species of rice seeds were captured in 2 hyperspectral data cubes and 2 high resolution RGB images each containing 48 different seeds. In total, the dataset consists of fully registered RGB and HSI images ~(385 – 1000) nm of 8640 seeds (90 varieties × 96 seeds). In this study the seeds were manually positioned on the white sheet to avoid overlaps or touching boundaries between them. In practice, one can conceive a conveyor belt arrangement where seeds are mechanically spread individually under appropriate and consistent illumination for imaging.
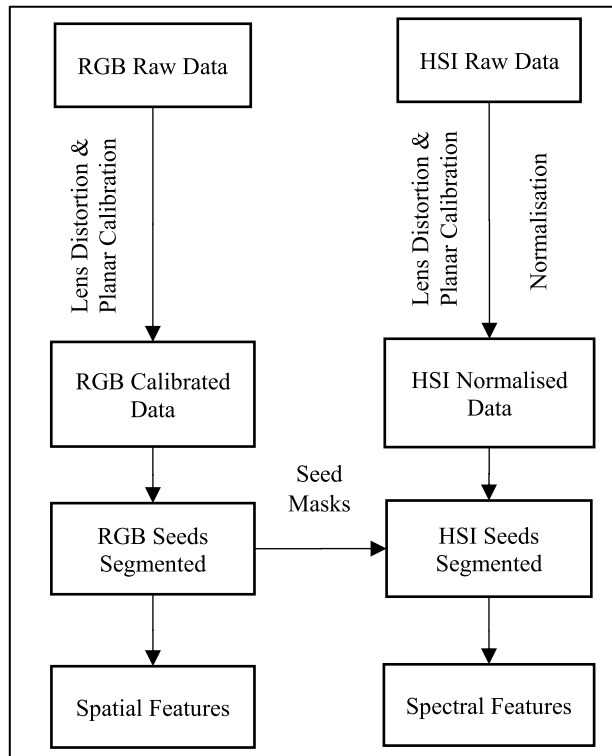


**FIGURE 3.** A block diagram showing the stages of the data processing task.

The data processing steps to extract spatial and spectral features is illustrated in Fig. 3. Initially, both image modalities are calibrated for lens and planar effects using the rotation and transformation matrix obtained from the checkerboard pattern as described in Section III-A. Then, the HSI data cube is normalised as described in Section III-D. After this, the processing paths diverge slightly. The RGB data are segmented using the process described in Section III-C and binary masks for each rice seed in the RGB image are used to extract the high resolution spatial features as described in Section III-D.

The masks obtained from the RGB segmentation are then transformed to the HSI space (using the calibration matrix obtained from imaging the checkboard pattern) to segment the rice seeds in the HSI data. The segmented seeds are subsequently used to extract the spectral features as described in Section III-E.

### C. CALIBRATION PROCEDURES AND RICE SEED SEGMENTATION

The rice seed segmentation is performed on the high spatial resolution RGB images to ensure that the complete kernel is captured. The proposed procedure consists of the following steps. First, the R-channel of the RGB image is extracted; the R-Channel is chosen because it offers the highest contrast to the background. A top-hat transform [24] of the R-channel is then computed and thresholded to obtain the binary images.

### D. SPATIAL FEATURE EXTRACTION

Trained personnel screen rice seeds by manually analysing their spatial features. In this work, spatial features of the separated seeds are extracted from RGB image masks to encapsulate the expert knowledge. The features extracted are selected due to their effectiveness for discriminating among species, as shown in recent work, [6] and [7]. The morphological feature vector $f$ for a single kernel contains the following six attributes:

$f_1$: Area; the number of pixels inside a seed kernel

$f_2, f_3$: Major Axis Length and Minor Axis Length; specify the length (in pixels) of the major and minor axis of the ellipse that has the same normalised second central moment at the region of the sample seed

$f_4$: Aspect Ratio; the ratio of Minor Axis Length over the Major Axis length ($\frac{f_3}{f_2}$)

$f_5$: Perimeter over Area Ratio; $\frac{Perimter}{Area}$, where Perimeter is number of pixels along the seed boundary; and Area is the $f_1$ feature.

$f_6$: Eccentricity; FociDistance/MajorAxisLength (i.e. the eccentricity of the ellipse with the same second moments as the seed kernel region) is the distance between two foci of the ellipse, and the major axis length of the ellipse (i.e. feature).
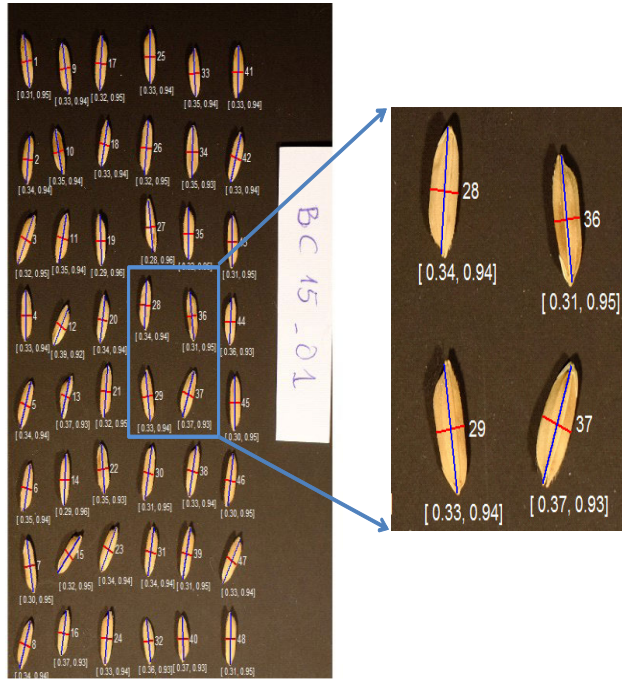


**FIGURE 4. Results of the spatial feature extraction. The ID number of each seed is shown to its right. For illustration purposes, only features f4 and f6 are shown below each kernel. The MajorAxisLength f2 and MinorAxisLength f3 are marked with a red and a blue line, respectively.**

Fig. 4 illustrates output of the spatial feature extraction for a sample image that contains 48 seeds. It is noted that because the spatial features are extracted using the high spatial resolution images, they are expected to be more accurate than those reported in [2], [15] where the spatial features are extracted from the HSI system which has lower spatial resolution.

### E. SPECTRAL FEATURE EXTRACTION

Extracting the spectral information consists of two main phases: data correction, and feature extraction. These are illustrated in Fig. 3. Firstly, to reduce the variation in the acquired reflectance values among measurements, the collected data are corrected following the approach laid out in [25]. Let $y$ denote a data cube consisting of reflectance values as a two-parameter set:

$$y_\lambda(x), \quad x \in X, \ \lambda \in \Lambda \tag{1}$$

where $\lambda$ represents a wavelength belonging to $\Lambda$, that is a set of the wavelengths at VIS/NIR range $\sim (385-1000)$ nm and $x$ represents a pixel in $X$ where $X$ is 2-D coordinate by row $m$ and column $n$. For each specific wavelength, the array of

reflectance values can be regarded as an image where spatial relationships between the pixel reflectance values have meanings. At each $x$, the raw reflectance value could vary due to different lighting conditions. To reduce the variation in the acquired reflectance values due to illumination variations or unique pixel response, the data were scaled relative to known max reflectance value; the raw data are normalised using:

$$y\lambda(x) := \frac{y_{raw}, \lambda(x) - b(n, \lambda)}{w(n, \lambda) - b(n, \lambda)}, \quad \lambda \in \Lambda \tag{2}$$

where $b(n, \lambda)$ and $w(n, \lambda)$ are the reflectance values of reference dark and white images. The dark reference is acquired by covering the lens-cap and the white reference is a 100% reflective spectralon tile which is a highly reflective Lambertian scatter commonly used to calibrate HSI systems. For each $\lambda$, $b(n)$ and $w(n)$ are averaged on reflectance values at column $n$ along the white tile's height dimension. After normalizing the spectral data, the rice seed masks obtained from the above description of the rice seeds segmentation procedure are transformed to the HSI plane using a transformation matrix including rotation and translation operators to segment the seed samples on the spectral images, as shown in Fig. 3. Fig. 5 clearly illustrates the advantages of the proposed segmentation method compared to the conventional work (e.g., in [2], [15]) where the seed segments are extracted directly from the low spatial resolution images collected by HSI system. When the seeds are segmented using only the HSI data, in many cases, the segments include some pixels at boundary regions of the shadow rather than the pure spectra of the seeds themselves. Subsequently, any measurements of the morphological features made using only spectral image segmentation could become inaccurate and spectra of non-rice-seed pixels will be included in the analysis. Hence, in this work the seeds are segmented from the high resolution RGB image to ensure that the spatial and spectral features are correctly included in the analysis.

Based on the segmented seed samples on a hyperspectral data cube, spectral information from every pixel of the seed regions is extracted. The mean spectrum of all pixels in each seed is then computed and used to determine the spectral features for that seed. As denoted in (2), a raw spectral feature vector of a rice seed sample is a set of $y_\lambda$ in which $\lambda$ is one of the 256 bands belonging to a range $\Lambda \sim (385-1000)$ nm for our VIS/NIR system.

### F. DIMENSIONALITY REDUCTION

Originally collected spectra are high dimensional feature sets. LDA is a statistical tool that can be used to reduce the dimensionality of data [26]. For this application, LDA is used to select features that account for the highest degree of variation among the 90 species in the dataset and select variables for classification or discriminant analysis of the rice seed species. In LDA, features are ranked based on the amount of variation they account for in the data. LDA is a supervised algorithm and so uses the labels provided in the dataset. In this work,
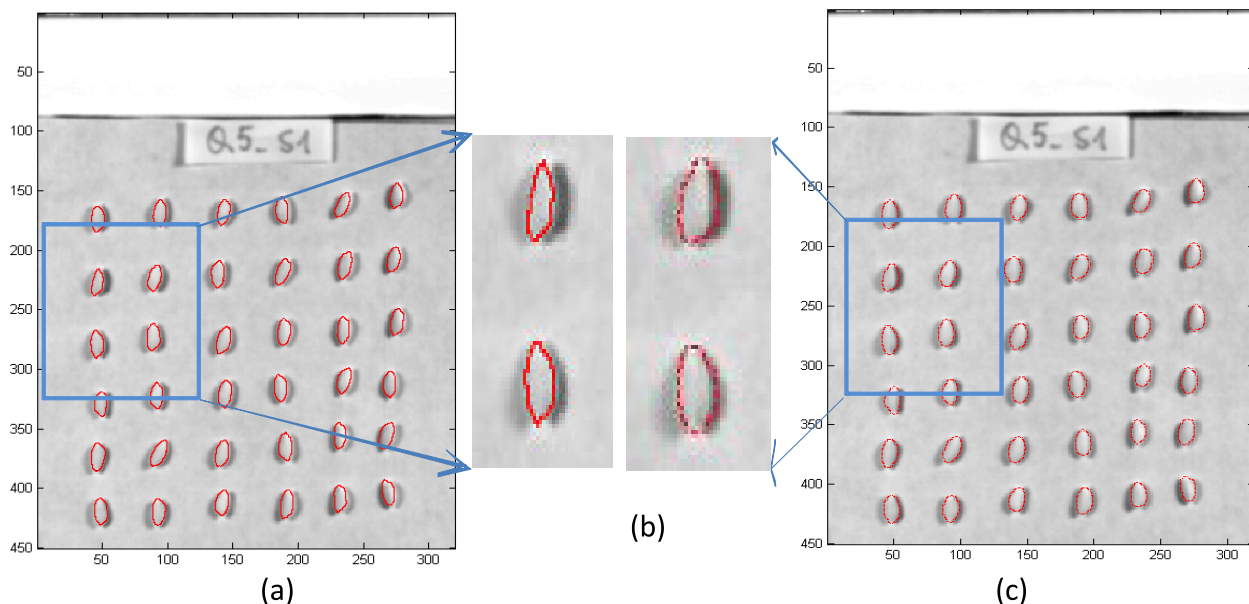
**FIGURE 5.** Comparing segmentation results of rice seeds on HSI image. (a) Using RGB image for reference (b) A Zoom-in version of two segmentation results (Red and white points are the results from (a) and (c) respectively) (c) Without using RGB image.

LDA was adopted because it maximizes the separability of the different rice seed species. LDA can be used to reduce the dimensionality of the spectral features (originally 256) in the dataset by maximizing the separation among the 90 species of rice seeds. The transform achieves this by maximizing the distance between the means of 90 species (interclass variance) and minimizing the variation within each category (intra-class variance) itself [27]. In section IV, the spatial features are combined with varying number of LDA features to determine the best feature combination scheme for performing rice seed classification using our dataset.

### G. RICE SEED VARIETY CLASSIFIER

The aim of variety classification is to detect seeds within a batch which are not of the species that the batch is claimed to represent. Models which are trained to perform this task could directly utilize the full-band wavelengths or only selected components (for spectral features) from the output of feature extraction or dimensionality reduction techniques adopted, and/or spatial context. Therefore, the features extracted are utilized in four different schemes: (1) spatial features only, (2) spectral features only, (3) a combination of all 256 spectral features and extracted spatial features; (4) a combination of LDA components extracted from the spectral data fused with the spatial features. For the task of rice seed variety classification, a Random Forest (RF), which has produced better classification results than many other classifiers including support vector machines and K-Nearest Neighbor in many related work [6], [9], [2], is trained for rice seed classification.

Random forests contain many decision trees which are grown from a bootstrap sample of the response variable. Trees are allowed to grow to a maximum extent without pruning

after selection of the best split from a random subset of features at each of the decision trees' nodes. Accumulating the results of each decision tree forms the basis on which random forests make predictions on new data. Random forests are able to process large datasets in a fast, effective way while achieving high classification accuracy. The number of decision trees and the ratio of the training to test samples that gave the best classification results during preliminary analysis are 500 and 4:1 respectively, and so are adopted for this work.

### IV. EXPERIMENTAL RESULTS AND ANALYSIS

In order to evaluate the performance of our RF classifier and the effectiveness of the approaches proposed in this work, three performance metrics are used. They are Precision, $P$, Recall, $R$, and $F_1$ Score.

$$P = \frac{t_p}{t_p + f_p} \tag{3}$$

$$R = \frac{t_p}{t_p + f_n} \tag{4}$$

where $t_p$ is the number of true positives, $f_p$ is the number of false positives, $t_n$ is the number of true negatives and $f_n$ is the number of false negatives. The $F_1$ score is the harmonic mean of the precision and recall,

$$F_1 \text{score} = 2 * \frac{P * R}{P + R} \tag{5}$$

Based on the collected datasets of ninety species, and in line with the objectives of this work, results and analysis are presented in this section under three different scenarios. The first and second scenarios aim to show the effectiveness of using spatial and spectral features extracted from a high

resolution RGB image fused with HSI data for rice seed classification. In the first scenario, the RF classifier was trained using all 90 species. Since we are also interested in exploring how the number of species in the dataset affects classifier performance, we considered a second scenario where 6 different subsets of the datasets were used to train the classifier separately and evaluate performance. Each of the 6 subsets of the data consists of 6 species (more than or similar to the typical number used in related work – see Section I, Table 1) drawn randomly from the 90 available. In the third scenario, we select a different 6 subsets of the data and vary the number of species in each subset from 6 – 90 to explore how increasing the number of species in the dataset affects classifier performance. For each sub-dataset, the performance of the random forest is reported and analysed.

**TABLE 2.** Classification results with and without dimensionality reduction.

| Feature Schemes | Average precision (%) | Average Recall (%) | Average F1 Score (%) |
|---|---|---|---|
| spatial | 16.33 | 16.57 | 15.96 |
| spectral | 34.93 | 35.86 | 34.46 |
| Spatial + spectral on full bands | 51.66 | 51.49 | 50.51 |
| Spatial + 85 LDA Components from Spectral | 79.64 | 78.80 | 78.27 |

### A. ANALYSING PERFORMANCE ON ALL 90 SPECIES

In this experiment, we trained 4 different RF models using different combinations of the data we gathered for training and testing. The classification results presented in Table 2, illustrate that the average precision, average recall and average $F_1$ score were lowest when using only spatial features to train our RF. The average precision, average recall and average $F_1$ score improved when only spectral features were used. As expected, higher average precision, average recall and average $F_1$ scores were achieved when the spatial and spectral features were combined and analysed together. Next, we used LDA to reduce the dimensionality of the spectral features (originally 256) in the dataset. The outputs of the LDA were separately combined with the spatial features extracted from our dataset of 90 species. These data were used to train the RF classifier. The classification results obtained when using this approach to differentiate between all 90 species. While we have discussed the average precision, recall and $F_1$ score here, we also show the individual scores for each of the 90 species in Table 3. Interestingly, the $F_1$ scores vary between 80-100%, and 90-100% for 45 and 22 varieties respectively. These results showed very high $F_1$ scores and are comparable with those reported in Table 1. These results are pointers to the potential of our new imaging modality (combining spatial features from RGB images and spectral features from hyperspectral images) for

rice seed discrimination even with a fairly large number of species. However, we also note that this technique, when used on the entire dataset of 90 species, performs better for some species than others as illustrated in Table 3. Further analysis carried out to find out what could be responsible for this is presented in section IV-C.

### B. ANALYSING SELECTED SUBSETS OF 6 SPECIES

In this scenario, we selected 6 subsets from our dataset in order to compare our algorithm with state-of-the-art techniques which (as shown in Table 1) tend to be evaluated on a small variety of species; typically, 5-6 with the exception of [11] (30 species) and [8] (754 species). 5 of the 6 subsets selected, each consisted of 6 randomly selected species while the 6th subset consists of species that exhibited the worst discrimination performance in Table 3. All the subsets are summarized in Table 4. We used LDA to extract features from the spectral data of each subset and combined the outputs with corresponding spatial features using the best feature combination scheme selected. The 6 subsets were used separately to train the RF classifier and the classification results are presented in Table 5. We also note that the computation time (segmentation and classification) is 0.53 s on a commodity hardware; Intel Core i7. The computation time is insignificant for practical application of screening rice seeds compared to a manual human screening which takes minutes. For the first 5 subsets studied, we observed that the average precision and average recall were significantly improved when compared to the performance of the 90 species classification results. Our novel approach produced very high average $F_1$ scores and outperformed equivalent scores reported for the state-of-the-art techniques listed in Table 1 except for those in [9], [14] and [16]. This experiment shows that, in line with state-of-the-art techniques for rice seed classification, very good results and elimination of impure species from rice seed samples can be achieved by taking advantage of spatial features from high spatial resolution images and combining them with spectral features from hyperspectral data cube. For the last subset of species, the average F1 score when employing the first classifier trained on 90 varieties is 45.54%. In contrast, when the RF classifier is trained specifically on the 6 species selected, the F1 score rises to 61.29%. This improvement comes as a result of the smaller dataset and suggests that when the species are known a-priori, a targeted classifier would be more appropriate.

### C. ANALYSING THE SUBSETS WITH VARYING SPECIES SIZES

From the results presented in Sections B and C, it is clear that the size of the dataset and variety of species considered in each experiment directly affects our RF classifier's performance. To explore this further, we selected a further 6 subsets of data from our set of 90 species. This time, we varied the number of species in each subset of data to include 6, 20, 40, 60, 80 and 90 different species to allow us explore exactly how varying number of species in a study

**TABLE 3.** Classification results with the best feature scheme, spatial + 85 LDA features, and 90 species.

| Labels | Precision | Recall | F1 score | Labels | Precision | Recall | F1 score | Labels | Precision | Recall | F1 score |
|---|---|---|---|---|---|---|---|---|---|---|---|
| GS55R | 100.00 | 100.00 | 100.00 | LTH35 | 89.47 | 80.95 | 85.00 | KN5 | 62.50 | 90.91 | 74.07 |
| TC10 | 100.00 | 100.00 | 100.00 | PD211 | 82.61 | 86.36 | 84.44 | TV2 | 78.57 | 68.75 | 73.33 |
| NV1 | 100.00 | 100.00 | 100.00 | LT183 | 75.00 | 95.45 | 84.00 | DTH155 | 81.25 | 65.00 | 72.22 |
| KL25 | 100.00 | 100.00 | 100.00 | PC10 | 81.82 | 85.71 | 83.72 | ND9 | 68.18 | 75.00 | 71.43 |
| N97 | 100.00 | 100.00 | 100.00 | HQ15 | 73.91 | 94.44 | 82.93 | KB6 | 73.33 | 68.75 | 70.97 |
| NPQ | 96.00 | 100.00 | 97.96 | KD18 | 87.50 | 77.78 | 82.35 | NTBH | 77.27 | 65.38 | 70.83 |
| NT16 | 95.24 | 100.00 | 97.56 | TB14 | 94.12 | 72.73 | 82.05 | GL301 | 60.87 | 82.35 | 70.00 |
| BT6 | 100.00 | 94.74 | 97.30 | NTP | 72.73 | 94.12 | 82.05 | NH92 | 66.67 | 71.43 | 68.97 |
| HS1 | 100.00 | 94.12 | 96.97 | NTHY | 76.47 | 86.67 | 81.25 | DMV58 | 66.67 | 71.43 | 68.97 |
| TXHQ | 93.75 | 100.00 | 96.77 | NDSLH | 76.47 | 86.67 | 81.25 | NBP | 75.00 | 62.50 | 68.18 |
| NCT | 100.00 | 92.86 | 96.30 | VS6 | 78.95 | 83.33 | 81.08 | NM142 | 66.67 | 66.67 | 66.67 |
| N54 | 92.31 | 100.00 | 96.00 | DA1 | 80.00 | 80.00 | 80.00 | LDA8 | 55.17 | 84.21 | 66.67 |
| CS6 | 93.33 | 93.33 | 93.33 | NC7 | 87.50 | 73.68 | 80.00 | NC2 | 54.17 | 81.25 | 65.00 |
| NPT3 | 93.33 | 93.33 | 93.33 | MH88 | 73.68 | 87.50 | 80.00 | HD1 | 60.00 | 70.59 | 64.86 |
| CL61 | 89.47 | 94.44 | 91.89 | NN4B | 79.17 | 79.17 | 79.17 | VP1 | 66.67 | 62.50 | 64.52 |
| DT66 | 85.00 | 100.00 | 91.89 | AH1000 | 70.83 | 89.47 | 79.07 | VS1 | 51.61 | 84.21 | 64.00 |
| VT8 | 100.00 | 84.21 | 91.43 | SVN1 | 81.82 | 75.00 | 78.26 | DT8 | 57.89 | 68.75 | 62.86 |
| SHPT1 | 95.45 | 87.50 | 91.30 | A128 | 84.21 | 72.73 | 78.05 | H229 | 69.23 | 56.25 | 62.07 |
| VH8 | 95.24 | 86.96 | 90.91 | CT286 | 80.00 | 76.19 | 78.05 | TX1 | 100.00 | 44.44 | 61.54 |
| N98 | 95.00 | 86.36 | 90.48 | CH12 | 76.19 | 80.00 | 78.05 | HP28 | 71.43 | 52.63 | 60.61 |
| NHN | 95.00 | 86.36 | 90.48 | TQ14 | 72.73 | 84.21 | 78.05 | MT151 | 100.00 | 42.86 | 60.00 |
| R068 | 91.67 | 88.00 | 89.80 | KB27 | 73.68 | 82.35 | 77.78 | BC15 | 56.25 | 64.29 | 60.00 |
| NDC1 | 95.45 | 84.00 | 89.36 | 91RH | 100.00 | 62.50 | 76.92 | CNC12 | 100.00 | 42.11 | 59.26 |
| BQ10 | 95.45 | 84.00 | 89.36 | CTX30 | 83.33 | 71.43 | 76.92 | KC111 | 72.73 | 50.00 | 59.26 |
| DTL2 | 87.50 | 91.30 | 89.36 | R998KBL | 66.67 | 90.00 | 76.60 | NBK | 69.23 | 47.37 | 56.25 |
| HN39 | 84.21 | 94.12 | 88.89 | HT18 | 81.25 | 72.22 | 76.47 | DT52 | 64.29 | 50.00 | 56.25 |
| HL | 85.71 | 90.00 | 87.80 | TC112 | 81.25 | 72.22 | 76.47 | NBT1 | 50.00 | 50.00 | 50.00 |
| DV108 | 80.00 | 94.12 | 86.49 | BTS7 | 66.67 | 88.89 | 76.19 | NPT1 | 40.00 | 40.00 | 40.00 |
| 9d | 86.36 | 86.36 | 86.36 | TQ36 | 64.00 | 94.12 | 76.19 | TB13 | 42.86 | 35.29 | 38.71 |
| NKB19 | 82.14 | 88.46 | 85.19 | VS5 | 76.19 | 76.19 | 76.19 | KB16 | 33.33 | 30.77 | 32.00 |

**TABLE 4.** Species contained in each sub-dataset.

| Subset | Species |
|---|---|
| 1 | HS1, CH12, AH1000, SVN1, 91RH, DT8 |
| 2 | TB14, N54, NKB19, HQ15, BT6, NC7 |
| 3 | KB6, AH1000, HQ15, TQ14, KL25, NHN |
| 4 | TC10, DTL2, KB16, BT6, KB27, CNC12 |
| 5 | CL61, NKB19, VH8, TX1, MT15, HL |
| 6 | NBK, DT52, NBT1, NPT1, TB13, KB16 |

**TABLE 5.** Classification results using the output of LDA and randomly drawn 6 species.

| Subsets | Average precision (%) | Average Recall (%) | Average F1 Score (%) | Min–Max F1 score (%) |
|---|---|---|---|---|
| 1 | 96.03 | 96.46 | 96.18 | 89.66-100 |
| 2 | 96.23 | 96.31 | 96.21 | 89.47-100 |
| 3 | 98.59 | 98.33 | 98.42 | 97.30-100 |
| 4 | 98.55 | 97.93 | 98.17 | 95.45-100 |
| 5 | 96.39 | 96.73 | 96.52 | 92.86-100 |
| 6 | 61.99 | 61.12 | 61.29 | 35.71 - 85.71 |

can influence our classifier's performance. Having prepared the data, we applied LDA to extract features from the spectral data of each of our data subsets of increasing size and we combined the outputs with corresponding spatial features. We then trained the RF classifier using the output of LDA (starting from 1 LDA component up to S-1, where S is the number of species in the dataset) combined with spatial features in order to determine the number of features that gives

the best performance. For each subset of the data considered, we obtained a plot of average F1 score against the number of LDA features used in the classification. The results of this are shown in Fig. 7. We also obtained the maximum
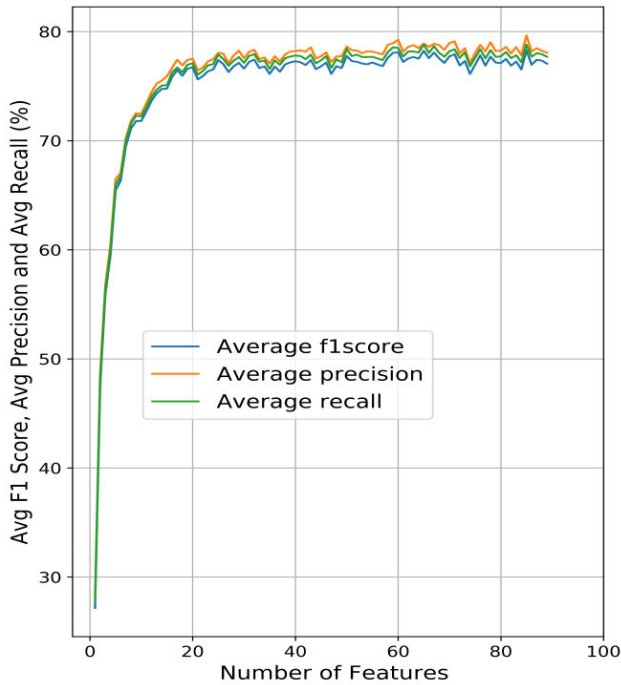
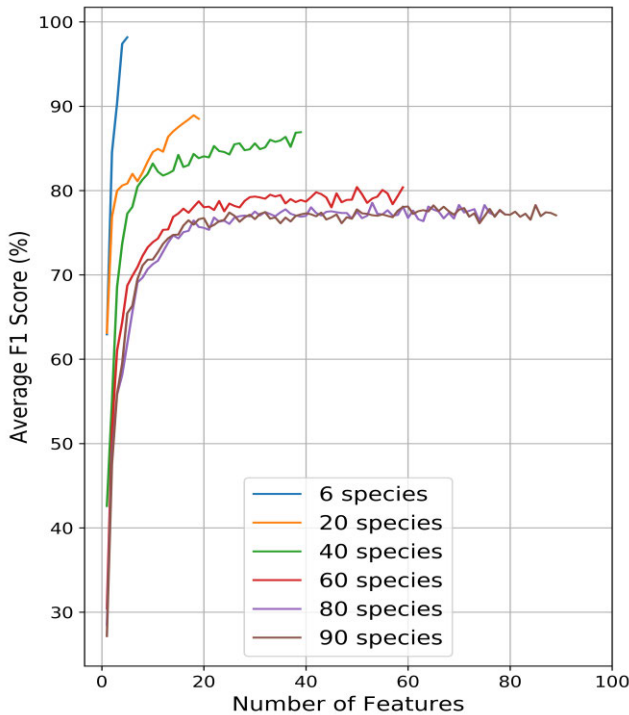**FIGURE 6.** Classification result using the outputs of LDA and 90 species.



**FIGURE 7.** Plots of average F1 scores against number of features for sub-datasets with species sizes of 6, 20, 40, 80, and 90.
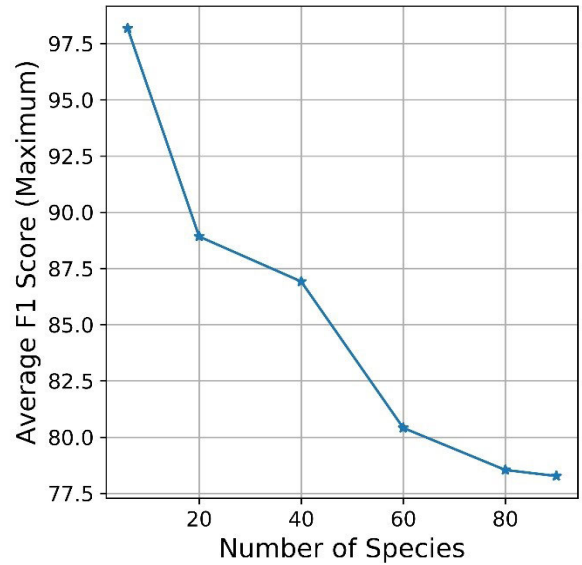


**FIGURE 8.** A plot of maximum average F1 score against number of species.
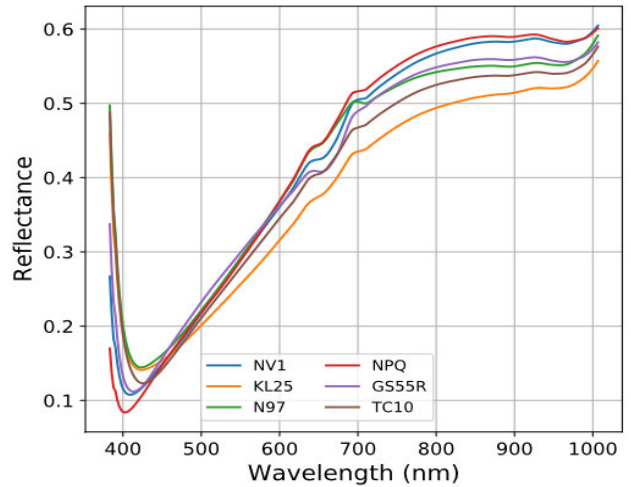


**FIGURE 9.** Average spectral profiles of some species with good classification results in Table 3.

average F1 score from each of the plots in Fig. 7 and used this information to plot the maximum average F1 score as a function of the number of species considered as shown in Fig. 8. From Fig. 8, we observed that the performance of

the RF classifier appears to be impacted by the number of species used in the study. The average F1 scores significantly dropped from 98.17% for 6 species to 78.27% for 90 species. One reason for this drop in classifier performance could be due to an increase in the level of similarity among the rice seeds species as the number of varieties also increases.

As a result, the need to investigate the influence of similarities among species of rice seeds on the classifiers' performance becomes imperative. Furthermore, we reported in the previous subsection that the approaches utilized in [9], [14], and [16] attained better results than ours. Similarity assessment of rice seeds species will also help to clarify whether the higher performances reported in these papers is due to the use of better feature combination schemes, better algorithms or inter/intra class variation
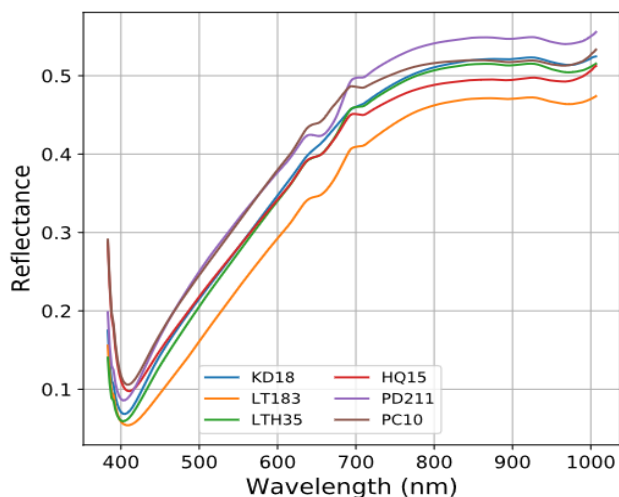
**FIGURE 10.** Average spectral profiles of some species with good classification results in Table 3.
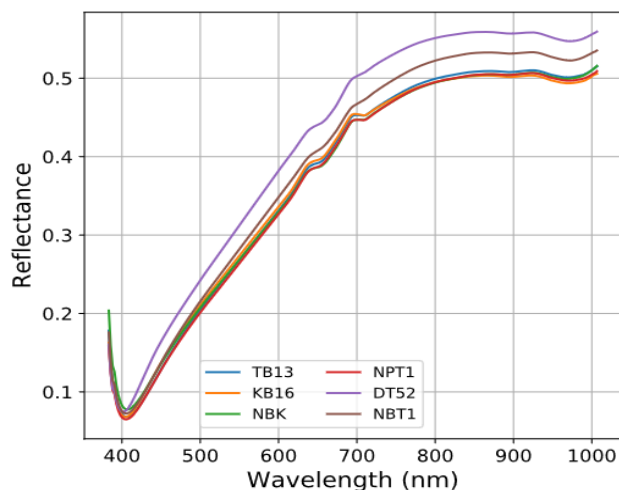


**FIGURE 12.** Average spectral profiles of some species with poor classification results in Table 3.
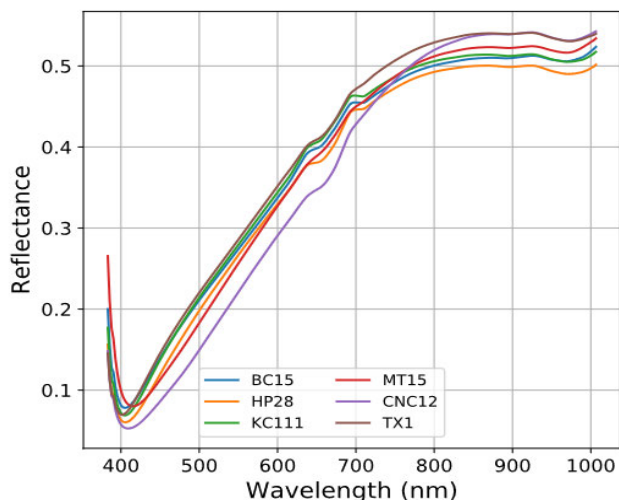


**FIGURE 11.** Average spectral profiles of some species with poor classification results in Table 3.

among species themselves. While in general the algorithm presented in this papers performs very well, our technique does have some limitations. This is clear by analysing situations where the method does not perform well. For example, some of the species which have poor classification such as species TB13, KB16, NBK and NPT1 all have virtually identical spectral profiles. This is clear with reference to Figures 9 - 12 and results in the misclassification of some seeds by the proposed method. Future work will be focused on the proposition of techniques to assess how similar these species of rice seeds are and to mitigate degrading effects on the classifiers' performance due to the level of similarities. While it is believed that misclassifications occur due to species exhibiting similar properties, it should be noted that that the observed drop in performance could also be due to the limited number of rice seeds/kernels available versus what

is required to cover the large feature space. Going forward, the effects of varying and, in particular, increasing the number of rice seeds per class will be explored.

## V. CONCLUSION

This paper presents a new RGB and HSI system for rice seed variety classification. RGB images and hyperspectral image data cubes which offer high spatial and spectral resolution respectively were acquired using the proposed system. A large number of rice seed species (90 specifically) were selected and the spatial and spectral features extracted from the acquired images and data cube constitute the dataset which was used for this work and is made publicly available. Experimental results show that very good classification results and elimination of impure species from rice seed samples can be achieved by taking advantage of spatial features from high spatial resolution images and fusing them with spectral features from hyperspectral data cubes. Suboptimal performance reported for some categories was linked to the use of samples with large number of species and similarities among species. Future work will focus on assessing the similarities among species of rice seeds and exploring ways in which the degrading effects on classifiers' performance due to seed similarty can be mitigated. We will also aim to explore how increasing the number of seeds per class during training can improve the results. Finally, we intend to evaluate the performance of the proposed techniques using an NIR spectral data (1000-1700nm). In this work, our novel approach was applied on a large set of 90 rice seeds provided by the National Center of Protection of New Varieties and Goods of Plants (NCPNVGGP) in Vietnam. This approach can be extended to datasets with larger number of species which are available at other agricultural institutes.

and the National Centre of Protection of New Varieties and Goods of Plants (NCPNVGGP) in Vietnam for providing the rice seeds for this work.

## REFERENCES

[1] Y. Ogawa, "Quality evaluation of Rice," in *Computer Vision Technology for Food Quality Evaluation*. Amsterdam, The Netherlands: Elsevier, 2016, pp. 377–400.

[2] H. Vu, C. Tachtatzis, P. Murray, D. Harle, T. K. Dao, T.-L. Le, I. Andonovic, and S. Marshall, "Spatial and spectral features utilization on a hyperspectral imaging system for Rice seed varietal purity inspection," in *Proc. IEEE RIVF Int. Conf. Comput. Commun. Technol., Res., Innov., Vis. Future (RIVF)*, Nov. 2016, pp. 169–174.

[3] L. Wang, D. Liu, H. Pu, D.-W. Sun, W. Gao, and Z. Xiong, "Use of hyperspectral imaging to discriminate the variety and quality of Rice," *Food Anal. Methods*, vol. 8, no. 2, pp. 515–523, Feb. 2015.

[4] K.-Y. Huang and M.-C. Chien, "A novel method of identifying paddy seed varieties," *Sensors*, vol. 17, no. 4, p. 809, Apr. 2017.

[5] C. Fang, L. Zhao-yan, and Y. Yi-bin, "Machine vision analysis of characteristics and image information base construction for hybrid Rice seed," *Rice Sci.*, vol. 12, no. 1, pp. 13–18, 2005.

[6] P. T. T. Hong, T. T. T. Hai, L. T. Lan, V. T. Hoang, V. Hai, and T. T. Nguyen, "Comparative study on vision based Rice seed varieties identification," in *Proc. 7th Int. Conf. Knowl. Syst. Eng. (KSE)*, Oct. 2015, pp. 377–382.

[7] Z.-Y. Liu, F. Cheng, Y.-B. Ying, and X.-Q. Rao, "Identification of Rice seed varieties using neural network," *J. Zhejiang Univ. Sci. B*, vol. 6B, no. 11, pp. 1095–1100, Nov. 2005.

[8] C. N. M. Peralta, J. P. Pabico, and V. Y. Mariano, "Modeling shapes using uniform cubic B-splines for Rice seed image analysis," in *Proc. IEEE 6th Int. Conf. Commun. Electron. (ICCE)*, Jul. 2016, pp. 326–331.

[9] W. Kong, C. Zhang, F. Liu, P. Nie, and Y. He, "Rice seed cultivar identification using near-infrared hyperspectral imaging and multivariate data analysis," *Sensors*, vol. 13, no. 7, pp. 8916–8927, Jul. 2013.

[10] D.-W. Sun, *Computer Vision Technology for Food Quality Evaluation*. Amsterdam, The Netherlands: Elsevier, 2008.

[11] T.-Y. Kuo, C.-L. Chung, S.-Y. Chen, H.-A. Lin, and Y.-F. Kuo, "Identifying Rice grains using image analysis and sparse-representation-based classification," *Comput. Electron. Agricult.*, vol. 127, pp. 716–725, Sep. 2016.

[12] A.-G. Ouyang, R.-J. Gao, Y.-D. Liu, X.-D. Sun, Y.-Y. Pan, and X.-L. Dong, "An automatic method for identifying different variety of Rice seeds using machine vision technology," in *Proc. 6th Int. Conf. Natural Comput.*, vol. 1, Aug. 2010, pp. 84–88.

[13] A. A. Aznan, I. H. Rukunudin, A. Y. M. Shakaff, R. Ruslan, A. Zakaria, and F. S. A. Saad, "The use of machine vision technique to classify cultivated Rice seed variety and weedy Rice seed variants for the seed industry," *Int. Food Res. J.*, vol. 23, pp. S31–S35, 2016.

[14] A. R. Pazoki, F. Farokhi, and Z. Pazoki, "Classification of Rice grain varieties using two artificial neural networks (mlp and neuro-fuzzy)," *J. Anim. Plant Sci.*, vol. 24, no. 1, pp. 336–343, 2014.

[15] J. Sun, X. Lu, H. Mao, X. Jin, and X. Wu, "A method for rapid identification of Rice origin by hyperspectral imaging technology," *J. Food Process Eng.*, vol. 40, no. 1, Feb. 2017, Art. no. e12297.

[16] K. R. Singh and S. Chaudhury, "Efficient technique for Rice grain classification using back-propagation neural network and wavelet decomposition," *IET Comput. Vis.*, vol. 10, no. 8, pp. 780–787, Dec. 2016.

[17] A. Polak, F. K. Coutts, P. Murray, and S. Marshall, "Use of hyperspectral imaging for cake moisture and hardness prediction," *IET Image Process.*, vol. 13, no. 7, pp. 1152–1160, May 2019.

[18] T. Qiao, J. Ren, C. Craigie, J. Zabalza, C. Maltin, and S. Marshall, "Quantitative prediction of beef quality using visible and NIR spectroscopy with large data samples under industry conditions," *J. Appl. Spectrosc.*, vol. 82, no. 1, pp. 137–144, Mar. 2015.

[19] T. Kelman, J. Ren, and S. Marshall, "Effective classification of Chinese tea samples in hyperspectral imaging," *Artif. Intell. Res.*, vol. 2, no. 4, pp. 1–13, 2013.

[20] F. S. Lai, I. Zayas, and Y. Pomeranz, "Application of pattern recognition techniques in the analysis of cereal grains," *Cereal Chem.*, vol. 63, no. 2, pp. 168–172, 1986.

[21] N. Sakai, S. Yonekawa, A. Matsuzaki, and H. Morishima, "Two-dimensional image analysis of the shape of Rice and its application to separating varieties," *J. Food Eng.*, vol. 27, no. 4, pp. 397–407, Jan. 1996.

[22] Y. Shao, C. Zhao, Y. Bao, and Y. He, "Quantification of nitrogen status in Rice by least squares support vector machines and reflectance spectroscopy," *Food Bioprocess Technol.*, vol. 5, no. 1, pp. 100–107, Jan. 2012.

[23] Z. Zhang, "A flexible new technique for camera calibration," Microsoft Res. Microsoft Corp. One Microsoft Way, Redmond, WA, USA, Tech. Rep. MSR-TR-98-71, 2000.

[24] X. Bai, "New class of top-hat transformation to enhance infrared small targets," *J. Electron. Imag.*, vol. 17, no. 3, Jul. 2008, Art. no. 030501.

[25] P. Geladi, J. Burger, and T. Lestander, "Hyperspectral imaging: Calibration problems and solutions," *Chemometrics Intell. Lab. Syst.*, vol. 72, no. 2, pp. 209–217, Jul. 2004.

[26] W. Li, F. Feng, H. Li, and Q. Du, "Discriminant analysis-based dimension reduction for hyperspectral image classification: A survey of the most recent advances and an experimental comparison of different techniques," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 1, pp. 15–34, Mar. 2018.

[27] A. Tharwat, T. Gaber, A. Ibrahim, and A. E. Hassanien, "Linear discriminant analysis: A detailed tutorial," *AIC*, vol. 30, no. 2, pp. 169–190, May 2017.

[28] A. Zimek, E. Schubert, and H.-P. Kriegel, "A survey on unsupervised outlier detection in high-dimensional numerical data," *Stat. Anal. Data Mining*, vol. 5, no. 5, pp. 363–387, Oct. 2012.

[29] R. J. Durrant and A. Kabán, "When is 'nearest neighbour' meaningful: A converse theorem and implications," *J. Complex.*, vol. 25, no. 4, pp. 385–397, Aug. 2009.

**SAMSON DAMILOLA FABIYI** (Student Member, IEEE) received the B.Eng. degree in electrical and electronic engineering from the University of Ilorin, Nigeria, in 2015, and the M.Sc. degree in electronic and electrical engineering from the University of Strathclyde, Glasgow, in 2018, where he is currently pursuing the Ph.D. degree with the Department of Electronic and Electrical Engineering. His research interests include signal/image processing, hyperspectral imaging, machine learning, and the Internet of Things and applications in precision agriculture.

**HAI VU** received the B.E. degree in electronic and telecommunications and the M.E. degree in information processing and communication from the Hanoi University of Science and Technology (HUST), in 1999 and 2002, respectively, and the Ph.D. degree in computer science from Osaka University, Japan, in 2009. He has been a Lecturer and a Researcher with the Department of Computer Vision, MICA International Research Institute (HUST - Grenoble INP), since 2012. His current research interests are in computer vision, pattern recognition, particularly, applying these techniques in agricultural engineering, medical imaging, and human-computer interactions.

**CHRISTOS TACHTATZIS** received the B.Eng. (Hons.) degree in communication systems engineering from the University of Portsmouth, U.K., in 2001, and the M.Sc. and Ph.D. degrees in communications, control and digital signal processing and electronic and electrical engineering from the University of Strathclyde, Glasgow, U.K., in 2002 and 2008, respectively. From 2008 to 2016, he held a Postdoctoral Researcher positions at the University of Strathclyde, U.K., and the Letterkenny Institute of Technology, Ireland. Since 2016, he has been a Lecturer with the Department of Electronic and Electrical Engineering. His research interests include distributed sensing, communication networks, edge computing analytics, computer vision, and artificial intelligence. He is a member of the IET. He received the Strathclyde Chancellor's Fellow Award, in 2016.

**PAUL MURRAY** received the M.Eng. degree in electronic and electrical engineering from the University of Strathclyde, Glasgow, U.K., in 2008, and the Ph.D. degree from the University of Strathclyde, in 2012. He is currently a Lecturer with the University of Strathclyde. His research interests include image processing, hyperspectral imaging and analysis, feature extraction, and machine learning.

**DAVID HARLE** (Member, IEEE) is currently a Senior Lecturer with the Broadband Research Group, Department of Electronic and Electrical Engineering. His current research interests within the Broadband & Optical Networks group ARE concerned with performance evaluation, design and management issues associated with current and future wireless, broadband and optical communication systems. Of particular interest are network survivability and the operation of routing protocols in multiservice environments. He is the author of over 100 international conference and journal publications, one technical text book, as well as three patents and undergraduate texts.

**TRUNG KIEN DAO** received the B.S. and M.S. degrees in information technology from the University of Pierre and Marie Curie, France, in 2004 and 2006, respectively, and the Ph.D. degree in mechanical and automation engineering from Da-Yeh University, Taiwan, in 2010. He works as a Lecturer-Researcher at the Department of Pervasive Spaces and Interaction, MICA International Research Institute (HUST - Grenoble INP), Hanoi University of Science and Technology (HUST), since 2010. His current research interests include the area of ubiquitous computing, particularly multimodal localization, location-based services, context awareness, and user adaptation in pervasive environments.

**IVAN ANDONOVIC** (Senior Member, IEEE) has held a Royal Society Industrial Fellowship in collaboration with British Telecommunications (BT) Labs investigating novel approaches to broadband networking. He has edited two books and authored/coauthored six chapters in books and over 380 journal and conference papers and secured funding for research and development in excess of £10M. He is a Fellow of the IET. He was a member of flagship Scottish Enterprise (Government agency for economic growth) team of the Intermediary Technology Institutes (ITIs), aimed at bridging the gap between basic research and company growth, has been a Visiting Scientist at the Communications Research Laboratories of Japan, a Visiting Professor with the City University of Hong Kong and Princeton University, the Topical Editor for the IEEE Transactions on Communications and the Technical Programme Co-Chair for the IEEE International Conference in Communications (ICC07).

**JINCHANG REN** (Senior Member, IEEE) received the B.E. degree in computer software, the M.Eng. degree in image processing, and the D.Eng. degree in computer vision from Northwestern Polytechnical University, Xi'an, China, and the Ph.D. degree in electronic imaging and media communication from the University of Bradford, Bradford, U.K. He is currently a Reader with the Centre in Signal and Image Processing (CeSIP), and also the Deputy Director of the Strathclyde Hyperspectral Imaging Centre, University of Strathclyde, Glasgow, U.K. His research interests focus mainly on hyperspectral imaging, image processing, computer vision, big data analytics, and machine learning. He has published over 280 peer reviewed journal/conferences papers, and acts as an Associate Editor for several international journals, including the IEEE JSTARS and the *Journal of The Franklin Institute*.

**STEPHEN MARSHALL** received the B.Sc. degree in electrical and electronic engineering from the University of Nottingham and the Ph.D. degree in image processing from the University of Strathclyde, U.K. He is currently a Professor with the Department of Electronic and Electrical Engineering, University of Strathclyde. With more than 150 papers published, his research activities focus in nonlinear image processing and hyperspectral imaging. He is a Fellow of the IET.

• • •